

Logic and Argumentation

87
Studies
in Logic

Reason to Dissent

**Proceedings of the 3rd
European Conference on
Argumentation, Volume III**

Editors
Catarina Dutilh Novaes
Henrike Jansen
Jan Albert van Laar
Bart Verheij

Studies in Logic

Logic and Argumentation

Volume 87

Reason to Dissent

Proceedings of the 3rd European
Conference on Argumentation

Volume III

Volume 68

Logic and Philosophy of Logic. Recent Trends in Latin America and Spain
Max A. Freund, Max Fernández de Castro and Marco Ruffino, eds.

Volume 79

Games Iteration Numbers. A Philosophical Introduction to Computability
Theory

Luca M. Possati

Volume 80

Logics of Proofs and Justifications
Roman Kuznets and Thomas Studer

Volume 81

Factual and Plausible Reasoning
David Billington

Volume 82

Formal Logic: Classical Problems and Proofs
Luis M. Augusto

Volume 83

Reasoning: Games, Cognition, Logic
Mariusz Urbański, Tomasz Skura and Paweł Łupkowski, eds.

Volume 84

Witness Theory. Notes on λ -calculus and Logic
Adrian Rezuş

Volume 85

Reason to Dissent. Proceedings of the 3rd European Conference on
Argumentation. Volume I
Catarina Dutilh Novaes, Henrike Jansen, Jan Albert van Laar and Bart
Verheij, eds.

Volume 86

Reason to Dissent. Proceedings of the 3rd European Conference on
Argumentation. Volume II
Catarina Dutilh Novaes, Henrike Jansen, Jan Albert van Laar and Bart
Verheij, eds.

Volume 87

Reason to Dissent. Proceedings of the 3rd European Conference on
Argumentation. Volume III
Catarina Dutilh Novaes, Henrike Jansen, Jan Albert van Laar and Bart
Verheij, eds.

Volume 87

Reason to Dissent. Proceedings of the 3rd European Conference on
Argumentation. Volume III
Catarina Dutilh Novaes, Henrike Jansen, Jan Albert van Laar and Bart
Verheij, eds.

Studies in Logic Series Editor
Dov Gabbay

dov.gabbay@kcl.ac.uk

Reason to Dissent
Proceedings of the 3rd European
Conference on Argumentation
Volume III

Edited by
Catarina Dutilh Novaes,
Henrike Jansen,
Jan Albert van Laar
and
Bart Verheij

© Individual author and College Publications, 2020
All rights reserved.

ISBN 978-1-84890-333-3

College Publications
Scientific Director: Dov Gabbay
Managing Director: Jane Spurr

<http://www.collegepublications.co.uk>

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise without prior permission, in writing, from the publisher.

Conference Organization

Organising committee

Catarina Dutilh Novaes (Vrije Universiteit Amsterdam)
Henrike Jansen (Leiden University)
Jan Albert Van Laar (chair) (University of Groningen)
Bart Verheij (University of Groningen)

Programme committee

Dale Hample (chair) (University of Maryland)
Henrike Jansen (secretary) (Leiden University)
Marcin Lewinski (Universidade Nova de Lisboa)
Juho Ritola (University of Turku)
Frank Zenker (Lund University)

ECA steering committee

Lilian Bermejo Luque (University of Granada)
Katarzyna Budzyńska (Warsaw University of Technology)
Henrike Jansen (Leiden University)
Marcin Koszowy (Warsaw University of Technology)
Jan Albert Van Laar (deputy chair) (University of Groningen)
Marcin Lewiński (Universidade Nova de Lisboa)
Dima Mohammed (Universidade Nova de Lisboa)
Steve Oswald (University of Fribourg)
Fabio Paglieri (chair) (ISTC-CNR)
Juho Ritola (University of Turku)
Sara Rubinelli (University of Lucerne)
Frank Zenker (University of Lund)

Scientific Panel

Mark Aakhus (Rutgers University)
Andrew Aberdein (Florida Institute of Technology)
Scott Aikin (Vanderbilt University)
Mehmet Ali Uzelgun (Universidade NOVA de Lisboa)
Ruth Amossy (Tel-Aviv University)
Corina Andone (University of Amsterdam)
Michael Baumtrog (Ryerson University)
Gregor Betz (KIT Karlsruher)
Sarah Bigi (Università Cattolica del Sacro Cuore)
Anthony Blair (University of Windsor)
Patrick Bondy (Cornell University)
Alessandra Von Burg (Wake Forest University)
John Casey (Northeastern Illinois University)
Daniel Cohen (Colby College)
Louis De Saussure (University of Neuchatel)

Marianne Doury (Paris Descartes University)
 Michel Dufour (Université Sorbonne-Nouvelle)
 Isabela Fairclough (University of Cantral Lancashire)
 Eveline Feteris (University of Amsterdam)
 Joana Garmendia (University of the Basque Country)
 Bart Garssen (University of Amsterdam)
 Ingeborg van der Geest (University of Amsterdam)
 David Godden (Michigan State University)
 G.C. Goddu (University of Richmond)
 Jean Goodwin (North Carolina State University)
 Floriana Grasso (University of Liverpool)
 Sara Greco (Università della Svizzera Italiana)
 Leo Groarke (Trent University)
 Thierry Herman (Universities of Neuchatel and lausanne)
 Martin Hinton (University of Lodz)
 David Hitchcock (McMaster University)
 Hans Hoeken (Utrecht University)
 Michael Hoffmann (Georgia Institute of Technology)
 Michael Hoppmann (Northeastern University)
 Jos Hornikx (Radboud University)
 Catherine Hundleby (University of Windsor)
 Constanza Ihnen (University of Chile, Law Faculty, Institute of Argumentation)
 Beth Innocenti (University of Kansas)
 Jérôme Jacquin (University of Lausanne)
 Henrike Jansen (Leiden University)
 Amy Johnson (University of Oklahoma)
 Manfred Kienpointner (University of Innsbruck)
 Gabrijela Kisicek (University of Zagreb)
 Susan Kline (The Ohio State University)
 Bart van Klink (Vrije Universiteit Amsterdam)
 Manfred Kraus (University of Tübingen)
 Maarten van Leeuwen (Leiden University)
 Christoph Lumer (University of Siena)
 Fabrizio Macagno (Universidade NOVA de Lisboa)
 Andrei Moldovan (University of Salamanca)
 Paula Olmos (Universidad Autónoma de Madrid)
 Rudi Palmieri (University of Liverpool)
 Steven Patterson (Marygrove College)
 Niki Pfeifer (Munich Center for Mathematical Philosophy)
 Gilbert Plumer (Law School Admission Council) (retired)
 Lotte van Poppel (University of Amsterdam)
 Henry Prakken (University of Groningen)
 Martin Reisigl (University of Bern)
 Adam Richards (Texas Christian University)
 Juho Ritola (University of Helsinki)
 Andrea Rocci (Università della Svizzera Italiana)
 Harvey Siegel (University of Miami)
 Francisca Snoeck Henkemans (University of Amsterdam)
 Manfred Stede (University of Potsdam)

Katharina Stevens (University of Lethbridge)
Christopher Tindale (University of Windsor)
Assimakis Tseronis (Örebro University)
Mariusz Urbanski (Adam Mickiewicz University)
Hans Vilhelm Hansen (University of Windsor)
Laura Vincze (University of Pisa)
Jacky Visser (University of Dundee)
Lena Wahlberg (Lund University)
Harry Weger (University of Central Florida)
David Williams (Florida Atlantic University)
Harald Wohlrapp (Universität Hamburg)
Carly Woods (University of Maryland)
Yun Xie (Sun Yat-sen University)
Minghui Xiong (Sun Yat-sen University)
Igor Z. Zagar (Educational Research Institute)
Marta Zampa (Zurich University of Applied Sciences)
David Zarefsky (Northwestern University)
Gábor Zemplén (Eötvös Loránd University)

Preface

After two successful editions held in Lisbon in 2015 and Fribourg in 2017, ECA was hosted in 2019 by the Faculty of Philosophy of the University of Groningen, on 24-27 June 2019. These three volumes contain the Proceedings of this third edition of the conference series, whose special theme was Reason to Dissent.

The European Conference on Argumentation (ECA) is a pan-European biennial initiative aiming to consolidate and advance various strands of research on argumentation and reasoning by gathering scholars from a range of disciplines. While based in Europe, ECA involves and encourages participation by argumentation scholars from all over the world; it welcomes submissions linked to argumentation studies in general, in addition to those tackling the conference theme. The 2019 Groningen edition focused on dissent. The goal was to inquire into the virtues and vices of dissent, criticism, disagreement, objections, and controversy in light of legitimizing policy decisions, justifying beliefs, proving theorems, defending standpoints, or strengthening informed consent. It is well known that dissent may hinder the cooperation and reciprocity required for reason-based deliberation and decision-making. But then again, dissent also produces the kind of scrutiny and criticism required for reliable and robust outcomes. How much dissent does an argumentative practice require? What kinds of dissent should we promote, or discourage? How to deal with dissent virtuously? How to exploit dissent in artificial arguers? How has dissent been conceptualized in the history of rhetoric, dialectic and logic? The papers in these three volumes discuss these and other questions pertaining to argumentation and dissent (among other themes).

ECA 2019 had 224 participants and 188 paper presentations, a clear indication that ECA continues to fulfill its role as a key platform of scholarly exchange in the field. These three volumes reflect the current state of the art in argumentation scholarship in general.

The proceedings contain papers that were accepted based on abstract submissions; each submission was thoroughly evaluated by three reviewers of our scientific board—for a full list of ECA committees, see www.ecargument.org. Volume I gathers 25 long papers and associated commentaries, together with 9 papers presented in the thematic panels that were held during ECA2019. Volumes II and III gather 69 regular papers that were presented during the conference.

Many people have contributed to the success of ECA 2019, and for the completion of the Proceedings. First of all, we must thank all members of our Scientific Panel and of our Programme Committee, thanks to whom we were able to select papers of the highest quality. In Groningen, thanks to those who provided organizational support, in particular the team of student assistants (especially Johan Rodenburg) who ensured that the conference was a pleasant experience to all participants. Our heartfelt thanks go to Jelmer van der Linde and Annet Onnes, who accomplished the gigantic task of putting all the papers together into these three volumes, and assisted us throughout in the process of producing the Proceedings. Thanks also to the European Research Council for generously supporting the production of the Proceedings by means of grant ERC-17-CoG 771074 for the project ‘The Social Epistemology of Argumentation’ (PI C. Dutilh Novaes).

The next edition of ECA will take place in Rome in 2021, and we look forward to seeing the ECA community gathering again for another successful event.

Catarina Dutilh Novaes, Henrike Jansen, Jan Albert van Laar, Bart Verheij

Keynote Speakers

Critical thinking as discourse

Deanna Kuhn – Columbia University

Less than it is an individual ability or skill, critical thinking is a dialogic practice people engage in and commit to, initially interactively and then in interiorized form with the other only implicit. An argument depends for its meaning on how others respond (Gergen, 2015). In advancing arguments, well-practiced thinkers anticipate their defeasibility as a consequence of others' objections, in addition envisioning their own potential rebuttals. Whether in external or interiorized form, the dialogic process creates something new, while itself undergoing development.

This perspective may be useful in sharpening definition of the construct of critical thinking and in so doing help to bring together the largely separate strands of work examining it as a theoretical construct, a measurable skill, and an educational objective. Implications for education follow. How might critical thinking as a shared practice be engaged in within educational settings in ways that will best support its development? One step is to privilege frequent practice of direct peer-to-peer discourse. A second is to take advantage of the leveraging power of dialog as a bridge to individual argument – one affording students' argumentative writing a well-envisioned audience and purpose. Illustrations of this bridging power are presented. Finally, implications for assessment of critical thinking are noted and a case made for the value of people's committing to a high standard of critical thinking as a shared and interactive practice.

Revisiting *Apologie de la polémique*: about some “felicity conditions” allowing for coexistence in dissent

Ruth Amossy – Tel-Aviv University

In my book entitled *Apologie de la polémique* (2014), I claimed that polemical discourse fulfils various social functions, among which “coexistence in dissensus” seems the most important. It means not only that disagreement is the basis of life in society, and the principle on which argumentation as a common, rational search for the reasonable, is built. It also signifies that agreement cannot always be reached in democratic societies recognizing the importance of diversity and difference, so that disagreement has to be managed through verbal confrontations, namely, agonistic discussions and polemical exchanges. It thus appears that the latter, though generally blamed for its radicalization and polarization, plays an important role in the public sphere. Among others, public polemics helps opposite parties to voice conflicting opinions and fight for antagonistic solutions without recurring to arms. To use Chantal Mouffe's words, it transforms “enemies” to be destroyed into “adversaries” who have a right to speak. Beside other social functions discussed in the book, polemics authorizes what the French call a “*vivre-ensemble*” – the possibility for people who do not share the same opinions, if not the same premises, to share the same national space and live together without outbursts of violence.

However, the emphasis on dissent and its polemical management is not without raising multiple questions concerning the conditions of possibility and the limits of the so-called coexistence in dissent. Obviously, the use of polemical discourse is not enough to prevent citizens from physically fighting each other and even, sometimes, to dispel the specter of civil war. Outbursts of violence against refugees regularly occur in Germany where the polemical discussion is vivid. In France, the polemical exchanges on Emmanuel Macrons' reforms and the authorized street demonstrations did not prevent urban violence. Even if polemical campaign discourse is tolerated, it did not prevent armed confrontations in certain African countries such as Ivory Coast. What, then, are the "felicity conditions" needed in order for public polemics to secure a peaceful "living together" in the framework of persistent and sometimes deep disagreements that can hardly be avoided in the democratic space? My contention is that to answer this question, it is necessary to explore polemical confrontations in their institutional framework, and to examine the functioning of polemical discourse in relation to the political, forensic and cultural factors that determine (at least partly) its degree of success. After synthesizing the finding of my first research into dissent and its polemical management, I will try – on the basis of a few case studies – to gather some of the "felicity conditions" necessary to make coexistence in dissent possible.

Dissent needed: argumentation for AI and law applications

Katie Atkinson – University of Liverpool

As technological advances in artificial intelligence are being turned into deployed products, societal questions are being raised about the need for AI tools to be able to explain their decisions to humans. This need becomes even more pressing when AI technologies are applied in domains where critical decisions are made that can result in a significant effect upon individuals or groups in society. One such domain is law, where there is a thriving market developing in support tools for assisting with a variety of legal tasks carried out within law firms and the wider legal sector. Law is a domain rich in argumentation and support tools that are used to aid legal decision making should similarly be able to explain why a particular outcome of a decision has been reached, and not an alternative outcome. Dissent needs to be captured and revealed within AI reasoners to ensure that the decision space is explored from different perspectives, if AI tools are to be deployed effectively to assist with legal reasoning tasks. In this talk I will discuss a body of work on computational models of argument for legal reasoning and show how dissent features within this work to promote scrutability of AI decision making.

Contents

Conference Organization	v
Preface	ix
Keynote Speakers	xi
 Regular papers	 1
The Spaces and Places of Argumentation: Developing a Spatial Model of Argumentation	3
<i>Steven Johnson</i>	
Aggressiveness as a Qualifier of Dissent in Interpersonal Arguing . . .	23
<i>Iryna Khomenko</i>	
Auditory Arguments – Importance of Sound in an Argumentative Discourse (An Empirical Study)	35
<i>Gabrijela Kišiček</i>	
Be reasonable! Ways to React to Cases of Presumed Unreason	49
<i>Erik C. W. Krabbe, Jan Albert van Laar</i>	
Epistemic Injustice and Deep Disagreement	63
<i>Thirza Lagewaard</i>	
Argumentative Patterns of Right-Wing Populism: The Example of the “Alternative for Germany”	77
<i>David Lanius</i>	
A Decision Tree for Annotating Argumentation Scheme Corpora . . .	97
<i>John Lawrence, Jacky Visser, Douglas Walton, Chris Reed</i>	
Philosophical Foundation of Reasonableness in Mencius’s Argumentative Discourse: Based on the use of dissociation	115
<i>Yan Linqiong, Xiong Minghui</i>	
Strategic Manoeuvring with the Construction “We (All) Know that X” in Parliamentary Debates	127
<i>Matthijs Looij, Bertijn van der Steenhoven</i>	

Argument, Justification and Inquiry: a Pragmatist Approach	143
<i>Federico E. López</i>	
Dissent as a Voice of Support: The Influence of Argument Types in Providing Support to Those Who have Experienced Racial Discrimination	153
<i>Uttara Manohar, Susan L. Kline</i>	
Holism of Reasons and its Consequences for Argumentation Theory .	167
<i>Hubert Marraud</i>	
“Why did not your correspondent make an honest inquiry before so writing?”: The Text Structure and Discourse of Disagreement in Irish letters to the Editor	181
<i>Davide Mazzi</i>	
An Argumentative Approach to the Justification of Abduction	195
<i>Paula Olmos</i>	
When You Can’t Trust What You Don’t Know: Organ and Tissue Donation, Public Trust, the Voice of Dissent and Informed Consent	213
<i>Harmony Peach</i>	
“The worst ever conceived by a man of genius. ” Hume’s Probability Argument in A Treatise	225
<i>Jeanne Peijnenburg, David Atkinson</i>	
Character, Dog Whistles, and the Limits of Charity	239
<i>Kathryn Phillips</i>	
The Rhetorical Potential of Metaphor: An Experimental Study on the Effect of Metaphors in Argumentation	251
<i>Roosmaryn Pilgram</i>	
Separate Opinions as an Argumentative Activity Type	267
<i>H. José Plug</i>	
Can Literary Fiction be Suppositional Reasoning?	279
<i>Gilbert Plumer</i>	
Conspiracy Theories and Reasonable Disagreement	291
<i>Juho Ritola</i>	
Inference to the Best Metaphor	305
<i>Cristián Santibáñez</i>	
Various Efforts of Enhancing Real World Online Discussions	311
<i>Alexander Schneider, Christian Meter</i>	

Beyond Randomized Clinical Trials: Emerging Innovations in Reasoning About Health	325
<i>Jodi Schneider, Sally Jackson</i>	
Multimodal Argumentation and Dissent – a Perspective of Multimodal Critical Discourse Analysis	339
<i>Andrea Sabine Sedlaczek</i>	
Analogical Argumentation in Philosophical Thought Experiments . .	355
<i>Stefan Sleeuw</i>	
Detecting and Handling Disagreement in Multi-party Health Coaching	369
<i>Mark Snaith, Alison Pease, Dominic De Franco</i>	
Explaining Away and the Propensity Interpretation of Probability: the Case of Unequal Priors	385
<i>Alice Liefgreen, Marko Tešić</i>	
Responsible Agent Deliberation	405
<i>Douglas Walton, Alice Toniolo</i>	
Disagreement in Aristotle’s Topics and its Latin Medieval Reception .	421
<i>Gustavo Fernández Walker, Ana María Mora-Márquez</i>	
Conductive Arguments. Towards a Solution of the Riddles	435
<i>Harald R. Wohlrapp</i>	
A Ludological Perspective on the Shape of Argument: Collaborative Assent to Dissenting Opposition	449
<i>Michael A. Yong-Set</i>	
Towards the Justification of Warrant in a Legal Hard Case: How to Justify a Legal Rule by the Backing	463
<i>Shiyang Yu, Xi Chen</i>	
How Should We Classify Argument Schemes?	477
<i>David Zarefsky</i>	

Regular papers

The spaces and places of argumentation: Developing a spatial model of argumentation

STEVEN L. JOHNSON

*Department of Communication, University of Alaska Anchorage,
USA*

sljohnson@alaska.edu

In this paper, I argue that the cognition on which argumentation depends is intimately related to spatial reasoning, which in turn opens the door to new approaches to modeling argumentation. Argument modeling has proved useful for argumentation studies and pedagogy but largely is limited to conceiving of arguments as static, isolated phenomena. Building on the successes of argument mapping, this paper proposes a model for mapping the space of argumentation and imagines the use of such a tool in argument pedagogy.

KEYWORDS: abstract reasoning, argument, argument mapping, argumentation, argumentation pedagogy, cognition, space, spatial reasoning.

1. INTRODUCTION

Argumentation is goal-driven; most arguments exist for the ultimate purpose of convincing an audience to accept or reject a proposition. The discipline of argumentation studies recognizes this imperative as a call for attention to the normative pragmatics of argumentation: insights about argumentation are valuable insofar as they offer guidance to arguers seeking this goal. The intimate relationship between argumentation theory and practice is the product of this perspective.

Hence the practical focus of argumentation theory. This focus has produced a wealth of highly-utilitarian resources to improve arguer's ability to argue effectively, resources that are in demand now more than ever. (Kuhn, 2009) While advancement in teaching people to argue effectively has been steady, a need for further development of effective pedagogical tools and techniques continues to exist. (Kuhn, Hemberger, & Khait, 2017; Millard & Menzies, 2015)

Attention to understanding how we conceive of arguments and their functioning will continue to produce such advancements. In particular, the present survey attempts to understand the cognitive space

within which arguments function with an eye toward improving our ability to argue effectively.

2. COGNITION AND SPACE

The study of human cognition increasingly is converging around the conviction that understanding how humans think requires attention to the evolutionary forces that shape our mechanisms of cognition. Such an idea seems, if not inevitable, at least overdue. The notion that *homo sapiens* are the product of the challenges we faced and the adaptations that helped us overcome those challenges obviously is not a new idea; application of that perspective to the machinery of cognition is opening doors to a richer understanding of the means by which we think.

Of the physiological and psychological mechanisms that natural selection produced, Pinker (1999) argues that the visual perception organs and their functions are foundational to human cognition: “The repercussions reach to the rest of our psyche. We are primates—highly visual creatures—with minds that evolved around this remarkable sense.” (p. 214) Given how important recognizing threats and opportunities is to the selection of those who survive, it’s no surprise that the underlying organs devoted to these tasks would come to dominate our thinking. Visual perception is directly connected to spatial cognition; we rely upon both to develop our understanding of the world around us and where we are in it. Pinker’s point is that sense of sight and the spatial cognition attached to it are both the product of our environment and, consequently, the foundation of our approach to thinking about that environment.

This connection between the visual and the rest of our mind is increasingly being explored by those interested in one area of cognition in particular: abstract thought. Casasanto (2009) explains: “Many linguists have noted that when people talk about states, possessions, ideas and desires, they do so by co-opting the language of intuitive physics. In particular, words borrowed from physical domains of space, force, and motion, give rise to linguistic metaphors for countless abstract ideas.” (p. 456) Other contributions to the study of the relationship between thinking and space come from cognitive science. Talmy (1988) explores the relationship between our conceptions of force and our linguistic representations of abstract ideas. Gärdenfors (2004, 2014) offers a complete theory of conceptual formation explained by the geometry of conceptual spaces.

In these and other cases, insight into how cognition functions is gained by examining the relationship between what had long been seen as different ways of thinking: spatial cognition was one mode of thought; abstract reasoning was another. These scholars are challenging that assumption: could it be that abstract reasoning is the direct descendent

of spatial reasoning? It's one thing to suggest that relationships exist between different modes of thought and the organs on which we rely to produce that thought. It's quite another to suggest that the cognitive organs on which we rely to undertake different operations might not be distinct at all but are, in fact, different versions of the same organ. Pinker (2009) explains:

Evolutionary change often works by copying body parts and tinkering with the copy. For example, insects' mouth parts are modified legs. A similar process may have given us our language of thought. Suppose ancestral circuits for reasoning about space and force were copied, the copy's connections to the eyes and muscles were severed, and references to the physical world were bleached out. The circuits could serve as scaffolding whose slots are filled with symbols for more abstract concerns like states, possessions, ideas and desires. The circuits computational abilities, continuing to reckon about entities being in one state at a time, shifting from state to state, and overcoming entities with opposite valence. When the new, abstract domain has a logical structure that mirrors objects in motion—a traffic light has one color at a time but flips between them; contested social interactions are determined by the stronger of two wills—the old circuits can do useful inferential work. They divulge their ancestry as space- and force-simulators by the metaphors they invite, a kind of vestigial cognitive organ. (p. 355-356)

Dubbed *exaptation* by Stephen Gould and colleagues (1982), the pressing of cognitive processes into service for which they were not originally adapted may help explain why metaphors of space and force are so prevalent in our languages. The prevalence of these metaphors lead Pinker (2009) to conclude that their use wasn't merely a coincidence but revealed something far more significant: "Models of space and force don't act like figures of speech intended to convey new insights; they seem closer to the medium of thought itself." (p. 357)

New research seems to support this contention. Building on Nobel Prize-winning work on the functioning of specialized "grid" and "place" neurons that function as the mind's "GPS system" for navigating physical space, Bellmund et. al. (2018) (in cooperation with the aforementioned Gärdenfors) have used neural imaging to demonstrate that these systems function similarly when employed for abstract thought. "The hippocampus' place and grid cells, in other words, map not only physical space but conceptual space. It appears that our representation of objects and concepts is very tightly linked with our representation of space." (Rajagopalan, 2019)

3. ARGUMENTS, ARGUMENTATION AND SPATIALITY

It should, at this point, come as little surprise that the prevalence of linguistic metaphors related to space and force that generally are pervasive in our languages are similarly prevalent in the language we use to talk about argumentation. Indeed, a cursory examination of the lexicon of argumentation studies reveals that those who argue *stake out* positions, define the *boundaries* of conflicts, *frame* controversies, argue from *standpoints* and expect that claims will rest upon acceptable *grounds*, all while seeking to *move* audiences, *advance* positions, *sway* opponents, *follow* lines of argument, take logical *leaps*, and engage in *strategic maneuvering*. The relationship between spatial reckoning and making arguments—at least as revealed by linguistic metaphors—is intimate.

This intimacy is based, in part, on the inherent logic of spatial relations. Pinker (2009) notices this when discussing the use of visualization to solve logical problems:

[Visualizing logical problems] supplies many truths of geometry for free. For example, left-to-right arrangement in space is transitive: if A is to the left of B, and B is to the left of C, then A is to the left of C. Any lookup mechanism that finds the locations of shapes in the array will automatically respect transitivity; the architecture of the medium leaves it no choice. (p. 291)

Consideration of the models used to represent relationships in set theory demonstrate the same principle. As Venn (1880) noticed, the relationships between concepts that share (or are distinguished by) characteristics can be represented by overlapping circles; Venn diagrams are spatial and, therefore, logical. While a thorough exploration of the spatial dimensions of arguments and argumentation is beyond the scope of my present effort, suffice it to say that the evidence for the connection is intriguing and that there may be much to be gained by more fully exploring the spatiality of argumentation.

One area of argumentation theory that has capitalized on and contributed to our understanding of this phenomenon is the mapping (or diagramming or visualization) of arguments. In its most basic form, argument mapping attempts to make plain the critical elements of arguments and the relationships between those elements. Most argument mapping follows a very similar approach: by representing arguments' elements and relationships graphically, an argument map distills the complexity of natural language argumentation to a standard form whose operations can be observed, understood and critiqued. Argument mapping emerged in parallel to the greater attention directed toward argumentation theory. (Goodwin, 2000) Such approaches have

demonstrated great utility in improving students' critical thinking skills (van Gelder, 2015) and iterative, guided practice in the parsing and diagramming of arguments has produced abundant evidence of impressive pedagogical outcomes. (Davies, 2012; Dwyer, Hogan, & Stewart, 2012; Kunsch, Schnarr, & van Tyle, 2014; Rider & Thomason, 2014; van Gelder, 2003, 2005; van Gelder, Bissett, & Cumming, 2004)

Broadly speaking, extant approaches to argument mapping fall into two categories that parallel those introduced by Woods (1995) in his discussion of the methods of analyzing fallacies. Woods (1995) employs the metaphor of a butcher preparing meat for sale to distinguish between arguments as they occur in natural language and those prepared for analysis. Groarke (2019) explains: "[A]rguments *on the hoof* are arguments as they appear in their real life contexts. One *dresses* them to identify and isolate their key components in a way that prepares the way for argument evaluation." (pt. 3.0; emphasis added) This distinction between "dressed" arguments and arguments "on the hoof" may be used to distinguish the current approaches to argument mapping: some argument maps function to reduce complex arguments to their essential elements; other approaches attempt to capture the complexity of natural language argumentation in its extant state. Woods' metaphor will also serve the basis for my contention that there is another, unexplored aspect of argumentation yet to be mapped. Before I turn my attention to this *terra incognita* of argument mapping, I'll review both the major contributions and state-of-the-art of both existing approaches.

3.1 Mapping "dressed" arguments

Argument mapping developed in parallel to the growing interest in the process and product of argumentation. Early evidence of the inclination toward mapping may be found in Whatley's *Elements of Logic* (1897), where he advises that "many students probably will find it a very clear and convenient mode of exhibiting the logical analysis of the course of argument, to draw it out in the form of a Tree, or Logical Division." (p. 253) Contributions to the approach from Wigmore (1913), Beardsley (1959) and Toulmin (1958) helped to cement mapping arguments as a useful prerequisite to studying and teaching argumentation. These early efforts—and the modern versions based on these early developments discussed below—largely take the same approach: arguments are distilled to their essential components; those components and the relationship between them are then laid out in a graphical representation. Relationships are typically depicted as hierarchical, demonstrating the connections between superior, subordinate and coordinate ideas.

More recently, argument mapping has moved from the static paper-and-pencil approach to the more dynamic digital environment.

Known generally as Computer-Supported Argument Visualization (CSAV), the effort is a significant leap forward for argument mapping. It's not necessary to trace the history of CSAV here in order to understand where it fits in the development of argument mapping; Buckingham Shum (2003) has produced a very good record of its development. As Kirshner et. al. (2012) argue in the preface to their book on the subject, the potential benefits of using dynamic, adaptive models of argument include nothing less than creating a common language for management of humanity's most pressing problems.

3.2 Mapping arguments "on the hoof"

A second important effort at argument mapping constitutes a more holistic approach to mapping arguments though, unlike mapping Woods' "dressed" arguments, this effort focuses on the larger natural language context in which arguments are exchanged. As an example, consider the project undertaken by Horn et. al. (2003) to map not just individual arguments but the entire debate of a complex controversy. Their effort is impressive: mapping the manifold interactions between participants over time produces rich, intricate and holistic visual representations of debates that include decades of contributions from hundreds of arguers. Yoshimi (2004) differentiates their effort from argument mapping, explaining that

... whereas argument diagrams relate premises and conclusions *within* an argument (allowing one to distinguish divergent, convergent, linked and serial arguments, among others), debate maps relate whole arguments (allowing one to distinguish different forms of thread, debate, and position). Thus, every node on a debate-map can be represented by its own argument map, resulting in a graph of graphs. (p. 3)

While the scope of "debate mapping" certainly differs from most other argument mapping schemes, the fundamental approach is decidedly similar to other argument-mapping techniques. While I in no way intend to diminish the importance of their work or to understate their contributions to the methods of evaluating controversies at the systems level, it would appear that their "graph of graphs" represents a difference of scale rather than substance. The result of their efforts—often comprising several poster-sized displays for a single debate—is nothing short of comprehensive.

To advance, the field of argument mapping must continue down this road, turning its attention to argumentation as process and designing maps, models and tools that capture the dynamic, situated and interactive operations of arguments within argumentation, particularly as that dynamism, situation and interactivity relates to the space in which

argumentation occurs. Exploring the uncharted terrain on which arguments roam would be a step in the right direction.

4. A SPATIAL MODEL OF ARGUMENTATION

In his explanation of the metaphor distinguishing arguments “on the hoof” from those “dressed” to better understand their constituent parts, Woods (1995) offers an intriguing footnote. The butcher analogy, says Woods, is “... a gory metaphor. In the fateful passage from range to abattoir to table, beef on the hoof is ‘dressed’ before it is submitted to the final cut (reconstruction) in the meat department of Safeway.” Though Woods is primarily interested in the passage of arguments from their natural language form to the reconstructed, analytic representation of those arguments, the “range” on which those arguments roam is equally worth exploring. It is to this territory to which I’ll now turn my attention.

4.1 The territory of argument

To conceive of argumentation spatially demands examination of a couple of underlying assumptions. First, we conceive of arguments as possessing materiality. The perception of materiality is reflected in the previously-discussed metaphoric frames we use to describe arguments and corresponds with our inclination to imbue arguments with seemingly-material features: elements, dimensions, valence, boundaries, force, etc. That we conceive of arguments as possessed of materiality also invites consideration of what we know of the perception of material things. Gestalt psychology identifies “figure-ground differentiation” as a basic principle of perception. (Schacter, Gilbert, & Wegner, 2010) Arguments, as phenomena to which we impart the characteristics of material objects, are subject to the same perceptual principles: conceived of as material objects, we assume that arguments have both definition (as discrete objects, separate from other discrete objects) and exist in a context against which those discrete objects may be observed (a “space” in which arguments exist).

Following on from the first, the second assumption underpinning the spatiality of argumentation is that arguments’ existence in space permits them to interact. Arguments may be located, manipulated and oriented relative to the space in which they exist and relative to other arguments. This interaction is the product of energy applied through argumentation and influences the arguments themselves and space within which those arguments interact. Through the argumentative practices of construction, refutation and comparison, our conception of argumentation is predicated on the assumption that arguments’ operation affects other arguments and the context in which they exist. Representing this context and the interactions it contains in a graphic

model may lend insight into the function of arguments and the operations of argumentation.

These two assumptions indicate that a spatial model of argumentation should reflect the dimensions of *space* and *force*. *Space* primarily is concerned with the demarcations necessary to conceptualize arguments and the exchange of those arguments within argumentation; *force* acknowledges the energy applied to and within that space as manifest by the interaction between arguments.

4.2 Space & Argumentation

Argumentative space is initially demarcated by contest over and convergence on the substance and focus of the dispute. When an arguer advances an argument, they demarcate the space of argumentation either implicitly (by defining the terms of their argumentative efforts or making the substance and focus of the controversy known via the arguments they advance) or explicitly (by identifying or accepting the proposition offered for the controversy). The defined space of argumentation may be thought of as a field of play or, as noted by Goodwin (2002), as “territory” defined by the exchange of arguments; in either case, the space of argumentation may be modeled simply, as represented by Figure 3 below.



Figure 3 – Representing the space of controversy

Even at this rudimentary level, the model has utility. Boundaries around the controversy space define relevance: those subjects contested by the arguers (and all potential subjects they may contest in the course of the argumentation) are included within the bounds of the controversy; irrelevant subjects are outside the bounds of the controversy. In this way, the spatiality of argumentation functions much like a concept, grouping similar ideas together in coherent constructs. Unlike simple concepts,

however, the space of argumentation makes possible a particular cognitive functioning defined by and necessary for argumentation.

The concept of *audience acceptance* is fundamental to understanding the space of argumentation. Audience acceptance is a presumed goal of argument; to select information, identify relationships between different kinds of information, structure the presentation of that information and articulate the whole effort in a way that increases the chance that the audience will accept the novel or controversial is the essence of argumentation. Pursuit of audience acceptance defines the efforts of those advancing, critiquing and evaluating arguments in an exchange. *Audience acceptance*, as it constitutes the space within which argumentation occurs, may be defined only as it exists in tension with that which *has not* been accepted. Inch et. al. (2006) explain: “Because arguments occur only when people disagree with each other, we can imagine a line that separates the statements of belief and value with which an audience agrees from those with which it disagrees. In other words, for any issue we can think of, there is a line that separates what we are willing to accept from what we are unwilling to accept.” (pp. 46-47) This tension helps to elaborate on the distinction between simple conceptual space and that cognitive space defined by and dedicated to argumentation. The functional distinction between simple conceptual space and the unique attributes of argumentative space echoes Gärdenfors’ (2004) discussion of the function of nonmonotonic inference in conceptual spaces. Concepts inherently exhibit nonmonotonic inferential features, according to Gärdenfors, insofar as they are constantly challenged by new information in both propositional form (new facts that challenge the boundaries of existing concepts) and conceptual form (deductions of characteristics from inclusion of a new object into an existing category). (pp. 102, 126-131) In argumentative space, this nonmonotonicity is determinate: a space for argumentation doesn’t exist without tension between acceptance and rejection of the proposition that defines the space.

Building on the notion of the boundary between accepting and rejecting arguments, the model should reflect that the space allocated to each participant in a controversy is the product of the audience’s acceptance of an arguer’s position; the more willing the audience is to accept the arguer’s position, the more “territory” that position occupies. The representation of controversy in Figure 3 doesn’t capture the unique dynamism of argumentative space; to do so, more detail must be added to the model, as shown in Figure 4 below.

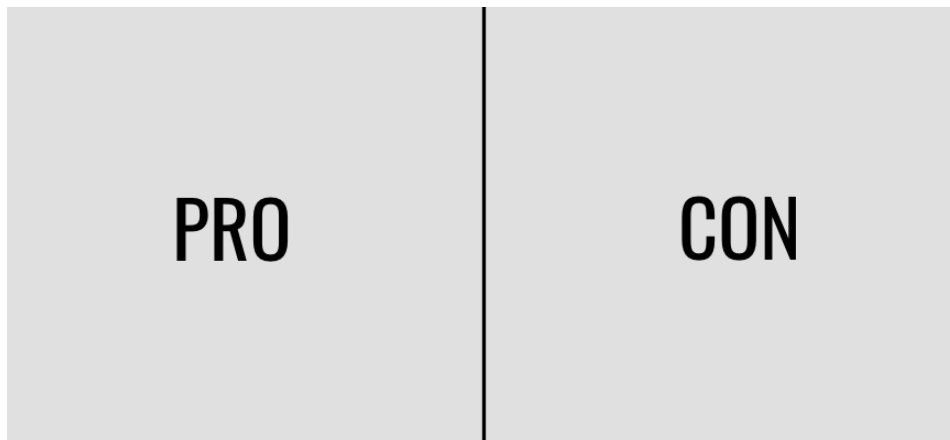


Figure 4 – Dividing the space of controversy into Pro and Con allocations

Here a line dividing Pro territory from Con territory has been added to suggest the space allocated to each side of the controversy. Like the larger space defined by the substance of the controversy, this division allocates territory to those arguments available to Pro and Con sides. The model gains complexity by recognizing that the territory is defined by the goal of the argumentation to be conducted within that space. The contest over the territory allocated to the Pro and Con is the essence of argumentation; that controversy exists both assumes the boundaries around the space within which arguments will be exchanged and the division of that space into (at least) Pro and Con subdivisions.

But the argumentative space itself may be further subdivided to represent areas of focus in argumentation. Generally referred to as *issues*, these subdivisions are similar in function to the general argumentative space defined by a proposition but different in scale as they divide and organize the more precise areas of argumentation within a broader controversy. Issues, according to Goodwin (2002), “... help us organize arguing at all its levels. Issues structure the exchange of individual premise/conclusion units (arguments in the narrowest sense), the fairly pointed, one-to-one exchange of multiple such units in a debate, and the wide-ranging, intermittent and many-to-many exchanges of a social controversy, not to mention fights and other less tightly organized argumentative interactions.” (p. 84). From this perspective, issues are the scaffolding of controversy, giving structure and focus to the exchange of arguments. Like the general territory of argumentation defined by a proposition, issues demarcate boundaries around subjects of sub-controversies within the larger territory of the general controversy. In the present model, issues divide the general territory into areas of more narrow focus, as Figure 5 illustrates.

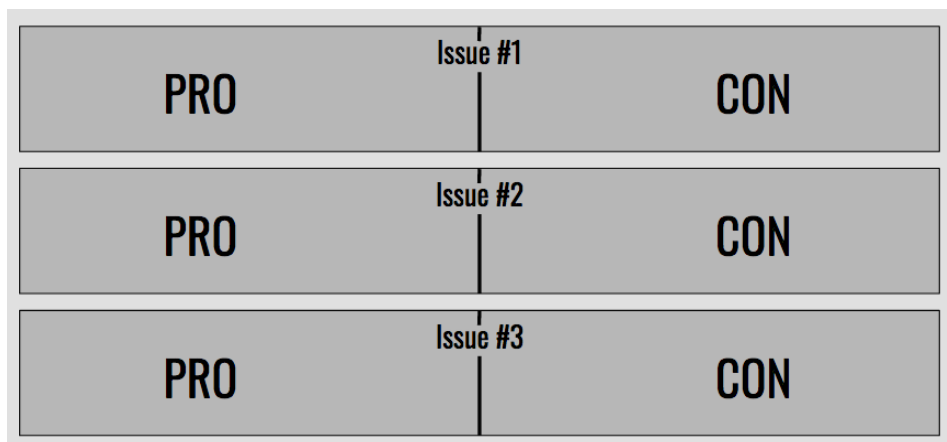


Figure 5 – Subdivision of the argumentative space into issues

Thus, the modeling of argumentation space reveals that space to be the product of the structural functions of reasoning. Insofar as we conceptualize and draw boundaries around the subject of controversy, which is in turn defined by the tension between competing sides in the controversy and organized by the issues considered in service to the larger controversy, the modeling of argumentative space provides insight into critical elements of argumentation.

The utility of the spatial model extends beyond the mere organization of arguments in a controversy. To understand how a spatial model of argumentation may further represent fundamental aspects of argumentation, we must now turn our attention to the second element of argumentation that must be illustrated by a model: *force*.

4.3 Force and Argumentation

Recall that the earlier discussion of efforts to model and map arguments revealed a widespread metaphoric frame employed to discuss arguments; namely, that *argumentation is movement*. Many models of argument already embrace this metaphor. By way of example, Toulmin's definition of argument made clear that motion is the essence of argument, leading an audience from data through warrant to claim. Foss, Foss and Trapp (1991) explain:

[Toulmin's] layout is based on an analog of motion: "an argument is *movement* from accepted *data*, through a *warrant*, to a *claim*." (Brockreide & Eningher, 1960) Making an argument is, therefore, analogous to taking a trip. One is trying to "get someplace" from "someplace else." (p. 100)

To be comprehensive, a model of argumentation must account for this movement; the present model will do so by illustrating the effect of *force* as it relates to the operation and exchange of arguments.

The next iteration of the model adds indications of force to represent the respective sides' goal in the exchange: each side endeavors to move the line dividing Pro from Con ground in an effort to occupy a greater amount of the audience's acceptance. The model incorporates arrows, illustrated in Figure 6, to represent this effort.

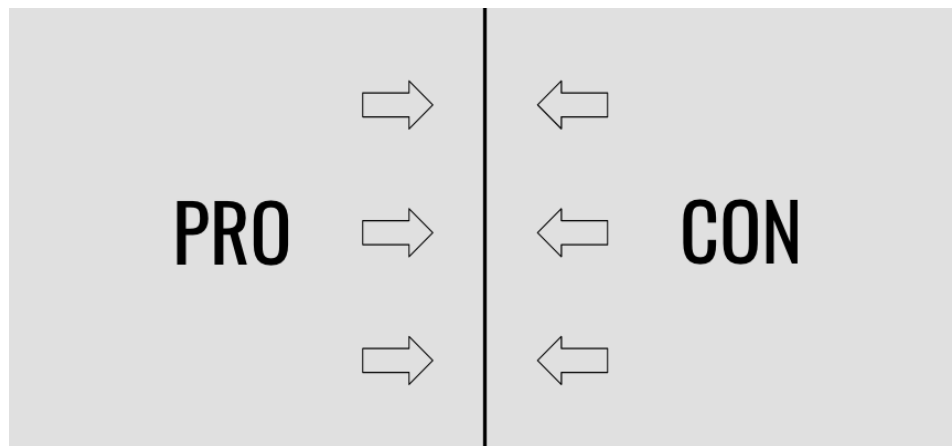


Figure 6 – Arrows illustrate the tension between the Pro and Con efforts

The tension created by these competing efforts represents the “clash” of positions advanced by the Pro and Con. This tension both defines the space (as discussed in the section above) and represents the potential energy inherent in argumentation: the effort to capture the acceptance of the audience defines argumentation and creates the space within which that argumentation plays out.

When applied to further structural refinements in the model, the representation of force also holds relevance. The issues that structure the sub-territory of argumentation within a controversy are subject to forces similar to those that act in the general space of argumentation. Figure 7 makes this clear.

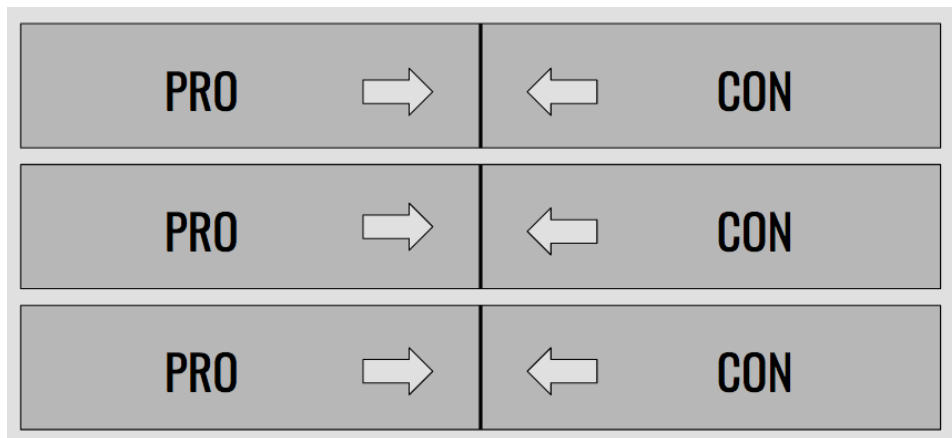


Figure 7 – Arrows indicate the goal of Pro and Con within the issue subdivisions

Here, the force of the Pro and Con efforts to capture territory within each of the issues is represented by arrows. Each side endeavors to establish arguments that will prove convincing to their audience while diminishing the force of their opponent's arguments; that effort acts upon the space of argumentation to allocate ground favorably to one side or the other. Figure 8 below reveals the measure of this effort; specifically, that the allocation of acceptance between Pro and Con's arguments is the result of the audience's *preference* for one side's arguments over the other.

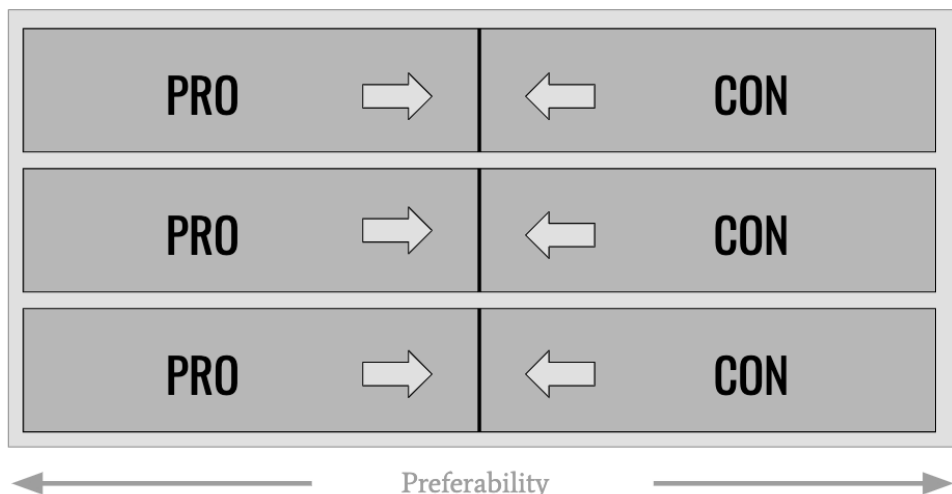


Figure 8 – Preference allocates territory within issues

The exchange of arguments within issues produces a preference on the part of the audience and, correspondingly, an allocation of territory based on that preference. We may refer to the goal of this effort *distribution*,

where gains in one side's territory correspond with losses to the other's, as is made clear in Figure 9.

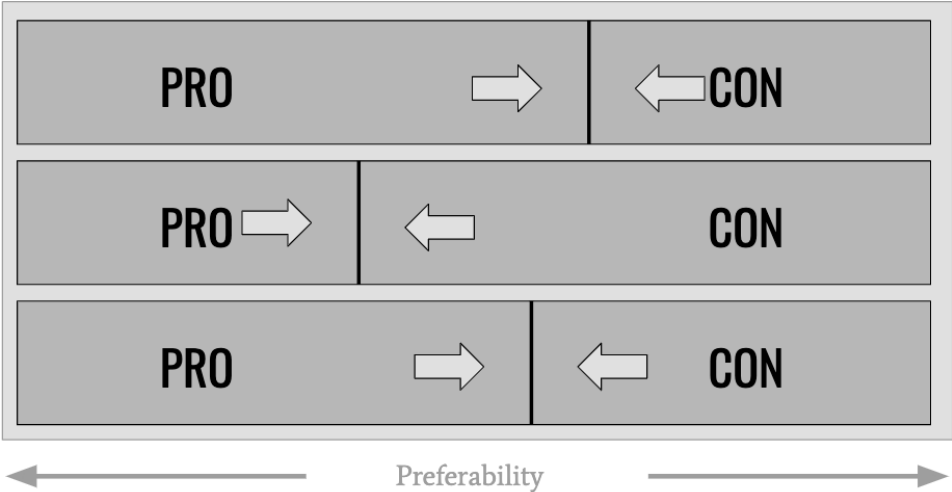


Figure 9 – Territory within issues is distributed by the audience's preference

Force acts upon the relationship between issues as well. Successful arguers attend not only to advancing arguments (within an appropriate issue to gain territory) and critiquing their opponent's arguments (in an effort to prevent them from doing the same), they also must be concerned with elevating the issues on which they're likely to prevail to a greater level of relative significance than other, competing issues. Goodwin (2002) observes that issues are not all created equally; some are more important than others: "It is very common to speak of issues on some sort of scale: that is, as more or less *big, important, major, prominent, significant, key, central, basic, fundamental, vital, pressing, necessary* and so on. ... [A]n issue is big (etc.) because the *fact of contention* matters more." (p. 85) The force of an arguer's effort to establish the *significance* of one issue over another may be modeled as illustrated in Figure 10.

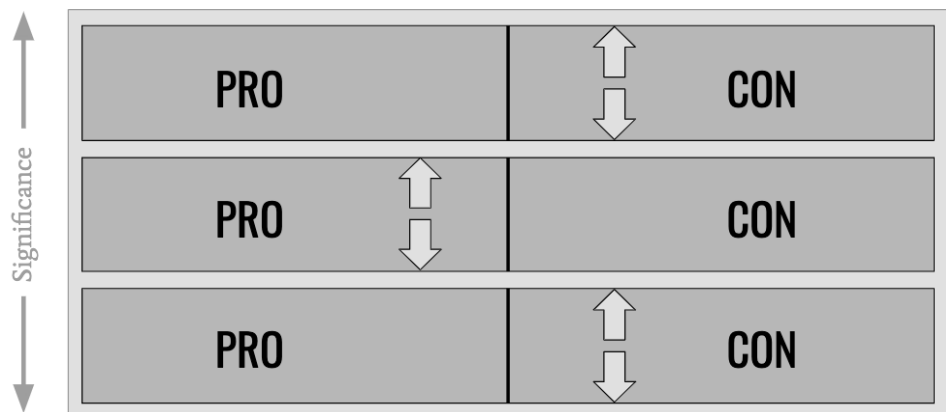


Figure 10 – Comparison of issues establishes the significance of each

The goal of this effort is *displacement* of other issues. If the amount of (audience acceptance) territory available to occupy is fixed by the boundaries of the controversy, each increase of an issue's size will correspond with a decrease in the size of other issues. When an arguer convinces an audience of the significance of one issue over others, the territory is allocated accordingly, as Figure 11 shows.

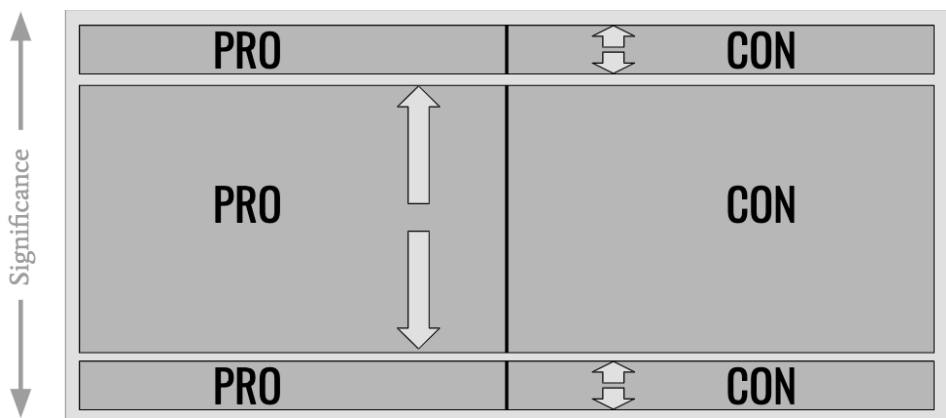


Figure 11 – Comparison displaces one issue in favor of another according to the audience's perception of significance

5. CONCLUSION

Through the integration of representations of both *space* and *force*, the utility of the model becomes more evident. Argumentation, as a communication phenomenon, may be extraordinarily complex. The uncertain nature of the subject matter, the partisan nature of the participants, the various strategies and tactics employed (or misemployed) by advocates and the potentially diverse perspectives of an audience create potential for a befuddling exchange. As such, the pedagogy of argumentation presents unique challenges.

Clear, functional models may help alleviate that confusion, illustrate basic principles of argumentation, structure practice in the exchange of arguments and provide a common frame of reference for discussions about those argumentative efforts. Consider, for example, an exchange of arguments between students for the proposition “Tobacco products should be banned.” Using the model, Figure 12 illustrates how the exchange may be scaffolded for novice arguers, providing guidance and predictability that will enhance their efforts to engage in a productive exchange of arguments.

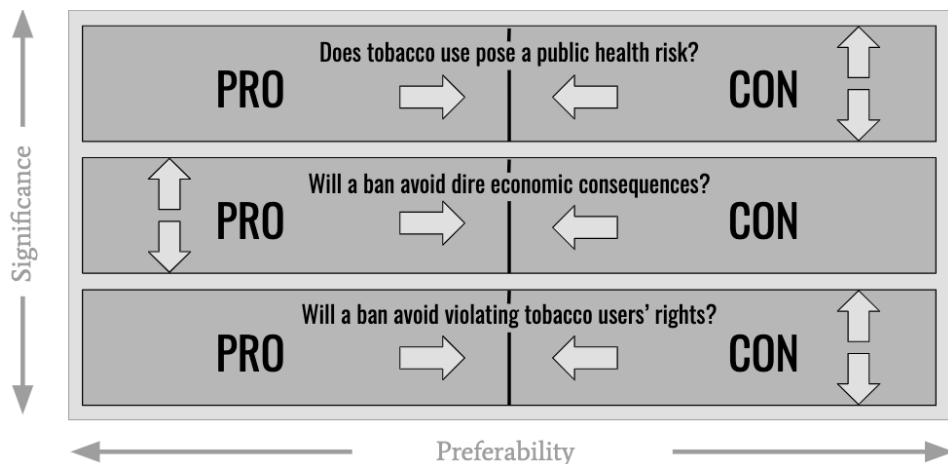


Figure 12 - Scaffolding an exchange of arguments for novice arguers

From the certainty provided by this structure for the exchange, the students can proceed to offer their arguments for and critique their opponents' arguments against the proposition utilizing a structure that maximizes the precision of those efforts. As the debate unfolds, the audience is able to follow more clearly the arguments exchanged, having already been provided with a “map” of the “territory” the debaters intend to cover. Further, this approach imagines that the audience could use the model to represent their individual assessment of the exchange,

allocating territory to reflect their perceptions of the preferability of positions articulated by each side as well as the significance of each issue debated. Figure 13 illustrates how the arguments made by Pro and Con in the tobacco ban debate may be rendered by an audience member to allocate their acceptance of the arguers’ efforts:

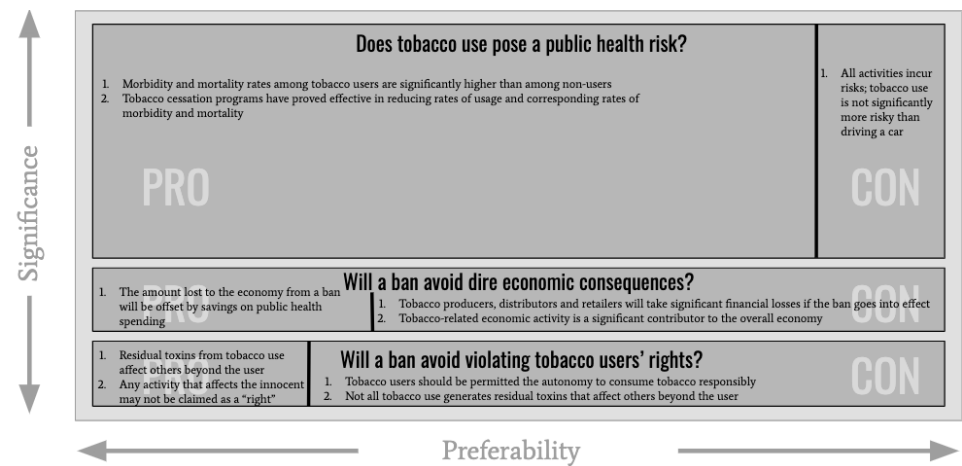


Figure 13 – The “territory” of the audience’s acceptance after the debate

This demonstrates only one of the ways that a model of argumentation may contribute argumentation pedagogy. Other considerations—such as defining the terms of a proposition, identifying and building strategy around anticipated issues, analyzing the audience’s presumption by considering their allocation of acceptance relative to the proposition and many other such theories and practices—may be made more accessible to students with such a model. In all, and as is the case with the majority of argumentation theory, such insights are the product of interaction between the insights that emerge as argumentation is more fully explored and the adjustments to practice and pedagogy that necessarily follow.

REFERENCES

Audrey, W. J. (2018). How to Promote Enlightened Debate Online. *The Chronicle of Higher Education*. Retrieved from <http://search.proquest.com/docview/2032480155?accountid=14473>

Beardsley, M. C. (1959). *Practical Logic*. Prentice-Hall.

Bellmund, J. L. S., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, 362(6415). <https://doi.org/10.1126/science.aat6766>

- Casasanto, D. (2009). Space for thinking. In *Language, cognition and space: State of the art and new directions* (pp. 453–478). Equinox Publishing.
- Chaudoin, S., Shapiro, J. N., & Tingley, D. (2014). Revolutionizing Teaching and Research with a Structured Debate Platform¹. *Journal of Political Science*, 58, 1064–1082.
- Davies, M. (2012). Computer-Aided Mapping and the Teaching of Critical Thinking: Part I. *Inquiry: Critical Thinking Across the Disciplines*, 27(2), 15–30.
- Dwyer, C. P., Hogan, M. J., & Stewart, I. (2012). An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning environments. *Metacognition and Learning*, 7(3). <https://doi.org/10.1007/s11409-012-9092-1>
- Foss, S. K., Foss, K. A., & Trapp, R. (1991). *Contemporary perspectives on rhetoric* (Vol. 12). Prospect Heights, IL: Waveland Press, Inc.
- Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. MIT Press.
- Gärdenfors, P. (2014). *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. MIT Press.
- Goodwin, J. (2000). Wigmore's Chart Method. *Informal Logic*, 20(3). <https://doi.org/10.22329/il.v20i3.2278>
- Goodwin, J. (2002). Designing Issues. In H. P. Van Eemeren F. H. (Ed.), *Dialectic and Rhetoric: The Warp and Woof of Argumentation Analysis* (pp. 81–96).
- Gould, S. J., & Vrba, E. S. (1982). Exaptation—a Missing Term in the Science of Form. *Paleobiology*, Vol. 8, pp. 4–15. <https://doi.org/10.1017/s0094837300004310>
- Groarke, L. (2019). Informal Logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019). Retrieved from <https://plato.stanford.edu/archives/sum2019/entries/logic-informal/>
- Horn, R. E. (2003). Infrastructure for Navigating Interdisciplinary Debates: Critical Decisions for Representing Argumentation. In P. A. Kirschner, S. J. Buckingham Shum, & C. S. Carr (Eds.), *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making* (pp. 165–184). London: Springer London.
- Inch, E. S., Warnick, B., & Endres, D. (2006). *Critical Thinking and Communication: The Use of Reason in Argument*. Pearson/Allyn & Bacon.
- Kirschner, P. A., Buckingham-Shum, S. J., & Carr, C. S. (2012). *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Springer Science & Business Media.
- Kuhn, D. (2009). *Education for Thinking*. Harvard University Press.
- Kuhn, D., Hemberger, L., & Khait, V. (2017). *Argue with me: Argument as a path to developing students' thinking and writing*. Routledge.
- Kunsch, D. W., Schnarr, K., & van Tyle, R. (2014). The Use of Argument Mapping to Enhance Critical Thinking Skills in Business Education. *Journal of Education for Business*, 89(8), 403–410.
- Millard, W., & Menzies, L. (2015). *The State of Speaking in Our Schools*. Retrieved from Voice 21 website: <https://www.lkmco.org/wp-content/uploads/2016/11/Oracy-Report-Final.pdf>
- Pinker, S. (2009). *How the Mind Works*. W. W. Norton & Company.

- Rajagopalan, A. (2019, February 8). New Evidence for the Strange Geometry of Thought. *Nautilus*. Retrieved from <http://nautil.us/blog/new-evidence-for-the-geometry-of-thought>
- Rider, Y., & Thomason, N. (2014). Cognitive and pedagogical benefits of argument mapping: LAMP guides the way to better thinking. In *Knowledge cartography* (pp. 113–134). Springer.
- Schacter, D. L., Gilbert, D. T., & Wegner, D. M. (2010). *Psychology*. W. H. Freeman.
- Shum, S. B. (2003). The Roots of Computer Supported Argument Visualization. In Paul A. Kirschner, Simon J. Buckingham Shum, Chad S. Carr (Ed.), *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making* (pp. 3–24).
- Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, 12(1), 49–100.
- Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge University Press.
- van Gelder, T. (2003). Enhancing Deliberation Through Computer Supported Argument Visualization. In P. A. Kirschner, S. J. Buckingham Shum, & C. S. Carr (Eds.), *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making* (pp. 97–115). London: Springer London.
- van Gelder, T. (2005). Teaching Critical Thinking: Some Lessons From Cognitive Science. *College Teaching*, 53(1), 41–48.
- van Gelder, T. (2015). Using Argument Mapping to Improve Critical Thinking Skills. In M. Davies & R. Barnett (Eds.), *The Palgrave Handbook of Critical Thinking in Higher Education* (pp. 183–192). New York: Palgrave Macmillan US.
- van Gelder, T., Bissett, M., & Cumming, G. (2004). Cultivating expertise in informal reasoning. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 58(2), 142–152.
- Venn, J. (1880). I. On the diagrammatic and mechanical representation of propositions and reasonings. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(59), 1–18.
- Whately, R. (1897). *Elements of Logic*. Longman, Green, Longman, Roberts and Green.
- Wigmore, J. H. (1913). *The Principles of Judicial Proof: As Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials*. Little, Brown,.
- Woods, J. (1995). Fearful symmetry. In Hans V. Hansen and Robert C. Pinto (Ed.), *Fallacies: Classical and contemporary readings* (pp. 181–193). University Park, PA: Pennsylvania State University Press.
- Yoshimi, J. (2004). Mapping the Structure of Debate. *Informal Logic*, 24(1). <https://doi.org/10.22329/il.v24i1.2130>
- zzyzx. (2018, March 12). Should voters in the UK have a final vote on the Brexit deal?. Retrieved May 22, 2019, from Kialo website: <https://www.kialo.com/if-there-was-a-second-referendum-and-britain-voted-this-time-to-remain-it-would-not-dissolve-uk-euro-scepticism-or-solve-the-10250.1338?path=10250.0~10250.1-10250.1338>

Aggressiveness as a qualifier of dissent in interpersonal arguing

KHOMENKO IRYNA

Taras Shevchenko National University of Kyiv, UA

khomenkoi.ukr1@gmail.com

The purpose is to examine aggressiveness as a qualifier of dissent in interpersonal arguing. This research is based on the results of the cross-cultural project, which empirical part included a nationwide survey conducted in Ukraine in 2018. Aggressiveness was the characteristic, which was explored in this project. The new experiment was conducted within the previous project in 2019. In the paper, I analyse its findings related to the general attitudes of Ukrainians to non-verbal aggression in interpersonal communication.

KEYWORDS: aggressiveness, argumentation, empiricalization, empirical methodology, interpersonal arguing, Ukrainian identity

1. INTRODUCTION

In these times argumentation studies involves theoretical as well as empirical approaches. However, if we look back in the 20th century we can see investigations in argumentation sphere were mainly aimed at theoretical proposals and their philosophical support. Argumentation scholars mostly focused on the development of fundamental principles, ideas, and theoretical models. However, it should be noted that Toulmin (Toulmin, 1958) and Perelman considered some aspects of empirical study of argumentation

In 1958, Toulmin, the founder of the working logic, in his epoch-making book *The Uses of Argument* proposed a procedural model of argumentation [Toulmin, 1958]. He emphasized that:

a radical re-ordering of logical theory is needed in order to bring it more nearly into the line with critical practice [Toulmin, 1958: 253; 2003: 234].

It means that his model could work adequately in the different areas of argumentative reality. In this regard, Toulmin assumed that not only a theoretical component is relevant to the study of argumentation but also an empirical one is needed. Justifying this view, he claimed:

logic ...may have to become less of an a priori subject than it has recently been... Accepting the need to begin by collecting for study the actual forms of argument current in any field, our starting point will be confessedly empirical" [Toulmin, 1958: 257; 2003: 236-238].

Besides, it should be stressed that Toulmin connected the empirical component with the historical one. He believes:

not only will logic have to become more empirical; it will inevitably tend to be more historical... In the natural science, for instance, men such as Kepler, Newton, Lavoisier, Darwin and Freud have transformed not only our beliefs but also our ways of arguing and our standards of relevance and proof... Grotius and Bentham, Euclid and Gauss, have performed the same double feat for us in other fields" [Toulmin, 1958: 257; 2003: 237].

In fact, here Toulmin bearded in mind that empirical database for scholars may be the history of thought in general and the history of science in particular. Perelman and Olbrechts-Tyteca, who are among the co-founder of the modern theory of argumentation, supported Toulmin's view. They claimed that the theoretical concepts of their treatment, called new rhetoric, had to base on the empirical observation [Perelman & Olbrechts-Tyteca, 1958]. Unfortunately, the empirical dimension in their treatment was not developed thoroughly.

Only more recently empiricalization has become a new trend among researchers. For understanding the first steps of empiricalization it would be appropriate to give a perfect analogy:

Like Frege's theory of logic was founded upon a descriptive analysis of mathematical reasoning, they founded their argumentation theory on a descriptive analysis of reasoning with value judgments in the fields of law, history, philosophy, and literature" [Eemeren, 2015: 5-6].

Eemeren considers empiricalization as one of the main prospects of current argumentation study (see: Eemeren, 2015; 2017). In his opinion:

Three major developments in the treatment of argumentation have begun to materialize that open new avenues for research.

Although they differ in shape, these developments can be observed across a broad spectrum of theoretical approaches. The three developments I have in mind can be designated as empiricalization, contextualization, and formalization of the treatment of argumentation (Eemeren, 2015, p. 5).

Let us deal with what is empiricalization of the treatment of argumentation. It should be noted that a key feature of empirical research is to use empirical evidence as a way of gaining results. The term *empirical evidence* refers to the systematic collection and analysis of data related to the field of argumentation. Scholars may involve a variety of instruments (Khomenko, 2018). For example, in empirical research of various historical texts, Finocchiaro used the method of alternative conclusion, active evaluation, ad hominem argument, the method of counterexample, the principle of charity, and explanation of the error in reasoning. In his mind, the first three methods are the most relevant for his historical-textual approach of argumentation (Finocchiaro, 1980; 1994; 2005; 2010). Besides, Finocchiaro elaborated the method of meta-argumentation based on the principle of interpretation and evaluation (Finocchiaro, 2013).

Also, scholars often use a survey as an instrument for collecting data in such research (Hample, 2003; 2005; 2018). It bases on the questionnaire, which is a research instrument consisting of a series of questions for gathering information from respondents. Data can be collected relatively quickly, using a standard set of questions. A survey is useful for a large number of participants.

Empirical evidence can be analyzed quantitatively as well as qualitatively. Many argumentation scholars combine both forms of analysis to better answer questions which cannot be studied in artificial laboratory settings.

These days empirical methodology is very popular in the field of cross-cultural research, particularly it can be reasonably applied to the goal of understanding how people with different cultural traditions relate to face to face arguing. Among the appropriate approaches in this field is the one selected here. Its orientation is to examine predispositions and understandings regarding interpersonal arguing. Its main objective is to collect and analyze the data summarizing fundamental orientations to arguing among respondents. All of the instruments of this research originated in the US. It should be stressed that besides elucidating the argumentation predispositions in a certain country this approach advances the general project of comparing argumentation in various countries across the globe (Hample, 2018). To date, the United States, Chile, China, France, India, Malaysia, Mexico, Netherlands, Portugal, South Korea, Turkey, United Arab Emirates and other countries have

already participated in this global project (see, Hample & Anagondahalli, 2015; Hample & Rapanta, 2015; Lewiński et al., 2018).

In 2018 Ukraine joined this international project. It was the first comprehensive attempt to understand and assess the sentiments of Ukrainians towards face-to-face arguing (Khomenko & Hample, 2019). The main objective of the cross-cultural project was to collect and analyse data summarizing fundamental orientations to arguing among Ukrainians. After its completion, we believed it essential to continue working towards improving our knowledge in this research field. At the beginning of 2018 such frame as aggressiveness was selected as a subject matter of the new project.

Thus, the goal of this paper is to present received under the empirical research key findings, discoveries, and generalization concerning Ukrainian predispositions and understandings regarding aggressiveness in face to face arguing.

Empirical research (2019) was conducted to seek the views of young Ukrainians on aggressiveness in verbal and non-verbal argumentation. Here it is presented its key findings concerning non-verbal argumentation.

2. TOWARDS THE HISTORY OF THE PROJECT

2.1 Prehistory. The Ukraine-US cross-cultural project (2018)

In 2018 a nationwide survey was conducted in Ukraine within the international cross-cultural project. One of the main research questions was the following: how does Ukrainian and US respondents compare in their average responses concerning arguing motivations, understandings, and reactions? In this regard, we chose to pose such research questions.

RQ1: How do Ukrainian and US respondents compare in their average responses to items concerning arguing motivations, understandings, and reactions?

RQ2: Do Ukrainian men and women differ in their arguing motivations, understandings, and reactions?

RQ3: Do Ukrainian respondents who chose the Ukrainian language version of the survey differ in their arguing motivations, understandings, and reactions, compared to those who chose the Russian language version?

RQ4: What are the internal associations among arguing motivations, understandings, and reactions for the Ukrainian sample?

After data collection and its analysis, we concluded that Ukrainians do not like to be involved in face to face arguing because it is considered to be a hostile incursion into their personal world or conflict

and aggressive activity that tends to destroy their life and aggravate the interpersonal relations. If they are obliged to participate in such form of communication they mainly will not behave aggressively, offend people, incite hatred and provoke violence. The majority of Ukrainians are predominantly tolerant, peaceful, friendly, and open to others. In most cases, they try to be polite, to express respect for others and to argue constructively.

At the same time, understanding of face to face arguing solely as a destructive conflict tending to damage relationships may be considered as the reason why Ukrainians are less inclined to understand arguments as professionals do.

2.2 The project "Aggressiveness in interpersonal arguing" (2019)¹

At the beginning of 2019, a new research project concerning Ukrainian youth attitudes to face-to-face arguing was opened. Its subject matter was such frame of arguing as aggressiveness. Some ideas of argumentation scholars a sufficiently strong basis of this project (see, Hample, 2003, 2005, 2018; Hample & Irions, 2015; Hample & Cionea, 2010; Hample, Han & Payne, 2010; Hample & Dallinger, 1995; Infante & Wigley; Infante & Rancer, 1982). Taking into account that verbal aggressiveness has been already studied within the previous project we primarily focused on non-verbal argumentation.

Thus, the main objective of this project was to explore the general attitudes of Ukrainian youth to aggressiveness in face to face arguing. In this regard, we chose to pose such research questions.

RQ1. What are Ukrainian youth attitudes towards aggressiveness within interpersonal communication?

RQ2. Do the results of the present project (2019) confirm the findings on aggressiveness received within the survey (2018)?

RQ3. What are the reasons for increasing the level of aggressiveness in face to face arguing?

3. SAMPLE

The empirical part of the project included the survey and experiment conducted in March 2019. 14 respondents were interviewed. Among them were 11 women and 3 men. All respondents (100%) were enrolled in Taras Shevchenko National University of Kyiv (Ukraine) at the time they completed the survey and participated in the experiment. They were

¹ The empirical part of the project was conducted with the assistance of Kateryna Bura (Master student, Philosophy Faculty, Taras Shevchenko National University of Kyiv, Ukraine)

undergraduates and PhD students from the Philosophy Faculty and the History Faculty. 3 undergraduates and 11 PhD students were involved in the poll.

Ukraine has the following regional composition: West, East, Centre, North, and South regions. The respondents presented all regions and the city of Kyiv, the capital of Ukraine, except the Autonomous Republic of Crimea and uncontrolled territories of Donetsk and Luhansk regions.

The survey and experiment were carried out at Taras Shevchenko National University of Kyiv. The statistical error² of the survey does not exceed 0.3%.

4. RESULTS

An absolute majority of the respondents (71.5%) reported they did not like the performance (in one degree or another). In contrast, only a quarter of the interviewees (28.5%) enjoyed it. To the question what do you think is the video about the following answers were prevalent in the survey: (1) face-to-face arguing between very close people; (2) a splash of aggression and anger between arguers; (3) inability to control own emotions within an aggressive arguing.

While the respondents watching the performance, they wanted to get away from the scream of actors, to stop the torrent of hatred and aggression. Describing their feelings, the respondents used such words as offensively, obnoxiously, anxiously, pain, anger, disquiet, annoyance, nervousness, etc. Nevertheless, some of the respondents asserted that the video generated their interest and it was surprising.

Almost all respondents agreed that aggression occurred in the performance. It was expressed in screaming, violation of personal space, an attempt to win the partner emotionally. Also, the respondents drew attention to negative (aggressive) changes in voice tone, facial expressions and body movements. In general, the video was described by the following adjectives: aggressive, animal, annoying, wild, emotional, wrathful, strained, stressful, and gruelling. However, there were such characteristics as interesting, genuine, edifying, profound, and useful.

Let us move on to the issue of how music tracks affected the respondent's experience of the performance. The analysis of the survey data has shown that 75% (3 members) of the respondents from group1,

² The statistical error was calculated by the formula that corresponds to the "standard deviation" rule, according to which the statistical error can be estimated at the 95% confidence interval, so the error will depend only on the sample size: $\varepsilon = 1/\sqrt{n}$, where ε is the statistical error, n is the sample size.

listened to Chopin's track, got more aggressive experience, concerning with the performance.

It should be noted that everyone listened to Chopin (5 people) know "Nocturne in E-flat major, Op. 9, No. 2" earlier and liked it. The participants indicated that the music track put them in a resting state, helped to relax and feel good. However, despite this recognition, the track was a reason strongly negative attitude to the video they just have seen. Thus, it is possible to conclude that a positive attitude to the track, as well as the fact that the respondents were previously familiar with it, became that context increased the overall negative assessment of the performance. The state of calm and pleasure changed completely in the other direction during watching of the video. All these led to increased aggression and anxiety of the respondents and their negative evaluation of the performance in general.

The second group listened to Symphony No. 2 in D minor (Op. 40) by Prokofiev. Four out of five participants did not know this track. Only one person liked it, two members of this group had a neutral attitude, and yet two did not like it at all. On the one hand, two people asserted that the musical track influenced their ways to perceive the performance. On the other hand, three members of this group believed that music did not matter. At the same time, they described their feelings after listening like fear, anxiety, panic, displeasure, irritation, and anger. In this regard, it can possible to conclude that Prokofiev's track influenced on the respondent's mental well-being while watching the performance. Their aggressiveness was increasing.

The third group listened to the sounds of nature (White noise) – neutral sounds from everyday life. All participants indicated that audio and video were emotionally separated; music had no impact on their mental well-being while watching the performance. Here in contrast to the previous groups, the film was viewed without a specific musical setting.

5. DISCUSSION

Before the experiment, the respondents engaged in the discussion concerning their arguing motivations, understandings, and reactions. First up, they considered various ways of determining aggression in face-to-face arguing. The analysis of the respondent's views has shown that the absolute majority of interviewees were sure that aggressiveness primarily reflects the emotional state of the person.

Describing aggressive behaviour in interpersonal communication, the students connected it with (1) negative emotional impact (threats, provocations) on arguers; (2) destabilization of opponent's mental state (psychological suppression); (2) numerous non-

verbal factors concerning incredibly inappropriate body language of arguers (3) loss of self-control, provoked by outside factors, (4) verbal rejection of the opponent's point of view. Based on the definitions expressed during the discussion, we can conclude that Ukrainian youth links aggressiveness with an unusually powerful emotional influence on face-to-face arguing in general and arguers specifically. The respondents pointed out that rational approach to aggressiveness in interpersonal communication is almost impossible.

Further, the students discussed key factors affected on the level of aggression in face to face arguing and its manifestation in non-verbal argumentation. The interviewees identified that non-verbal manifestations of aggression are, on the one hand, overreacting, rude, and lewd gestures, and, on the other hand, a total absence of gestures. Also, it can be a raised voice tone; strong facial expressions, violation of personal space, contemptuous and deceptive behaviour.

Additionally, the respondents discussed several outside factors. They do not link with face to face arguing directly but can affect the level of aggression of the participants. These factors could be divided into subjective and objective. The first group includes such traits as bad physical and mental health of arguers (physical exhaustion, psychological stress, personal problems leading to powerful negative emotion, etc.). The second group involves visual points (irritating visuals, too bright light, etc.) and audio factors (annoying sounds, too loud music, very strange noise, etc.). Also, the respondents incorporated to outside factors too close physical contact, the previous negative experience of communication with the opponent, intervention of third parties, bad weather conditions (uncomfortable temperature, atmospheric pressure, etc.), volatile political situation, the use of obscene language, etc.

Summarizing, all respondents concluded that aggressiveness in face to face arguing cause hatred, anger and rejection of the opponent's view on the problem under discussion. They believed that aggressiveness in interpersonal communication is a hostile incursion into their personal world, leading to conflict and serious dissents between arguers. Such aggressive activity tends to destroy their life and aggravate the interpersonal relations.

In case participation in face-to-face arguing the students asserted that their behaviour mostly is not aggressive. They do not offend people, incite hatred or anger, and provoke violence. The respondents identified themselves as peaceful and friendly. They try usually to maintain stability and avoid conflicts.

6. CONCLUSION

In this paper, we consider the key findings of an extensive empirical study concerning varied aspects of aggressiveness in face to face arguing. It has focused on the survey and experiment conducted in March 2019. The participants were undergraduates and PhD students studying philosophy, political science, and history at Taras Shevchenko National University of Kyiv.

Focusing on young people aged 18 to 25, which present all regions of Ukraine except the Autonomous Republic of Crimea and uncontrolled territories of Donetsk and Luhansk regions the survey provides a mini-snapshot of Ukrainian youth, born between the mid- 90s and mid-2000s.

The first research question of our project was the following, what are Ukrainian youth attitudes towards aggressiveness within interpersonal communication? The analysis of received data has shown that Ukrainian youth has a strongly negative attitude to disputes in which aggressive emotions predominate. Aggression is considered as something unacceptable for the human person. According to the research, aggressive behaviour in face-to-face arguing often causes emotional pain, deep stress and other negative impacts on the mental health of arguers. All of these points lead to quarrel instead of conflict resolution. Therefore, Ukrainian youth identify aggressiveness as a reason for improving disagreement in face to face arguing. According to the majority of responses, Ukrainian students are predominantly tolerant, peaceful, and friendly. They prefer to express respect and sympathy for others and to argue constructively or do not enter the arguing at all.

As can be seen, there are large areas of overlap between the data of two surveys, conducted in 2018 and 2019.

Further, we were interested in the issue: what are the reasons for increasing the level of aggressiveness in face to face arguing? For identifying some of them, we experimented with such argumentative tools as different music tracks and the video performance AAA-AAA by Marina Abramovic & Ulay. While this experiment the last was seen as a model of face to face arguing. This research made it possible to identify the direct and indirect musical impact on the aggressiveness levels.

In the first case listening to aggressive music (Prokofiev's Symphony No. 2 in D minor. Op. 40) contributed to the strengthening of aggressiveness while watching video performance AAA-AAA by Marina Abramovic & Ulay. This result was expected for us.

The second group of the respondents listened to calm and pleasant music (Chopin's Nocturne in E-flat major, Op. 9, No. 2). There we have received the results impressed us. Calm and relaxing music increase

of aggressiveness within watching the performance. The respondents highlighted that it was the effect of dramatic contrast.

Above we present the main results of our research, discoveries and generalizations that have attracted our attention. However, we understand that it is only the tip of the iceberg and further detailed analysis of Ukrainian youth attitudes towards aggressiveness in interpersonal arguing should occur in future.

REFERENCES

- Eemeren, F. H. van. (2015). Bingo! Promising developments in argumentation theory. In *Reflections on theoretical issues in argumentation theory* (pp. 3-25). Heidelberg etc.: Springer.
- Eemeren, F. H. van. (2017). Argumentation theory and argumentative practices: A Vital but Complex Relationship. *Informal Logic*, 37 (4), 322-350.
- Finocchiaro, M. A. (1980). *Galileo and the art of reasoning: rhetorical foundations of logic and scientific method*. Boston, MA: Reidel.
- Finocchiaro, M. A. (1994). Two empirical approaches to the study of reasoning. *Informal Logic*, 16, 1-21.
- Finocchiaro, M. A. (2005). *Arguments about arguments. Systematic, critical and historical essays in logical theory*. Cambridge: Cambridge University Press.
- Finocchiaro, M. A. (2010). *Defending Copernicus and Galileo. Critical reasoning in the two affairs*. Dordrecht: Springer.
- Finocchiaro, M. A. (2013). *Meta-argumentation. An approach to logic an argumentation theory*. London: Colledge Publications.
- Hample, D. (2003). In *Arguing skill. Handbook of communication and social interaction skills* (pp. 439-478). Mahwah, NJ: Lawrence Erlbaum.
- Hample, D. (2005). *Arguing: exchanging reasons face to face*. Mahwah, NJ: Lawrence Erlbaum.
- Hample, D. (2018). *Interpersonal arguing*, New York: Peter Lang.
- Hample, D., & Anagondahalli, D. (2015). Understandings of arguing in India and the United States: Argument frames, personalization of conflict, argumentativeness, and verbal aggressiveness. *Journal of Intercultural Communication Research*. 44 (1), 1-26.
- Hample, D., & Cionea, I. A. (2010). Taking conflict personally and its connections with Aggressiveness. In *Arguments, aggression, and conflict: New directions in theory and research* (pp. 372-387). New York, NY: Routledge, Taylor, and Francis.
- Hample, D., & Dallinger, J. M. A. (1995). Lewinian perspective on taking conflict personally: Revision, refinement, and validation of the instrument. *Communication Quarterly*, 43 (3), 297-319.
- Hample, D., & Han, B., & Payne, D. (2010). The aggressiveness of playful arguments. *Argumentation*, 24 (4), 405-421.
- Hample, D., & Irions, A. L. (2015). Arguing to display identity. *Argumentation*, 29 (4), 389-416.
- Hample D., & Rapanta, Ch. (2015). Orientations to interpersonal arguing

- in the United Arab Emirates, with comparisons to the United States, China, and India. *Journal of Intercultural Communication Research*, 44(4), 263-287.
- Infante, D. A., & Rancer, A. S. (1982). A conceptualization and measure of argumentativeness. *Journal of Personality Assessment*, 46, 72-80.
- Infante, D. A., & Wigley Ch. J. Verbal aggressiveness: an interpersonal model and measure. *Communication Monographs*, 53, 61-69.
- Khomenko, I. (2018). Empiricalization as a Trend in Argumentation Study. *Future Human Image*. 10, 20-28.
- Khomenko, I. & Hample, D. (2019). Comparative analysis of Arguing in Ukraine and the USA (pp. 634 – 645). In Garssen, B, Godden, D., Mitchell, G.R.,
- Wagemans, J.H.M. (Eds.), *Proceedings of the Ninth Conference of the International Society for the Study of Argumentation*. Amsterdam: Sic Sat.
- Lewiński, M., & Hample D., & Sàágua J., & Mohammed D. (2018). Arguing in Portugal: Across-cultural analysis. *Journal of International and Intercultural Communication*, 233-253.
- Perelman, Ch., & Olbrechts-Tyteca, L. (1958). *Traité de l'argumentation: la nouvelle rhétorique*. Paris: Presses Universitaires de France La Nouvelle Rhétorique.
- Toulmin, St. E. (1958.) *The uses of argument*. Cambridge: Cambridge University Press, 1958.
- Toulmin, St. E. (2003). *The uses of argument*. Updated edn. Cambridge: Cambridge University Press, 2003.

Auditory arguments – importance of sound in an argumentative discourse (An empirical study)

GABRIJELA KIŠIČEK
University of Zagreb
gkisicek@ffzg.hr

Successfully navigating the social world requires that people accurately use nonverbal information to guide their behaviour.
(Bjornsdottir, Alaei & Rule, 2017)

1. PROSODIC FEATURES IN PERSUASION PROCES

Nonverbal communication is an important part of the persuasion process and extensive research in social psychology has examined how the nonverbal behaviour of a communicator can affect recipients' attitudes and attitude change via their effects on a recipient's perceptions of source credibility, attractiveness, or power (e.g., Aguinis et al 1998; Burgoon, Birk & Pfau, 1990). According to the Guyer et. al (2019) it has been proven that when a person is either unable and/or unmotivated to think carefully (i.e., low elaboration conditions), incidental emotions aroused by nonverbal features of the message source are either misattributed to one's attitude (i.e., positive feelings reflect a positive attitude), to the message (i.e., feeling good signals agreement with the message), or to the attitude object (i.e., the object makes me feel good, so I must like it). Although these accounts differ in their explanations, each agree that when attitudes change under low-thinking, the direction of persuasion is consistent with the direction implied by the nonverbal feature of the source.

One aspect of nonverbal behaviour that this paper is interested in is prosody or paralanguage or vocal cues (i.e. voice quality, tempo, loudness, fluency, pitch and pitch range etc.). Numerous empirical studies confirmed that voices prompt spontaneous evaluations related to attractiveness, and to character traits such as trustworthiness and dominance (Willis & Todorov 2006; Vukovic et al. 2011). These evaluations are highly consistent across observers (Oosterhof & Todorov 2008). Rezlescu et al. (2015: 367) and confirmed that perception of

trustworthiness and the credibility of the source (speaker) is highly influenced by vocal cues: "Voices, just like faces, can lead to formation of consistent trait impressions of trustworthiness, attractiveness, and dominance".

For all those involved in communication, public speaking, rhetoric but also in argumentation, it is important to take these insights into account. For instance, vocal cues are source of many stereotypes which is highly used in advertising industry: certain voice types are used to advertise certain products (examples in Kišiček 2014; 2016; Groarke & Kišiček, 2016 etc.) but it is also important for certain professions (e.g. telephone salespersons, radio or television news journalists) and certain professional situations (e.g. job interviews). Pittam et al. (1987, 1989) investigated social perceptions among Australians whose speech is thought to contain a high degree of nasality. Highly nasal voices were rated as being lower in "status" (occupation, ambitious, intelligent, educated, influential), lower in social solidarity (friendly, sympathetic, likeable, trustworthy, helpful), and were negatively correlated with perceptions of persuasiveness. Bloom & Zajac (1999: 279) claim that nasality as highly negative voice quality is often intentionally decreased to change perception of the speaker:

For example, in training for job interviews, applicants can adjust nasality of voice to possibly increase perceptions of competence, warmth, and persuasiveness, and to possibly decrease perceptions of arrogance or weakness.

Public speaking trainers and rhetoric scholars are also very aware that vocal cues may affect speaker's personality traits as well as her emotional state. Therefore, it has to be taken into account in the realm of politics (practice or pedagogy) and even in judicial discourse especially when vocal cues can signal deception e.g. vocal tension, pauses, speech errors (Davis et al., 2006; DePaulo et al., 2003; DePaulo, Stone, & Lassiter, 1985; Mann et al., 2004; Vrij et al., 2000). It can be concluded that these findings are useful in all situations in which credibility and trustworthiness of a speaker are crucial.

Invoking a specific emotional response from the audience is also essential in persuasion process. According to Aristotle *pathos* is considered as mean of persuasion in which speaker is creating a certain disposition in the audience. However, it can be developed not just by verbal means but also by nonverbal means of communication (especially prosodic features). Expressing emotions (by speaker) and recognizing speaker's emotions (by audience) via prosodic features is universal across cultures and will certainly have role in "putting audience in the right frame of mind". Nonverbal communication scholars claim that

human expressive behaviours which communicate joy, anger, disgust, sadness, and fear are thought to possess certain invariant properties which allow them to be recognized independent of culture and learning (Ekman et al 1980). Recent reviews (Juslin & Laukka 2003; Laukka 2008) have shown that vocal expressions of these emotions (e.g., anger, fear, happiness, sadness) are generally recognized with accuracy above chance, also cross-culturally, and are associated with relatively distinct acoustic characteristics. Having these findings in mind we can imagine that verbally expressed emotional appeals can be accompanied with corresponding vocally expressed emotions of the speaker and this can contribute to developing *pathos*. When a speaker intends to create empathy, anger or fear in the audience, prosodic features will have important impact. And it can even become an essential part of an argumentative discourse (Groarke & Kišiček, 2018).

2. AUDITORY ARGUMENTS – PROSODIC FEATURES AS PART OF A *LOGOS*

Recent research in the realm of argumentation theory introduced a concept of auditory arguments (Groarke & Kišiček, 2016; Groarke & Kišiček, 2018; Groarke, 2018; Kišiček, 2019) which Groarke (2018) defines as “an attempt to provide rational evidence for a conclusion using non-verbal sounds instead of or (more frequently) in addition to words”.

It means that sounds both human (prosodic features) and non-human (e.g. different sound alarms) may serve as part of a *logos* and contribute in (re)constructing the argument. Kišiček (2019) provided several examples of different argument schemes where sound is one of the premises (argument from sign, argument from consequences, argument from correlation to cause...). The specific scheme this paper is investigating is the argument from sign in so called testimonial claims.

As Govier (1993, p. 93) explains:

Testimonial claims are especially important for a variety of reasons. Human knowledge is utterly dependent upon our acceptance, much of the time, of what other people tell us. Only thus can we learn language and pass on knowledge from generation to generation; only thus have we access to times, places, and cultures we do not and cannot experience ourselves.

For testimonial claims it is of great importance not only **what** is being said but also **how** it is said. For instance, imagine a man testifying a robbery, describing what he saw on the street, pointing out a person as an alleged robber but delivering his testimony with a “drunk sounding” manner of speech. Would his claim be considered as reliable? Probably not. Or one of the most common examples is irony. An irony is recognized via specific prosodic features. For instance, if we hear a man arguing how

domestic violence is one of the biggest problems of civilized world and we recognize signals of irony, that would instantly affect our perception of this man and the argument itself. We can find many examples in a real life arguing in which manner of speaking affects the strength of an argument as well as credibility of the speaker.

For the purpose of this research we assumed the credibility of a speaker (his *ethos*) being a part of an argument reconstruction in the scheme as following:

1. SPECIFIC PREMISE: Politician X has deep voice, speaks little bit louder, has no disfluencies.

GENERAL PREMISE: People who speak in specific style (deep voice, little louder, without disfluencies) are perceived as trustworthy, credible (based on previously mentioned research).

CONCLUSION: Politician X is trustworthy, credible, knows what he is saying.

Further on,

2. SPECIFIC PREMISE: Politician X is trustworthy, credible, knows what he is saying.

GENERAL PREMISE: Politician X is saying Y.

CONCLUSION: Therefore, Y is true.

We can say that argument scheme 1 is argument from sign while argument 2 is argument from authority. In many real life situations when we do not have speaker's CV in hand and we are not fully familiar with the topic (for instance, climate changes, dangers of oil exploration etc.) our decision might depend on our "perception" of speaker's credibility, our "perception" of his expertise and competence. Of course, the verbal part of his argument, the content of his speech is undoubtedly the most important but much of this credibility "perception" depends on the manner of speaking (i.e. prosodic features of his speech). Both rhetorical studies but especially argument analysis neglected these elements in their work and I believe it should be seriously considered. To make this point stronger I conducted an empirical research joining insights from nonverbal communication with argumentation analysis.

3. AUDITORY ARGUMENTS – METHODOLOGY OF AN EMPIRICAL RESEARCH

Previous research (Kišiček, 2014, Kišiček 2016; Van den Hoven & Kišiček, 2017, Groarke & Kišiček, 2019) analysed examples from the public discourse (from advertising genre, journalism, judicial discourse etc.) to demonstrate the importance of prosodic features in an argumentative discourse. In some examples creators knowingly used prosody to make their argument stronger while in other examples prosodic features contributed to the argumentative discourse even without author's

intention (especially example from Van den Hoven & Kišiček, 2017; Groarke & Kišiček, 2018). However, this research deliberately manipulated prosodic features of different speakers to discover whether they will have an impact on listener's attitude and attitude change regardless of the arguments themselves (meaning that arguments remained the same while speakers with different prosody changed).

Six speakers (3 males and 3 females) read the same text (a comment from weekly newspaper) with a strong anti-immigrant standpoint. The same argumentative discourse was therefore delivered in different manner and 87 listeners assessed it. The survey was conducted on-line so one listener heard only one speaker (not all of them!) because the main intention was not for the listener to compare different argument deliveries but to decide on the argument strength. At the beginning of the survey listeners were instructed to imagine that they are listening to a politician in a European Parliament arguing against immigrants. The intention was to get as closer as possible to a real life situation. Audience is listening politicians argue, debate, deliver their arguments of a certain standpoint and then decide who to trust or for which politician to vote. However, it has to be clear, this research has no intention in belittling the importance of the verbal argument. The research is constructed in a way to examine: will the manner of speaking i.e. **how** the argument is delivered have any influence on the listener?

3.1. Argumentative text

The text used for this research was a real newspaper comment in Croatia which got a lot of attention due to increasing number of immigrants illegally crossing the border between Bosnia and Herzegovina and Croatia (the first EU country they can enter). The author of the comment is one of the public persons recognized for anti-immigrant attitude. And the topic itself is more or less in the centre of attention in Croatia (as well as in EU).

Speaker's standpoint was clearly anti-immigrant claiming they should not be given asylum in Europe. The standpoint was dominantly supported with emotional appeals (appeal to fear). The first sentence in the text is an argument presented as a fact "immigrants are terrorizing people of domicile country" and corroborated with the examples of terror: robbing houses, begging, molesting minor girls; upsetting domicile people (*Bosnian citizens are upset: immigrants sleeping in parks, muggings, begging, sexual harassment, breaking and entering, violating basic social norms like urinating in parks, defecating on inappropriate places are only some of everyday offenses*). Further on, anti-immigrant standpoint is supported by police reports on confiscation of items which can serve as a weapon (*police forces in Bihać made a raid in immigrant's*

centres and found dozens of knives, bats, hammers and other breaking and entering tools). Author of the text also uses statistics (17% of immigrants have higher education while most of them are uneducated and only 15% of them are employed in the countries that gave them asylum while others receive welfare money. Immigrants have basic difficulties – they are unable to learn language and the effectiveness of their work is significantly lower than from European workers) to support the “fact” of immigrant uselessness in sense of labour (they are not educated and do not wish to work) and therefore they cannot contribute to asylum country. Even more, they are on country’s expense because 85% of immigrants are receiving welfare money.

To make his claim stronger, author also mentions several possible counter-arguments and then refutes it. One, and most commonly heard argument in countries of former Yugoslavia is an appeal to pity: “They are running away from war as we did once”. Author sees it as manipulation and emotional appeal claiming that most of the immigrants are not war refugees but economic migrants and fugitives from law i.e. criminals (*Although there is a complete chaos in Bosnia caused by immigrants, some media is justifying it and compare it with refugee wave using emotions as manipulation. However, truth is completely different. Unlike war refugees from Syria, Yemen, Iraq or Afghanistan, in this case we are talking about economic immigrants which are using every possible situation to enter EU and to escape from law in their own countries*).

Second counter argument (used by civil right organizations) is protecting immigrant’s human rights. Refutation of “civil rights argument” is pointing out hypocrisy of people representing civil rights (*Civil rights organizations claim that they require “civilized minimum” for immigrants and similarly like in Croatia, they offer their own houses to accept immigrants – but just until it actually comes to this – in practice their offer disappears*). Author refutes it with an *ad hominem* argument.

The text ends with numbers of immigrant which certain cities will have to accept under the EU directives (appeal to fear) referring to the text published in distinguished Serbian newspaper (*Serbian magazine “Today” published preliminary numbers of Demostat research about immigrants’ acceptance in different Bosnian cities: Velika Kladuša 3000 immigrants, Bihać 2500, Gradiška 1500, Banja Luka 1500, Bijeljina 2500, Travnik 2000, Sarajevo 1000, Trebinje, 2500 and most of them will be set forth Mostar, 3000 people. Of course, here is also some for Grude (1000), Čapljina (2000), Široki Brijeg (2000)*).

Argumentative text has a clear standpoint and provides different argument types to support it (appeal to fear, argument by example, statistics, arguments from authority, arguments from sign etc.) therefore, it was considered as a good corpus for the experiment.

3.2. Speakers

Argumentative text was read by 6 different speakers with different prosodic features which were deliberately manipulated to illustrate certain emotional state and personality traits of the speaker.

1. D.N. is a male speaker with higher pitch (higher than average) and softer voice quality. He also spoke with lower intensity (quieter) and with slow tempo. Argument was delivered with many disfluencies and illogical pauses. These prosodic features are connected with lower self-esteem, insecurity, lack of dominance and confidence, lack of authority.

2. D.S. is a male speaker with a low pitch and volume (very pleasant voice quality), voice type which would be considered as an attractive male voice. Text was delivered with good interpretation i.e. logical pauses and logical word emphasis, appropriate tempo and loudness. Based on prosodic features he would be perceived as competent, strong, confident, trustworthy.

3. J.B. is a male speaker with higher pitch and uneven pitch range (high intonation beginnings and endings) and very loose articulation which can be perceived as casual (even under the influence of alcohol), speech rate is changeable as well as the loudness. This kind of speaking style may be connected with the lack of seriousness, expertise, too casual, informal and intoxicated.

4. I.B. is a female speaker of average pitch but breathy voice quality. Text was delivered with slower tempo and less intensity (quieter) almost whispery. Combination of these prosodic features is connected with empathy, warmth and softness of character, lack of dominance and authority. These prosodic features are common in comforting situations, compassion and consolation.

5. I.C. is a female speaker of higher pitch and especially important is high intonation beginning and ending of almost every sentence and faster tempo. This type of speaking style is distinctive for positive emotional states like joy, happiness, carelessness. And it can be perceived as informal and not serious enough (even childish), cheerful and inappropriate for serious, official situations.

6. M.D. is a female speaker with lower pitch and hoarse voice quality. Text was delivered with somewhat faster tempo, louder and with more vocal tension. The most important prosodic feature in this example is staccato rhythm which is always connected with commanding style. Combination of these prosodic features results with the perception of anger as well as determines, self-confidence and rigidity.

Six speakers represented different character types (from high dominance to low dominance), different emotional states (from empathy to anger) and as a result, I believe, different persuasiveness power.

3.3. Evaluators and survey

As mentioned above, 87 (42 males, 45 females) evaluators heard the arguments and had to assess it. Listeners were ranging from the ages of 20 to 55, from different social statuses, different level of education (from high school education to PhD's), different professions (journalists, professors, plumbers, hairdressers...) and different origin (various Croatian cities and villages). On line survey provides an opportunity to reach different profiles of people. Survey was designed in a way that evaluators received a link and when they clicked on it one of 8 speakers (randomly chosen) was the one they heard. First, they needed to read the instruction:

Listen to the speaker and imagine you are hearing the politician speech on the immigrants' topics. Then answer several questions in the survey. Keep in mind that you will hear audio recordings so listen in a quiet place or take your hand phones.

Then listeners filled in demographic data, and answered preliminary questions as follows:

I. preliminary part (1 absolutely NO – 5 absolutely YES)

1. Do you support immigrants entering the Croatia
2. Do you think immigrants are safety risk for Croatia?
3. Should Croatia provide asylum for those who decide they want to stay?
4. Do you think that Muslim immigrants are endangering European values?
5. Are you afraid of immigrants?

The point of preliminary data was to test the attitude change. After preliminary questions were answered, evaluators listened to audio recording (duration was approximately 4 minutes) and then answered questions in the second part of the survey, as follows:

1. Is the standpoint clear (1 not at all – 5 completely)
2. How strong is argumentation? (1 very weak – 5 very strong)
3. How persuasive is the speaker (1 not at all – 5 very persuasive)
4. How much did your attitude toward immigrants changed? (1 – not at all – 5 completely)

4. RESULTS AND DISCUSSION

Results of this research showed that prosodic features of an argument delivery do have some influence on argument assessment and persuasive

power of the speaker. Preliminary questions on immigrant attitude revealed that evaluators are quite indecisive (average score for all answers was between 2 and 3) on immigrant policy which makes them good potential audience for attitude change (i.e. attitude shaping). They are on average little bit more inclined toward supporting immigrant policies and providing asylum for them and are on average not afraid of immigrants entering EU. Based on this preliminary results it can be expected that and argumentative text with a strong standpoint and supporting arguments will have influence on persuasion direction.

However, survey demonstrated that on average, attitudes of the evaluators did not change.

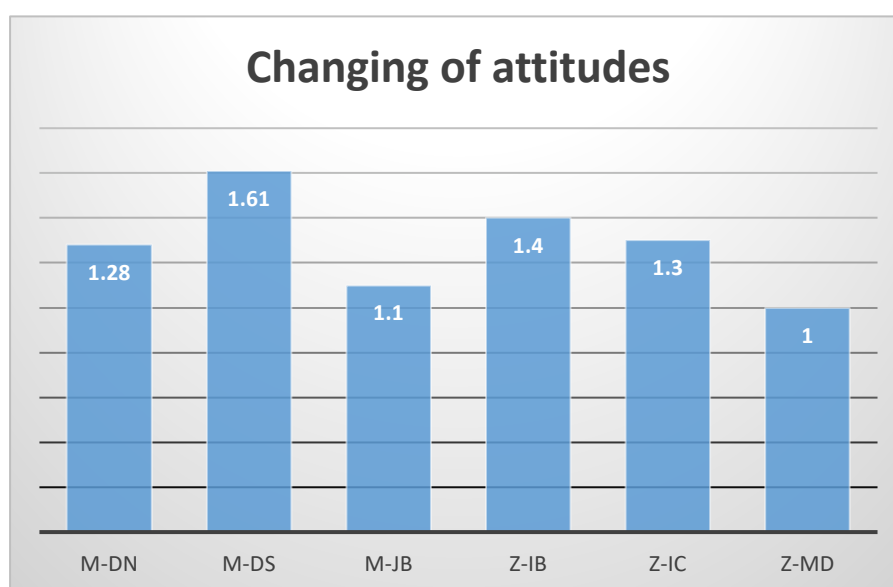


Figure 1 – How much did your attitude toward immigrants changed? (1 – not at all – 5 completely)

It is well known that changing of attitudes requires longer period of time and more exposure to continuous repetition of certain arguments but it is interesting to notice how the highest score for attitude change was in the example of the best combination of prosodic features (speaker D.S.). If we leave aside “attitude change”, we can look at the overall results for the “worst speaking style” and the “best speaking style” speakers.

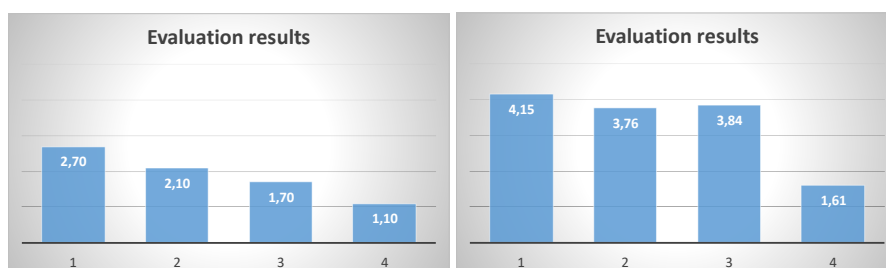


Fig. 2 Comparison in evaluation between the "worst speaking style" (J.B.) and "best speaking style" (D.S.)

J. B. was a male speaker with specific prosodic features (very loose articulation, changing of tempo and intensity, illogical pauses in text) which combined together resulted with a drunk sounding speech. He had the lowest score for the clarity of the standpoint (1) because it was very difficult to listen to the content of the speech. Prosodic features in his case were "too loud", "too informative" attracting attention more than the verbal message. The best speaker (D.S.) had high score for the clarity of the standpoint because all prosodic features in his case were "working together" with the verbal message i.e. did not draw attention from what has being said.

For the second question (argument strength) J.B. gain again the lowest score. The same arguments when delivered by J.B. were 2,1 while delivered by D.S. were 3,76. This shows how the manner of speaking did have influence on argument strength evaluation. Very similar was evaluation of the persuasiveness of the speaker (third question). J.B. was the least persuasive of all the speakers with 1,7 score while the most persuasive speaker of all was again D.S. with 3,84. This comparison, I think, confirms that prosodic features of delivery do have certain influence on evaluation of the argumentative discourse and persuasiveness of the speaker. By the term "evaluation" in this case I don't think on coherent, consistent argument evaluation as in argumentation theory but more of the everyday assessment which citizens do when listening the politicians. And this whole experiment was designed with intention to investigate does **how** we speak has any influence on evaluation of **what** we speak. And based on this comparison, the answer is yes.

As far as the other speakers are concerned, few more interesting results are worth mentioning. The second lowest result on speaker persuasiveness was the male speaker D.N. with prosodic features of an insecure man, low self-esteem and lack of confidence (weaker voice quality with higher pitch, slower tempo, longer and illogical pauses, quieter and non-fluent with speech errors and occasional stuttering). It confirmed previous research describing prosodic features of attractive

voices (Berry, 1991, 1992; Zuckerman & Driver, 1989, Zuckerman et al 1990, Zuckerman & Miyake, 1993) which are connected with personality traits such as confidence, self-esteem, determines and persuasiveness (Burgoon, Birk & Pfau, 1990). Speaker D.N. displayed all the opposite prosodic features which resulted with highly unattractive male voice and weak persuasiveness.

The most persuasive between female speakers was M.D. with prosodic features which are perceived as bossy (for female speakers), determined, strict and strong (staccato rhythm, louder, faster, harsh voice quality and lower pitch). However, as mentioned above, the changing of attitude result was lowest in her case. Perhaps, it can be explained that although speaker sounds persuasive, this combination of prosodic features is not considered attractive and correlates with the perception of negative personality traits such as “bossy”, “dictatorial”. It also confirms different criteria for male and female voices. According to Berry (1992) vocal attractiveness for male speakers is connected with traits such as competence, dominance, strength while for female speaker’s vocal attractiveness correlates with the perception of warmth, gentleness, honesty and kindness.

5. CONCLUSION

This paper presented results of an empirical research which main intention was to apply insights from nonverbal communication research to argumentation studies. Analysing, assessing and evaluating argumentative discourse neglected the influence of prosodic features which in some cases might have significant role in argument (re)construction. In real life argumentative situation, like in political debates, prosodic features can influence both persuasive power of the speaker and the perception of argument strength. This experiment revealed exactly that: the same argumentative discourse delivered by different speakers with different prosodic features was differently evaluated. Prosodic features which are connected with the perception of character traits such as credibility and trustworthiness insured the speaker more persuasive power and arguments were perceived as stronger. On the other hand, prosodic features connected with negative character traits resulted with the diminishing of persuasive power and argument strength. However, experiment also confirmed that attitudes do not change easily and if one wants to influence shaping or changing audience’s attitudes it depends mostly on argument itself (verbal part of the message). Weak arguments cannot become strong just by adopting prosodic features nor can strong argument be dismissed based on inappropriate prosody of the speaker. *Logos*, argument themselves,

content of the speech, verbal part of the message remains the most important part of the argumentative discourse but prosodic features which accompany it might have influence in overall argumentative process.

Based on the results of this empirical research it can be said that good arguments can be perceived as even stronger if delivered with favourable prosodic features (which contribute to the perception of trustworthiness, credibility and persuasiveness of the speaker) and weak arguments can be perceived as even weaker when delivered with unfavourable prosodic features (those connected with weakness of character, lack of competence and confidence).

ACKNOWLEDGEMENTS: I wish to thank all the speakers and evaluators who participated in this research. A special thanks to Jordan Bićanić, a colleague from the Department who helped in designing and conducting this experiment.

REFERENCES

- Aguinis, H., Simonsen, M. M., & Pierce, C. A. (1998). Effects of nonverbal behavior on perceptions of power bases. *The Journal of Social Psychology*, 138(4), 455-469.
- Berry, D. S. (1991). Accuracy in social perception: Contributions on facial and vocal information. *Journal of Personality and Social Psychology*, 62, 298-307.
- Berry, D. S. (1992). Vocal types and stereotypes: Joint effects of vocal attractiveness and vocal maturity on person perception. *Journal of Nonverbal Behaviour*, 18, 187-197.
- Bloom, K., Zajac, D.J. & Titus, J. (1999). The influence of nasality of voice on sex-stereotyped perceptions. *Journal of Nonverbal Behavior* 23(4), 271-281.
- Burgoon, J. K., Birk, T. & Pfau, M. (1990). Nonverbal Behaviours, Persuasion and Credibility. *Human Communication Research*, 17, (1), 140-169.
- Davis, M., Markus, K. A., & Walters, S. B., (2006). Judging Credibility of Criminal Suspect Statements: Does Mode of Presentation Matter. *Journal of Nonverbal Behavior*, 30, 181-198.
- DePaulo, B. M., Stone, J. I., & Lassiter, G. D. (1985). Deceiving and detection deceit. In B. R. Schlenker (Ed.), *The self and social life* (pp. 323-370). New York: McGraw-Hill.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74-118
- Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology*, 38(2), 270-292.

- Govier, T. (1993). When Logic Meets Politics: Testimony, Distrust, and Rhetorical Disadvantage. *Informal logic*, 15, (2), 93 -104.
- Groarke, L., & Kišiček, G. (2016). Compassion, Authority and Baby Talk: Prosody and Objectivity. In Benacquista, L. & Bondy, P. (Eds.) *Argumentation, Objectivity and Bias: Proceedings of the Ontario Society for the Study of Argumentation Conference*, Vol. 11. Windsor: University of Windsor.
- Groarke, L., & Kišiček, G. (2018). Sound Arguments: An Introduction to Auditory Argument. In Oswald, S., Maillat, D. (Ed). *Argumentation and Inference: Proceedings of 2nd European Conference on Argumentation*. London: Collage Publications, 177-199.
- Groarke, L. (2018). Auditory Arguments: The Logic of ‘Sound’ Arguments *Informal Logic*, 38, (3), 312-34.
- Guyer, J., Brinol, P., Petty, R., & Horcajo, J. (2019). Nonverbal behaviour of persuasive sources: A Multiple Process Analysis. *Journal of Nonverbal Behaviour*. Published on line 8th January. (Journal of Nonverbal Behaviour <https://doi.org/10.1007/s10919-018-00291>)
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129, 770–814.
- Kišiček, G. (2014). The role of paralinguistic features in the analysis of multimodal argumentation. *International Society for the Study of Argumentation (ISSA), 8th international conference on argumentation, at the University of Amsterdam, The Netherlands, July 1–4, 2014*.
- Kišiček, G. (2016). Prosodic features in the analysis of multimodal argumentation. In D. Mohamed & M. Lewinski (Eds.) *Argumentation and Reasoned Action* (pp. 629-643). Collage Publications: Milton Keynes.
- Kišiček, G. (2019). Auditory arguments – importance of sound in an Argumentative discourse. In Bart Garssen, David Godden, Gordon R. Mitchell, Jean H.M. Wagemans (Eds.) *Proceedings of the Ninth Conference of the International Society for the Study of Argumentation* (pp. 640-651), Sic Sat: Amsterdam.
- Laukka, P. (2008). Research on vocal expression of emotion: State of the art and future directions. In K. Izdebski (Ed.), *Emotions in the human voice. Vol 1. Foundations* (pp. 153–169). San Diego, CA: Plural Publishing.
- Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers’ ability to detect suspects’ lies. *Journal of Applied Psychology*, 89, 137–149.
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, 105(32), 11087–11092.
- Pittam, J. (1987). Listeners' evaluations of voice quality in Australian English speakers. *Language and Speech*, 30, 99-113.
- Pittam, J. (1989). The relationship between perceived persuasiveness of nasality and source characteristics for Australian and American listeners. *Journal of Social Psychology*, 130, 81-87.
- Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., Banissy, M. (2015). Dominant Voices and Attractive Faces: The Contribution of Visual and Auditory Information to Integrated Person Impressions. *Journal of Nonverbal Behavior*, 39, 355-370.

- Van den Hoven, P., & Kišiček, G. (2017). Processing multimodal legal discourse; the case of Stanley 'Tookie' Williams. In M. Manzin, F. Puppo & S. Tomasi (Eds.) *Studies on Argumentation & Legal Philosophy /2 Multimodality and Reasonableness in Judicial Rhetoric* (pp. 33-63). Università degli Studi di Trento: Trento.
- Vrij, A. (2000). *Detecting lies and deceit*. Chichester: John Wiley & Sons.
- Vukovic, J., Jones, B. C., Feinberg, D. R., DeBruine, L. M., Smith, F. G., Welling, L. L., & Little, A. C. (2011). Variation in perceptions of physical dominance and trustworthiness predicts individual differences in the effect of relationship context on women's preferences for masculine pitch in men's voices. *British Journal of Psychology*, 102(1), 37-48.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7), 592-598.
- Zuckerman, M. & Driver, R.E. (1989). What sounds beautiful is good: The Vocal attractiveness stereotype. *Journal of Nonverbal Behavior* 13, 2, 67 -77.
- Zuckerman, M., Hodgins, H., & Miyake, K. (1990). The vocal attractiveness stereotype: Replication and elaboration. *Journal of Nonverbal Behavior*, 14. 97-112.
- Zuckerman, M. & Miyake, K. (1993). The attractive voice: What makes it so? *Journal of Nonverbal Behavior*, 17, 119-135.

Be reasonable! Ways to react to cases of presumed unreason

ERIK C. W. KRABBE

*Faculty of Philosophy, University of Groningen, The Netherlands
e.c.w.krabbe@rug.nl*

JAN ALBERT VAN LAAR

*Faculty of Philosophy, University of Groningen, The Netherlands
j.a.van.laar@rug.nl*

Are we living in an age of unreason? And what to do about it? Can we combat unreason? We discuss situations in which one may presume to be confronted with unreasonable behavior by an interlocutor: fallacies, changing rules of the game, shifting to some other type of dialogue, and abandonment of reasonable dialogue. We recommend ways that could be helpful to obtain a return to reason. These possibilities lead us to a moderately optimistic conclusion.

KEYWORDS: Abandonment of reasonable dialogue, Dialogue shift, Fallacy, Game-change, Optimism

1. INTRODUCTION

The times seem to be changing. And it does not look like a change for the better, at least not to those who value reasonable argumentation as a means for peaceful conflict resolution. It seems that amidst the contemporary plethora of persuasive messages the use of reason is losing ground. Yet, it is nothing new to being confronted with attempts at persuasion from all sides. Forty years ago, Johnson and Blair wrote:

“As citizens we are constantly being offered persuasive rhetoric from a multitude of directions [...]. The teachers’ union, the school board, the city council, irate taxpayers, all are trying to gain your support for higher salaries, lower salaries; a strike, back-to-work legislation; city core redevelopment, rezoning for a suburban shopping mall, bikeathons want you to bike, telethons want you to phone in a pledge. [...] Groups and individuals incessantly vie for your

adherence to their way of seeing things, for your acceptance of their view of what is true, important or worth doing.” (Johnson and Blair 1983 [1977], p. viii)

In such circumstances, it wouldn’t be a good idea to give in to all these claims and therefore it becomes urgent to distinguish good arguments from bad arguments and to resist the latter. That is, we need to be capable of logical self-defense. To get the necessary skills, courses in informal logic, critical thinking, and argumentation theory can be helpful.

But the present situation seems harder to tackle than that described by Johnson and Blair: we seem nowadays to be bombarded by an indiscriminate avalanche of persuasive rubbish, not just fallacious arguments or inserted non-arguments but anything that is pseudo or fake: fake news, bullshit, crackpot theories, alternative facts, blunt inconsistencies and outright lies, seasoned by a sauce of mistrust and hatred. Unreason is ubiquitous, and we seem to get used to that as well; the observation is already old hat.

So are we really living in an age of unreason? What does that mean? Whither argumentation? Will it soon be ‘game over’ for any reasonable approach to our differences of opinion? Or can techniques of logical self-defense be bolstered so as to resist the avalanche?

Our paper is meant as a modest attempt to reflect on such issues. In Section 2 we discuss the concept of an age of unreason and defend the view that pessimism about the use of reason in our times is unwarranted. In Section 3 we resist the view that there is a struggle between two parties: the reasonable and the unreasonable. In Section 4 we describe several characteristic types of situation in which one may perceive a lack of reasonable behavior of one’s interlocutor and recommend certain ways of how to deal with these. Section 5 presents a moderately optimistic conclusion.

2. DO WE LIVE IN AN AGE OF UNREASON?

Donald Trump’s style of communication and argument is often perceived as exemplifying the heights of unreason. Here it looks as if all argumentation has been replaced by manipulation. In an article of little more than three pages Lakoff and Duran show how Trump uses language to frame and win debates and how he manipulates the press so as to inculcate his worldview. “Trump knows the press has a strong instinct to repeat his most outrageous claims, and this allows him to put the press to work as a marketing agency for his ideas.” (Lakoff and Duran 2018, p. 1)

Some of his linguistic manipulation techniques are:

- Weaponizing words (Hillary is always “crooked”, unwelcome news “fake news”, a threatening investigation “a witch-hunt”)
- Weaponized stereotypes (“...defaming entire groups of people as liars, rapists, terrorists...”)
- Weasel words (“...to avoid taking responsibility for a claim”: “Maybe”, “I don’t know”, “We’ll see”)
- Hyperbole (“great”, “terrific”, “the best/worst ever”, “a disaster”)
- Use of “winning” and “losing” (“Those who win deserve to win; those who lose deserve to lose”)
- Use of “America first” (“America is better than other countries, as shown by its wealth and power.”) (Lakoff and Duran 2018, p. 2)

Trump’s style is quite typical of those phenomena that make people believe that we have entered the age of unreason. We do not deny that these phenomena exist or that they can be upsetting. It is also upsetting that in public controversies people are gradually getting accustomed to being confronted with excessively unreasonable contributions. But even in these disturbing circumstances reasonable argument has not completely disappeared. Sometimes Trump gives us an argument. For instance, when he announced the US retraction from the Paris Climate Accord, he argued at the end of his speech as follows:

Example 1. Time to exit the Paris Accord

“The Paris Accord would undermine our economy, hamstring our workers, weaken our sovereignty, impose unacceptable legal risks, and put us at a permanent disadvantage to the other countries of the world. It is time to exit the Paris Accord – (applause) – and time to pursue a new deal that protects the environment, our companies, our citizens, and our country.” (The White House 2017, p. 7)

Now, one may criticize this argument on various accounts, but it can’t be denied that here we have an argument, and that Trump gave us at least five reasons (discussed earlier in the speech) why the Paris Accord would not be acceptable for the US.

Other examples of apparent unreason can be found in publications of European populist parties. Here also, we find that beside manipulation and bullshit, there are also arguments. Take, for instance, the German party *Alternative für Deutschland* [Alternative for Germany] (AfD). A meticulous argumentation analysis by David Lanius of their 2017 election platform reveals plenty of unreason: appeals to popular sentiments, prejudice, false or simplified statements, etc. Nevertheless, according to Lanius there are also many arguments:

“... the argumentation of the AfD is easy to grasp and can simply be put into a logically valid form.” (Lanius 2017, p. 29; our translation)

Given that the presence of lies, bullshit, appeal to popular sentiments does not exclude the presence of also some *prima facie* reasonable arguments, one may wonder whether unreason has really taken over. One may also doubt whether earlier ages fared much better than the present one. Is there any phenomenon that characterizes the supposed age of unreason and that did not occur in earlier times? If one wants to seriously investigate this question (which we won't), another question arises: When did the age of unreason start? In a paper on the terms “fake news” and “post-truth,” Joshua Habgood-Coote raises a similar question: When did the post-truth era start? After mentioning some options, going as far back as the Watergate scandal, he concludes:

“Most popular authors connect the era to the 2016 election, gesturing toward historical roots without providing any clarity about when it is supposed to have started [...]. The plethora of potential starting points suggests that ‘post-truth [era]’ has no clear extension. Everyone agrees that we are living in it, but no-one knows when it is supposed to have started.” (Habgood-Coote 2018, p. 10)

The same holds for the age of unreason, of which the post-truth era would be an aspect. The consequence is that the term “age of unreason” would be too vague to play a serious role in a philosophical discussion. Also, the idea of a golden age of norms and reason to which we could and should return is pure mythology. “There was never a golden age: the epistemic norms of democracy have never been realised in practice.” (Habgood-Coote 2018, p. 24).

But even so, the phenomena leading to the impression that there is more and more unreason are to be taken seriously. While all kinds of unreason, from slightly biased use of language to clear-cut bullshit and outright lies, may have been around through the ages, they also have always deserved to meet with criticism and opposition. Moreover, there may presently be an intensification of the symptoms of unreason due to technological innovations in communication technology. So there is enough reason for being on the alert.

Now should we be optimistic or pessimistic? Past experience shows that unreason has often been dealt with adequately. So why wouldn't we be able to do so in the future? On the other hand, there have also been many failures. So why wouldn't that happen again? We may opt for either optimism or pessimism or some mixture of both. These

observations may suffice to establish that it is at least a decent option to choose for a moderate optimism, and to continue the development of tools aimed at enhancing the use of reasonable argumentative exchanges in public life. Having reached this conclusion, it seems we could end the paper here. However there is more to say about what to do when confronted with unreason.

3. CAN WE COMBAT UNREASON?

The crucial question about unreasonable words or actions is: how should one react to them? Must we combine forces in a war on unreason? This martial metaphor may be used to strengthen our option for optimism: we are not alone in the struggle. In fact, many have decried the unreason of our times. But, even if, for the occasion, we condone the martial jargon, it may be asked – before we rush to battle: Who is the enemy? Is there really an opposing party of unreason that we, representing the party of reason, must combat?

Let us consider more precisely what is implied in assessing a statement, argument, or action as unreasonable. One cannot do so without, at the same time, assessing oneself as reasonable and possessing a sufficient amount of common sense to make the assessments. That everyone thinks of him- or herself as sufficiently provided with common sense is a well-known point made by Descartes at the very start of his *Discourse on the Method* (Descartes 1637). But actually, it is never excluded that in dismissing something as unreasonable we are ourselves missing the point and that what we perceive as unreason has a hidden rationality, as well as that what we see as our common sense points of view cover in reality our own pits of unreason. To admit that one could be wrong we see as a necessary component of a reasonable attitude. This does not mean that this fallibilistic insight need always come to expression. Rather it plays a role in the background and underlies one's willingness to change one's point of view if one becomes convinced of having been wrong.

For instance, you may be convinced that populist parties represent the pits of unreason. But, as Lanius has shown for the AfD, they also present arguments and so it would be reasonable to agree with them where these arguments are convincing. Similarly for Brexiteers and other populists. This doesn't mean that you join them; it may even be far removed from that. But it does mean, that in many cases, the combat metaphor is inappropriate. It suffices to oppose your interlocutors on those issues where you disagree with them.

In the case of Trump a combat metaphor seems out of place as well. Although he sells us a lot of bullshit (Kristiansen and Kaussler 2018), sometimes he really argues (above we gave an example of this).

So it could be reasonable to let him convince you on some points. At other points you may strongly oppose him.

The same holds for adherents of conspiracy theories and other crackpot views. After all there are also real conspiracies and views that were once considered outlandish have later gained acceptance. The appropriate mode for a reasonable opposition will not be to go into combat against such theories collectively but to judge each theory on its merits.

It seems then that to use a martial metaphor like “combating unreason” would be inappropriate or at least unnecessary (cf. Cohen 1995, Govier 1999, Ch. 4 and 14) and perhaps itself leading to a kind of unreason. Going around and blaming others for presenting “fake news” or “committing this or that fallacy” or “bullshitting” might create more heat than light, unless such claims are carefully underpinned. They could amount to no more than an unfair blaming strategy (Van Laar and Krabbe 2016). Certainly, it would be wrong to conceive of humanity as divided in two parties: the reasonable “us” and the unreasonable “they.” Everyone is sometimes reasonable and sometimes not. The whole combat metaphor, with its unnecessary military flavor, had better be dropped.

But what then can we do when we honestly perceive ourselves to be confronted by unreason?

4. WAYS TO REACT TO CASES OF PRESUMED UNREASON

What to do? It all depends on context! But let us discuss four characteristic types of situation that may occur in direct dialogical interaction, and see what the options are when you aim to support your side of a disagreement yet at the same time change the conversational setting so as to move towards overall reasonableness.

4.1 Dealing with isolated fallacies

You and your interlocutor are involved in a *persuasion dialogue*. That is, you both exchange arguments and critical considerations, within a dialogue that counts as a shared attempt to resolve your disagreements on the merits of both sides. The conversational contributions of your interlocutor provide no reason whatsoever to suppose that he or she wants to quit the persuasion dialogue. But then you notice that your interlocutor commits a fallacy by violating, willingly or unwillingly, a norm that is part of the very idea of resolving disagreements on the merits.

For instance, it could be that your interlocutor presents a fallacy, such as a Straw Man Fallacy, or a Fallacy of Loaded Terms. Very unfair.

Following the pragma-dialectical analysis of these fallacies, the interlocutor either violates the Rule for Critical Discussion according to which any critical response should genuinely relate to the commitments of the other party rather than to some distorted or fictitious version of it (Straw Man Fallacy), or the Rule for Critical Discussion according to which a defense should not falsely pretend to be based on shared starting points (Fallacy of Loaded Terms, as we think it can be understood within the pragma-dialectical approach to fallacies). What can you do if you want to support your side but also to redirect the course of dialogue into a more proper direction?

Acting reasonably, in our dialogical framework, implies that one is acting in such a way that the outcome of the dialogue will do full justice to the reasons available to the participants. When you assume that your interlocutor commits a fallacy, you are *prima facie* committed to ensure that this alleged norm violation does not impair the quality of the dialogue's outcome. As said, it all depends. Possibly, in your assessment there's no need to discuss the fallacy, as it will probably have no effect whatsoever, say because it concerns a minor issue, or because the addressees (you yourself or the attending audience) will not be led astray by it.

But then, possibly, leaving the fallacy untouched may have a distorting effect, in which case it needs to be defused. Otherwise, when one leaves the fallacy untouched, the fallacy may bias the outcome, or one may convey (inadvertently) the message that future fallacies will be left unchallenged as well. As we are not dealing, for now, with a setting in which the interlocutor systematically tries to exploit fallacies, but rather with a setting in which a presumably isolated, but noteworthy, fallacy occurs, we recommend to raise a point of order (Hamblin 1970) by identifying the contribution at hand, and by assessing it as a breach of an important norm for argumentative exchanges. Initiating such a *metadialogue* (Krabbe 2003) can be done in a blunt, insensitive, and arrogant manner – so that the fallacy charge puts the required spirit of cooperation at risk. But this is an issue of presentation and style, since there are also ways to express a fallacy charge with elegance, respect, and modesty. Thus, our recommendation would be that, if you need to bring up a charge of fallacy, you will do so in a rhetorically efficient way, so as to increase the likelihood that you and your interlocutor will come to agree either that the contribution was indeed fallacious or that, on the contrary, your fallacy charge cannot be sustained. After having made the necessary retractions, both of you should then return to the *ground level dialogue* as soon as possible.

Example 2. *Very high levels of intelligence*

"[Trump:] "I don't see" the devastating climate change effects warned about in the report. "One of the problems [is] that a lot of people like myself, we have very high levels of intelligence, but we're not necessarily such believers" [...]. "You look at our air and our water and it's right now at a record clean." The 1,600-page National Climate Assessment, issued by the Trump administration, details the climate and economic impacts U.S. residents will see if drastic action is not taken to address climate change." (Hayes 2018)

Trump can be seen as violating the pragma-dialectical Relevance Rule for critical discussion according to which one should advance argumentation relevant to one's standpoint (van Eemeren & Grootendorst 2004, p. 192) – because he appeals rather to his high level of intelligence than to substantial evidence or relevant expertise (a tactics known as the Ethical Fallacy or the *argumentum ad verecundiam*). Hayes's response can well be understood as criticizing Trump's inadmissible appeal to his personal qualities, but he does so in quite a detached and subtle manner by pointing to the contrast between Trump's personal estimation of the extent of climate change, and that of the National Climate Assessment's estimation issued by Trump's own administration. Hayes's approach seems adequate for dealing with occasional fallacies, but one may wonder how well it works in the case of Trump.

4.2 *Dealing with alternative views on rules*

Another way in which the contributions of your interlocutor may strike you as unreasonable, is when you notice that he tries to change the rules of game, or the way you presumed the rules were to be understood in the context at hand, and thereby to modify for the occasion the very idea of resolving disagreements on the merits (yet without going so far as to abandon the idea of obtaining a resolution by argumentation altogether).

For example, your interlocutor may flesh out "relevant to" in the Relevance Rule differently than you do. Or he may dismiss what you presumed to constitute a *prima facie* reliable kind of source of information – the press, climate science – as overall biased and untrustworthy, so that his specification of "appropriate argument scheme," a key concept in the Argument Scheme Rule,¹ differs from what could be expected when commencing the dialogue.

1 According to the Argument Scheme Rule, standpoints may be regarded as conclusively defended by argumentation not presented as based on formally

We recommend taking issue with your interlocutor on his views on the rules of the game, and to try to sort out collectively at a metalevel to what extent these are acceptable for both of you, and to do so even when your interlocutor, in a domineering manner, tries to present his views of the rules as *obviously* called for in the present context, rather than *as a proposal* to be discussed. Each proponent of a thesis who wants to propose an interpretation of the procedural rules must make sure that the addressee (the opponent) accepts that interpretation, given that argumentation can only be rationally convincing when starting from the (substantial and procedural) concessions of the addressee. But note that it may not be in the interest of reason if the addressee just clings to the rules as she has presumed and liked them. Such conservatism only counts as reasonable if it is or could be successfully defended in a metadialogue about the proponent's proposed interpretation. Acting reasonably also means acting according to a defensible dialogical procedure. Reason cannot be a monolith, and proposed procedural adaptations can sometimes withstand critical testing in an open-minded inquiry or a cooperative persuasion dialogue.

Example 3. *Big political agenda*

"WHAT MR. TRUMP SAID "Look, scientists also have a political agenda."

Asked about scientists who say hurricanes and other extreme weather events are worsening, Mr. Trump replied, "You'd have to show me the scientists because they have a very big political agenda."

THE FACTS

Scientists dispute that.

No doubt climate change has become politicized. And climate skeptics Sunday night cheered Mr. Trump's remark. But scientists took umbrage at the notion that their research has an agenda. Here are three in their own words:

Katharine Hayhoe, climate scientist, Texas Tech University: "A thermometer isn't Democrat or Republican. It doesn't give us a different answer depending on how we vote."

Andrew Dessler, climate scientist, Texas A&M University: "At its heart, this is just a wacky conspiracy theory," he wrote. "It's important to realize that there's never been a conspiracy by a

conclusive reasoning, only if the defense does take place by means of appropriate argument schemes that are correctly applied (see van Eemeren & Grootendorst 2004, p. 194).

huge field of science. And this would have to be an extremely massive conspiracy, considering the thousands of scientists working on this. On the other hand, there have been many examples (cigarettes, anyone?) where political advocates have tried to cast doubt on science that is extremely solid. That's what's going on here."

Donald Wuebbles, climate scientist, University of Illinois: "No scientists have political agendas. That's just an excuse."" (Friedman 2018)

The response by Hayhoe, though not incorrect, is weak since it is not very responsive to any considerations that might motivate an abandonment of appeals to the expertise of climate scientists. The response by Wuebbles might even be seen as scientific stonewalling. The response by Dessler misfires since it seems that Trump develops his stance without committing himself to there being any conspiracy by scientists – climate scientists might collectively have been led astray due to wrong incentives and social biases rather than by deliberate scheming. Yet, Dessler's response might be easily modified so as to provide a reasonable and convincing argument against the very idea that in general climate scientists are too politically driven to be taken seriously, and thereby in support of the *prima facie* reliability of arguments from expert opinion, also in the field of climate science.

4.3 *Dealing with shifts to other dialogue types*

Suppose you are, or thought you were, involved in a persuasion dialogue, but you start to doubt whether your interlocutor isn't trying to shift to some other type of dialogue. Such a shift away from a persuasion dialogue may be a *licit*, rather than an *illicit*, shift (Walton and Krabbe 1995, pp. 100-116), provided that the interlocutor does not conceal her attempt to shift but instead invites you to accept a proposal to turn to another type of dialogue, and thus in no way suggests that he succeeded in convincing you of his standpoint in the persuasion dialogue.

The parties may decide to postpone their persuasion dialogue but first turn to another type of dialogue in order to profit from the results of the latter, so that this other dialogue will be functionally embedded in the persuasion dialogue. For example, your interlocutor may catch you up on the news within an *information seeking dialogue*, before returning to the critical probing of the persuasion dialogue. Or, the two of you may try to find the correct answer to an open (undisputed) question within an *inquiry* – "what's the expected sea level rising by 2040?" – or to decide on a practical issue within a *deliberation dialogue* – "when will we set the deadline for the next National Climate

Assessment?" – before resuming the argumentative exchange. Further, you may try to come to a compromise agreement on some disputed issue in a *negotiation dialogue* before taking up the persuasion dialogue: "If you accept my definition of *green* then I'd be willing to accept your definition of *economically competitive*."

It is even possible, we think, that the parties licitly shift to an *eristic dialogue* – the kind of dialogue, such as a polemic altercation or a quarrel, which starts from a conflict and aims at no more than a reshaped relationship, such as a reshuffled intellectual or emotional hierarchy. Such a common dialogical goal still provides its participants with, admittedly minimal, norms for evaluating their contributions. Typically, one is allowed here to be quite impolite, to exploit rhetorical tricks, and to exert some emotional pressure. But within limits, beyond which contributions count as unseemly intimidation, harassment, or coercion. One requirement for eristic dialogue is that it be consensual, and that there is still some minimal level of mutual trust and cooperation. It is difficult to keep eristic dialogue under control, and clearly, it borders on unreason.

We recommend that you and your interlocutor take some time out to discuss in what type of dialogue you want to proceed. If, for instance, your interlocutor wants to negotiate, you may either agree or insist on first trying persuasion dialogue. You may be lenient, for example, when a climate skeptic teases you:

Example 4. *A big fat dose of global warming*

[Trump twitters:] "It's really cold outside, they are calling it a major freeze, weeks ahead of normal. Man, we could use a big fat dose of global warming!" (Cillizza 2017)

This ridicules the opposition and would be no good in a serious persuasion dialogue,² yet it can also be seen as shifting towards an eristic dialogue. It would be a waste of time to deal with this as something that needs serious refutation. Better to see it as a joke and respond in kind, or else to refuse to go along in that direction and insist on the importance of seriously discussing the issue of global heating.

4.4 *Dealing with abandonment of reasonable dialogue*

You begin to believe, or you even perceive, that your interlocutor is not, or no longer, committed to reason, because of a continual use of fallacies, bullshitting, and outright lies, in such a way that you cannot even interpret your interlocutor's behavior as an attempt to make a licit

² See on 'lost in the laugh,' Fearnside and Holther (1959).

shift to eristic dialogue. Therefore you are in doubt about the possibility of any serious dialogue with this interlocutor.

One option is to initiate a metadialogue about the nature of the interlocutor's previous more reasonable contributions (if any), and convince him or her to adopt or return to a kind of exchange in which you collaborate towards a reasonable outcome of some specific kind (a somewhat different kind of metadialogue than that in Krabbe 2003). Such a response fits the golden standard, yet chances are slim that it will make your interlocutor adopt or return to a more reasonable attitude. What is more, it requires an analytic approach, which may annoy members of your audience who perceive your expostulation as tiresome and pedantic.

A second option to be taken seriously is to retort in kind, so as to give the interlocutor an incentive to adopt, or to return to, a more reasonable attitude, in which case you follow suit (van Laar and Krabbe 2016; see also Jacobs 2009). You can use such a tit-for-tat strategy to communicate the message that it is expedient for your interlocutor to return to a reasonable dialogue – in which case you only *seem* simply to retort in kind, whereas, on another level, you really try to commence a metadialogue. However that may be, if, to your regret, your interlocutor decides to retaliate the battle is on and the dialogue off.

Thirdly, one may, more or less ostensibly, ignore the interlocutor's withdrawal from reasonable dialogue, and act as if the two of you are still engaged in a reasonable exchange. You can explain to your interlocutor at what points you remain unconvinced, and what faults he commits in the dialectic, yet without provoking him by labeling his behavior as exemplifying unreason. In this way, you may entice him to follow your good example, or you may at least put across your message to an audience of onlookers. Of course, you could fail to be persuasive after all, and onlookers might mistake your tolerance as a failure to stand your ground.

5. CONCLUSION

There is reason for a moderate optimism, which may encourage us to further investigate the merits of various critical reactions to cases of presumed unreason. But in this we should avoid to fall prey to the pitfall of considering our difficulties with presumed unreason as so unprecedented and exceptional that these would justify “the reasonable us” to enter into a combat using all and any means to beat “the unreasonable they”. Instead, we listed a number of situations in which one may come to suspect that the interlocutor is unreasonable, to some degree. For each of these situations we discussed at least one way to respond to the interlocutor that is critical and promotes an exchange of

reasons. But then, in line with our adherence to the idea that reasonable dialogue is inquisitive, and that one's own contributions should never be immune from criticism, it is no surprise that we have been unable to find any response that can be expected to settle the matter and to do away with unreason in general. In sum, when unreason seems ubiquitous, reasonable discussion may not be a panacea, but will still be a pertinent asset, worthwhile to be studied in dialectical argumentation theory.

ACKNOWLEDGEMENTS: We wish to thank the audience of a presentation of this paper for their questions and critical comments (3rd European Conference on Argumentation, Groningen, June 2019). We also acknowledge the support of COST Action CA17132 'European Network for Argumentation and Public Policy Analysis.'

REFERENCES

- Cillizza, Chris (2017). Donald Trump doesn't think much of climate change, in 20 quotes. CNN, August 8, 2017. Retrieved March 11 2019 from <https://edition.cnn.com/2017/08/08/politics/trump-global-warming/index.html>
- Cohen, Daniel (1995). Argument is war...and war is hell: Philosophy, education, and metaphors for argumentation. *Informal Logic* 17, 177-188.
- Descartes, René (1637). *Discours de la méthode: Pour bien conduire sa raison, & chercher la verité dans les sciences [...]* [Discourse on the Method: For Rightly Conducting One's Reason and Seeking the Truth in the Sciences [...]]. Leiden: Ian Maire.
- Fearnside, W. Ward, & Holther, William B. (1959). *Fallacy: The Counterfeit of Argument*. Englewood Cliffs: Prentice-Hall.
- Friedman, Lisa (2018). 'I Don't Know That It's Man-Made,' Trump Says of Climate Change. It Is. *The New York Times*, October 15, 2018. Retrieved from <https://www.nytimes.com/2018/10/15/climate/trump-climate-change-fact-check.html>, March 11, 2019
- Govier, Trudy (1999). *The Philosophy of Argument*. Edited by John Hoagland with a preface by J. Anthony Blair. Newport News VA: Vale press.
- Hayes, Christal (2018). Trump is skeptical of climate change report because of his 'high levels of intelligence', USA TODAY, November 27, 2018. Retrieved on March 11 2019 from <https://eu.usatoday.com/story/news/politics/2018/11/27/donald-trump-cites-high-levels-intelligence-climate-change-denial/2132448002/>
- Habgood-Coote, Joshua (2018). Stop talking about fake news! *Inquiry*, 1-33. DOI: 10.1080/0020174X.2018.1508363.

- Hamblin, Charles L. (1970). *Fallacies*. London: Methuen. Reprinted with a preface by John Plecnik and John Hoaglund: Newport News, VA: Vale Press, 1986.
- Jacobs, Scott (2009). Nonfallacious rhetorical design in argumentation. In Frans H. van Eemeren & Bart Garssen (Eds.), *Pondering on problems of argumentation: Twenty essays on theoretical issues* (pp. 55–78). (*Argumentation library* 14). Dordrecht: Springer.
- Johnson, Ralph H. and J. Anthony Blair (1983). *Logical Self-Defense*. 2nd ed. Toronto: McGraw-Hill Ryerson. First ed. 1977.
- Krabbe, Erik C. W. (2003). Metadialogues. In: Frans. H. van Eemeren, J. Anthony Blair, Charles A. Willard and A. Fransisca Snoeck Henkemans (Eds.), *Anyone Who has a View: Theoretical Contributions to the Study of Argumentation* (pp. 83-90), Dordrecht: Kluwer.
- Kristiansen & Kaussler (2018). The bullshit doctrine: Fabrications, Lies, and Nonsense in the Age of Trump. *Informal Logic*, 38, 13-52.
- Lakoff, George P. & Gil Duran (2018). Trump has turned words into weapons: And he's winning the linguistic war. *The Guardian* 13 June 2018.
- Lanius, David (2017). *Wie argumentieren Rechtspopulisten? Eine Argumentationsanalyse des AfD-Wahlprogramms* [How do rightwing populists argue? An argumentation analysis of the AfD election platform]. Karlsruhe: Institut für Technikzukunft (Diskussionspapiere 4). Retrieved on September 10 2019 from: <https://davidlanius.de/wie-argumentieren-rechtspopulisten>.
- Stanley, Jason (2015). *How Propaganda Works*. Princeton, NJ: Princeton University Press.
- The White House (2017) Statement by president Trump on the Paris climate accord. <https://www.whitehouse.gov/briefings-statements-president-trump-paris-climate-agreement/>
- van Eemeren, Frans H. & Rob Grootendorst (2004) *A Systematic Theory of Argumentation: The Pragma-dialectical Approach*. Cambridge: Cambridge University Press.
- van Laar, Jan Albert and Erik C. W. Krabbe (2016). Fair and Unfair Strategies in Public Controversies. *Journal of Argumentation in Context* 5 (3), 315-347. DOI 10.1075/jaic.5.3.04kra.
- Walton, Douglas N., & Erik C. W. Krabbe (1995). *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. Albany, NY: State University of New York Press.

Epistemic Injustice and Deep Disagreement

THIRZA LAGEWAARD
Vrije Universiteit, Amsterdam
t.j.lagewaard@vu.nl

This paper wants to contribute to a better understanding of 'deep disagreement' by arguing that sometimes, disagreements are deepened due to epistemic injustice. I explore a case of deep disagreement: the debate in the Netherlands about racism. This dispute should be understood as a deeper disagreement, because there is disagreement about what counts as evidence for the claim that racism is a significant issue in the Netherlands, due to both testimonial injustice and hermeneutical injustice

KEYWORDS: [Deep disagreement, Epistemic injustice, Epistemic principles, Hermeneutical Injustice, Netherlands, Racism, Testimonial injustice]

1. INTRODUCTION

My goal is to contribute to a better understanding of what has been called *deep disagreement*: disagreements which involve disagreement about underlying epistemic principles. Deep disagreements are not just a theoretical puzzle for social epistemologists. As Kappel (2012) and Lynch (2010) highlight, they can cause practical problems for collective decision making, because collective choices often depend on shared factual beliefs. This way, deep disagreements can hamper collective choices and policy making.

The main claim of this paper is that sometimes ordinary disagreements become deep as a result of epistemic injustice, i.e., injustice that occurs when someone is wronged specifically as an epistemic subject (Fricker 2013: 1320; Fricker 2017: 53). The paper thus explores a hitherto unnoticed connection between two phenomena that have received ample attention in recent social epistemology: (deep) disagreement and epistemic injustice.

The central idea is that when (pre-existing) epistemic injustice comes into play in a regular disagreement, this can lead to higher-order disagreement about what counts as evidence concerning the original

disagreement, which makes the disagreement deep. Introducing *injustice-based deep disagreement* highlights moral and political aspects of disagreements that might seem factual.

The plan is as follows: in section 2 I introduce and modify a common definition of deep disagreement and propose that the depth of disagreements is best understood as a matter of degree: disagreements can be more or less deep. Next, in section 3, I introduce and explore a case study of real-life disagreement: the disagreement about whether racism is a significant issue in the Netherlands, illustrated by the case of 'Black Pete'. As the Netherlands is often seen as a liberal and tolerant place, where one might expect questions about racism to be addressed in a cool and evidence-based manner, focusing on the debate on racism in this country will be especially helpful to illustrate my points. In section 4 and 5, I argue that there is disagreement about what counts as evidence in the case study because of two forms of epistemic injustice: testimonial and hermeneutical injustice. In section 6, I discuss how these epistemic injustices deepen the initial disagreement about racism and conclude that the intersection of disagreement and epistemic injustice is a fruitful area for future work in social epistemology.

2. WHAT ARE DEEP DISAGREEMENTS?

In this section, I discuss the way I'll characterize deep disagreement in this paper. Deep disagreement should be differentiated from 'regular' disagreement. In a regular disagreement, there is often a lot of background agreement about how to solve the disagreement at hand. For example, when disagreeing about which day of the week it is, both parties will agree on how to solve this disagreement (for example, by consulting a phone).

In a deep disagreement, there is also disagreement about how to solve the disagreement. Lynch (2010), Kappel (2017) and Matheson (2018) all define deep disagreement as disagreement about 'fundamental' or 'basic' epistemic principles. An *epistemic principle* tells us how we should form our beliefs. Such a principle concerns what counts as reliable evidence for what and/or what counts as justified belief regarding a certain domain (Lynch 2010). For example, the epistemic principles of tasseography tell us that we can gain justified beliefs about our fortunes by interpreting the patterns of coffee grounds. We all accept certain epistemic principles when forming and updating beliefs.

One way to distinguish between different epistemic principles is by separating *fundamental* principles from *derived* epistemic principles. As Matheson (2018: 3) puts it: 'Fundamental epistemic principles are simply basic; they are not derived from any other principle.' Examples

of such fundamental principles are those concerning visual perception, deduction or introspection. In the end, any arguments for the reliability of these principles will be circular (Fogelin 1985; Alston 1986; Feldman 2005). In this paper, I am assuming deep disagreement comes in gradations: they can also be about 'relatively fundamental' epistemic principles.

A deep disagreement, then, involves about epistemic principles. But not just any evidence: it's a disagreement about relatively fundamental epistemic principles.

This 'looser' characterization of deep disagreement is the one I'll use in this paper. In the next section, I'll explore a case of real-life deep disagreement: the debate about racism in the Netherlands, in order to show how epistemic injustice can deepen disagreement.

3. RACISM

In the Netherlands, there is disagreement about whether racism in Dutch society is a significant problem (Gorashi 2014; Wekker 2016; Essed 2018). While this may look like an ordinary disagreement that could be resolved easily by attending to the relevant data and experiences, I will show that it is in fact a deep disagreement, because there is underlying disagreement about what counts as evidence for the claim that racism is a significant issue in the Netherlands. Although this disagreement has moral aspects as well, my focus is on its epistemic aspects.

The disagreement on racism I am interested in here, then, is about whether everyday racism (Essed, 1991) and the systemic inequalities it produces are a significant problem in Dutch society. Racism is a significant issue, I propose, when people of color are structurally at a disadvantage as a result of bigger or smaller inequities in various parts of their lives due to racial discrimination.

There is ample statistical evidence that in Dutch society racism is indeed a significant problem. For example, people with a 'foreign' sounding name have a lower chance to be invited for a job interview than equally qualified people with a Dutch sounding name, even if these latter people have a criminal record (Van den Berg et al. 2017). In addition, people of color have a harder time renting houses or apartments (Rasit & Tielbeke 2018). Highly educated Dutch with a non-western background are more often unemployed (Huijnk et al. 2014) and youth with a migration background are, compared to their peers without migration background, suspected and convicted in higher numbers for the same kinds of offences.

However, due to the coded and ingrained nature of everyday racism, it's hard to point to specific actions and establish 'objectively'

that they are indeed clear cases of racism. This leaves room for disagreement. And indeed, there is plenty of disagreement about whether the above evidence establishes that racism is a significant problem. Many Dutch people tend to see themselves as tolerant and anti-racist (Wekker 2016: 1). Hondius (2012: 273) characterizes a broadly shared sentiment in Dutch society when she writes: 'Racism is simply 'not done', also meaning to suggest literally that it does not happen; it is considered self-evident that variety in skin tone is unimportant, irrelevant, and meaningless.' She adds that this denial of racism has the consequence that there is not a lot of debate about it in public discourse.

So far, the debate on racism might look like an ordinary disagreement that could be rationally resolved by apprising people of the relevant evidence and testimonies pertaining to the occurrence of everyday racism in Dutch society. However, the debate has become extremely polarized and entrenched and has led to political and even physical clashes. How did it get this far? I will now go on to argue that the disagreement about racism has deepened because it involves disagreement about what constitutes good evidence for the proposition that racism is a significant issue in the Netherlands. The main source of evidence that is contested are the testimonies of people who have first-hand experiences of racism. Because this source of evidence is contested, the disagreement about racism becomes deep. This deepening is caused by epistemic injustice, or so I will argue. The contesting of the testimony constitutes part of the racism that is debated. In section 6, I will present my argument in more detail, but first, I turn to epistemic injustice.

4. TESTIMONIAL INJUSTICE AND DEEP DISAGREEMENT

Much has been written about epistemic injustice following Fricker's (2007) introduction of the term (Kidd et al. 2017 provides an excellent overview).

In this section, I show how testimonial injustice can deepen disagreement. In the next, I'll do the same for hermeneutical injustice. The disagreement about racism in the Netherlands will continue to serve as my main illustration throughout.

I'll now go on to argue that the debate about racism in the Netherlands is plagued by testimonial injustice: the victims of racism aren't given enough credibility. Then, I show how this turns disagreement about racism into deep disagreement.

We should start by noting that sometimes, a non-dominant group can have an epistemic advantage over the dominant group when it comes to knowledge and understanding of oppression (Mills 2007, Dormandy

2018). This is because, As Berenstain (2016) and others note, someone who experiences a form of oppression, like racism, has one additional way of acquiring knowledge about oppression compared to others who don't suffer from it. Everyone can learn about oppression through testimony from those who are oppressed and through scientific evidence about it, but only the oppressed themselves can acquire knowledge about oppression through first-personal experience of it. The personal experience of a minority group member who experiences racism is a distinct source of evidence bearing on the racism-question, because this experience is significantly different from that of the majority in at least two ways: it is privileged and (partially) private.

Knowledge or justified belief about racism from first-personal experience is *privileged*, precisely because it concerns *first-personal* experiences of interpersonal interactions. Such experiences are privileged because of the social position of people who experience racism. As feminist epistemologists such as Harding (1993) and Pohlhaus (2011) have emphasized, your social position in the world shapes how you see the world. It shapes what you notice and what you pay attention to. Differences in experiences (due to differences in social position) can lead to differences in epistemic perspectives. For our case, this means that people who have personal experiences of racism have a perspective on the world in which racism is more salient than for people who do not experience racism. As these experiences of racism can be very consequential for their lives, people who experience racism are bound to be more sensitive and perceptive when it comes to racism. People who not experience racism, due to their social position, may have a 'blind spot of racial insensitivity' (Medina 2016: 185). From their (dominant) point of view, there might be a lack of evidence for the significance of racism.

In saying this, I am not claiming that privileged access is infallible or that one's own experiences are always the best guide to understanding racism. A person's own experience of a situation does not always accurately reflect that situation. Someone's emotions and previous experiences influence their more recent experiences. For example, they might experience a situation as more negative than it is, because they are tired. People who experience racism can be mistaken about what the situation actually is like. As a default, however, people will be in a better epistemic position when it comes to racism, than people who do not experience racism. Also, as the number of people reporting similar first-personal experiences increases, the scenario that *all* of these people are wrong about their own experiences *all* of the time becomes more and more unlikely.

Knowledge or justified belief about racism from first-personal experience could also be seen as *private*, at least to a degree. This means

that it cannot (easily) be shared what it is like to be the victim of racism. One might object to this that, say, white people can come to know what it is like to be the subject of racism by imagining it, because they might have experienced other forms of discrimination, such as those based on gender or age. This is too quick, however, because oppression does not produce equivalent results for all oppressed groups (Hills Collins 2007: 212). Sexism, for example, plays a different role in the lives of black and white women, because of their different races (Grillo & Wildman 1991: 399). Although different kinds of oppression might be connected, they but will be experienced very differently by different groups, making it hard to make useful comparisons (Hills Collins 2007: 210).

Because experiences of racism are both privileged and private, the evidence gathered by the minority who experience racism is different from the evidence that the majority can gather from *their* experience and from the scientific study of racism. Because of this, the white majority in the Netherlands has a different body of evidence than the minority who experience racism. Hence, the members of this latter group have an epistemic advantage over the majority on this specific issue.¹ For someone who doesn't experience racism, the *testimony* of those who have first-personal experience of racism ought to be an important source of evidence about whether racism is a significant problem. This is where the problem is located; it's the reliability of this evidence that is contested due to testimonial injustice, which deepens the disagreement.

Members of the dominant group could take the disagreement as good news: an opportunity to learn from others and to revise their beliefs accordingly (Christensen 2007). This way, by trusting that people of color possess important evidence about racism and accepting their testimony, the disagreement could be resolved. However, in the case of racism in the Netherlands – as well as in many others – this is not what happens. It's plausible that this is because of racial bias. Judgements about whether someone is a credible testifier are made very quickly and are based on appearances (Sperber et al 2010). Especially, we make social judgment about people based on their *faces* (Hugenberg & Wilson 2013) and we do this after being exposed to them for less than a second (Todorov et al. 2009). Someone's implicit race biases are a strong predictors of their evaluations of trustworthiness: a person with implicit race biases will have less social trust in a person of another race

¹ Of course, this isn't to say that each and every member of the minority will always have more or better evidence about racism than each and every majority member. The point is that, in general, the minority has direct access to important evidence that the majority lacks, or has only indirect second-hand access to.

(Stanley et al. 2011). These implicit racial biases are not rare, they occur widely. Research suggests that white perceivers often make negative social judgements about people with Afrocentric features and/or darker skin tones based on their facial appearances (Hugenberg & Wilson 2013: 171-173). So, people of color are often judged to be less trustworthy by white perceivers. It seems likely that this extends to judgments about trustworthiness of their testimony.

All this seems to support the claim that the testimony of people of color on racism is often not given the appropriate credibility *due to racial prejudice*, as is suggested by a.o. Mills (2007). Only some of the people who do not experience racism, take testimony about racism to be a weighty source of evidence.

We can now see how testimonial injustice deepens the disagreement on racism. As noted in section 2 above, a disagreement is deep when it involves disagreement about relatively fundamental epistemic principles. This is exactly the effect of testimonial injustice: it adds to the original disagreement a higher-order disagreement about epistemic principles governing the evaluation and uptake of testimony. In addition to the original disagreement about whether racism is a significant problem, there is now a further disagreement about whose testimony counts as good of evidence to settle this question or about how testimony by victims of racism ought to be weighed against other sources of evidence. More specifically: the majority who don't experience racism themselves implicitly or explicitly reject an epistemic principle that stipulates how testimony ought to be treated – or, even more precisely, they reject the application of this principle to the case at hand – whereas the minority who does experience racism firsthand takes (this application of) such a principle to be correct. The result is a lack of evidence for the significance of racism, from the point of view of the dominant group.²

One might object to this by denying that the relevant epistemic principles are 'relatively basic'. After all, the characterization of deep disagreement given above requires the disagreement to be about relatively basic epistemic principles. In response, note that it is hard to provide clear and objective general criteria for when principles are 'relatively basic'. But one strong reason to think that the principles at stake in the present case ought to count as relatively basic, is that it's difficult to see how someone who is doubtful of the probative value of testimony about firsthand experiences of racism could be convinced otherwise, without relying on claims about features of such testimony. That is, it's difficult to see how one could give a noncircular argument for the principles at stake. Firsthand experience is, by definition,

² This resembles white ignorance as described by Charles Mills (2007)

(partially) inaccessible to others. Someone who doesn't think firsthand experience is a privileged and private source of evidence will not be convinced when you point out that *undergoing* racism or other forms of oppression is different from merely observing it or learning about it through systematic scientific research, precisely because this latter point already assumes that there is something epistemically unique and important about firsthand experience.

I conclude that testimonial injustice causes the disagreement about racism to become deep. Let's turn to how hermeneutical injustice can deepen disagreement next.

5. HERMENEUTICAL INJUSTICE AND DEEP DISAGREEMENT

Hermeneutical injustice, too, can play a role in deepening disagreement. After a short detour through standpoint theory, I show how hermeneutical injustice can deepen the disagreement on racism.

As Pohlhaus (2012) describes, when such epistemic resources are formed, a dominant group will tend to have more influence than a non-dominant group. They will have a stronger influence on which epistemic resources are available and used. In this way, epistemic resources that are used to make sense of what goes on in a society, come to reflect the way the dominant group sees that society. They describe and make sense of the world largely from the *situation* or *epistemic perspective* of the dominant group. From the point of view of the non-dominant group, however, there may be gaps in the language, concepts, and criteria that are used to describe the world on a communal level (Pohlhaus, 2012). An example of this is the epistemic resource of racism, on which I will elaborate below.

With this in mind, I turn to hermeneutical injustice. As discussed above, the form of hermeneutical injustice I am focusing on occurs when a person is hindered in sharing conceptual resources which she herself possesses with people outside her group (Dotson 2012: 32; Fricker 2013: 1319; Fricker 2016: 166-167; Medina 2017: 43-44,).

Ignorance of concepts employed by marginally situated knowers need not be intentional and might be overcome. However, if members of a dominant group continue to refuse to do something about their ignorance when confronted by it – when they refuse to learn the conceptual resources they missed out on – the result is *willful* hermeneutical ignorance (Pohlhaus 2012).

How is hermeneutical injustice related to the debate on racism? The conceptual resources required for describing and making sense of racism (the skill to use the relevant concepts) are very well developed, in particular in communities that experience racism. A very general example is the aforementioned conceptualization of racism as 'everyday

racism'. Another example is the concept of 'institutional racism', which refers to racist practices of social and political institutions, like in healthcare policies or housing policies. Concepts like these are used to describe and recognize instances of racism and to make sense of them. They enable a perspective on the world in which racism is salient and can be described, discussed, and analyzed.

In the Netherlands, too, concepts describing racism are well established within certain minority communities. On a communal level and in public discourse, however, the occurrence of racism is often denied in the Netherlands (Ghorashi, 2014: 103). Often, the suggestion that racism is a problem in the Netherlands is met with a strong dismissive attitude or with utter silence. Using the word 'racism' is often seen as a way that migrants try to silence their opponents (Ghorashi 2014:113).

Hence, there is a lack of uptake in society at large of the rich epistemic resources available to describe 'racism'. This is a case of hermeneutical injustice.

That there is no uptake of the conceptual resources relevant to racism becomes clear when one considers that, in spite of minor shifts in recent years, racism is barely addressed in the Netherlands. There even seems to be an unwillingness to use the term 'racism' (Witte 2010: 17). According to Wekker (2016: 153-154), this is because racism is supposedly not an issue. If people disagree, they are accused of overreacting, being overly sensitive and being unable to take a joke (ibid:32). Because white people don't experience racism, the assumption that the Netherlands is not racist is the dominant way to reflect on Dutch society. As a result of this silence on racism, there is no shared vocabulary to talk about racism in the Netherlands. There is, from the dominant point of view, no need for elaborate conceptual resources to describe and make sense of experiences of racism.³

As a result, the group that is the object of racism is severely limited in their ability to discuss racism constructively. First, because there are no, or very little, shared epistemic resources because of the different social experiences of the two groups. Secondly, because people who experience racism have trouble fruitfully discussing their experiences with people outside their group, because of (willfull) hermeneutical ignorance. As Hondius (2014: 274) writes: 'What is not explicitly mentioned is hard to challenge.' This lack of discussion about racism further reinforces the false belief on the part of the dominant group that racism is not an issue in the Netherlands.

³ I'm overgeneralizing here to keep it simple. There are many positions in this debate and not all white Dutch people think the same way and not all Dutch non-white people think the same way.

In short, epistemic resources are developed and shared to reflect the experiences of communities, but when there is oppression, some communities might be epistemically left out in the process of developing or sharing these resources. They might form their own resources, reflecting *their* experiences, but they are not shared on a communal level. The result is a lack of shared vocabulary.

This lack, in turn, might strengthen the idea that first-personal testimony on racism is unreliable. As was discussed in the previous section, a member of a dominant group might judge someone who testifies on racism as unreliable because of identity-prejudices (testimonial injustice). But another reason to judge someone as unreliable is when their testimony does not make sense to you, which may be caused by a lack of shared concepts (hermeneutic injustice). For example, if a testifier asserts 'Black Pete is racism', when you take yourself to know that Black Pete is just part of an innocent tradition, this doesn't make sense. So why listen?

We can now see how hermeneutical injustice deepens disagreement. A lack of shared concepts to talk about racism leads to higher-order disagreement about what counts as credible and trustworthy testimony. The dominant group will live by the generally sensible epistemic principle that they judge testimony which is couched in concepts and terms that they don't understand as unreliable. But the higher-order disagreement that arises concerns the *application* of that principle to the case at hand. While the dominant group will see this case as a straightforward instance of testimony that makes little sense, the non-dominant group will see this application as unjustified, because the dominant epistemic agents ought to know better, or at least make an effort to educate themselves, rather than dismiss the testimony of non-dominant groups out of hand.

6. CONCLUSION

In the case study in this paper, there is disagreement about (the application of) epistemic principles concerning (A) whether private first-personal experience of racism is a weighty source of evidence in this domain (weightier than third-personal experience of the dominant group), (B) whether victims of racism count as important testifiers in this domain, and (C) how to assess testimony that is not (fully) intelligible to you because it employs concepts and terminology you are unfamiliar with. The dominant group can easily deny that there is anything new or relevant to be known about racism in the Netherlands by dismissing the relevant testimony and epistemic resources, which boils down to disagreement on the level of epistemic principles concerning (A), (B), and (C).

Because of epistemic injustice, it is hard to argue for or against the validity of epistemic principles concerning (A), (B), and (C) in a dispute-independent way. The epistemic injustices at work make it difficult or impossible to exchange epistemic reasons that are accepted by both parties, which shows that the relevant principles are indeed of the relatively fundamental sort required for disagreements to count as deep. This means the disagreement becomes very difficult to resolve. Non-dominant epistemic agents lack the perceived credibility and tools to convince dominant agents of the existence and nature of their experiences.

Although I have focused on the case of racism and Black Pete as a case study in this paper, I want to suggest that disagreements like these, which become deep due to epistemic injustice, can and do occur more widely. This discussion might be useful to look at other cases involving racism, like disagreement about the Black Lives Matter movement. But it might also be used as a lens to look at disagreements involving different kinds of epistemic oppression. When a disagreement appears to have become deep and involves a group that is on the receiving end of epistemic injustices, it might be a case of injustice-based deep disagreement. Think of disagreements where the testimony of women, disabled people, old people, and chronically ill people about their own experiences is contested. Attention to the details of such disagreements and empirical research on them, could show whether these disagreements are indeed deepened by epistemic injustices. Identifying such cases will be relevant in so far as rationality and morality might require different responses to them than to ordinary disagreements and ‘classic’ deep disagreements. Hence, understanding injustice-based deep disagreements seems to me to be a project not only of theoretical interest, but also of great practical and social relevance.

ACKNOWLEDGEMENTS: Jeroen de Ridder, Chris Ranalli, Rik Peels, René van Woudenberg, Catarina Dutilh Novaes, Jan-Willem Wieland, Mikkel Gerken, Jennifer Lackey, Christoph Jaeger, Klemens Kappel, Tamarinde Haven, Wout Bisschop and Gijsbert van der Brink.

REFERENCES

- Aikin, S. (2018). Dialecticality and Deep Disagreement. *Symposion: Theoretical and Applied Inquiries in Philosophy and Social Sciences*, 5 (2), 173-179.
- Alston, W. P. (1986). Epistemic circularity. *Philosophy and Phenomenological Research*, 47(1), 1-30.
- Christensen, D. (2007). Epistemology of Disagreement: The Good News. *Philosophical Review*, 116(2): 187-217.

- Coady, D. (2017). Epistemic Injustice as Distributive Injustice. In I.J. Kidd & J. Medina & G. Pohlhaus, Jr. (Eds.), *The Routledge Handbook of Epistemic Injustice* (pp.). New York: Routledge.
- Dormandy, K. (2018). Disagreement from the Religious Margins. *Res Philosophica*, 95, 371-395.
- Dotson, K. (2012). A Cautionary Tale: On Limiting Epistemic Oppression. *Frontiers*, 33(1), 24-47.
- Dotson, K. (2014). Conceptualizing epistemic oppression. *Social Epistemology*, 28(2), 115-138.
- Essed, P. 2008. Everyday Racism. In J. Hartwell Moore (Ed.) *Encyclopedia of Race and Racism* (pp. 447-449). Farmington Hills: Thomson Gale.
- Essed, P. (2018). *Alledaags Racisme*. (revised edition). Amsterdam: Feministische Uitgeverij Sara.
- Feldman, R. (2005). Deep Disagreement, Rational Resolutions, and Critical Thinking. *Informal Logic*, 25(1), 12-23.
- Feldman, R. (2006). Epistemological Puzzles About Disagreement. In: S. Hetherington (Ed.), *Epistemology Futures* (pp.). Oxford: Oxford University Press.
- Fricker, E. (1995). Critical notice: Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony. *Mind*, 104(414), 393-411.
- Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
- Fricker, M. (2013). Epistemic justice as a condition of political freedom? *Synthese*, 190(7), 1317-1332.
- Fricker, M. (2016). Epistemic Injustice and the Preservation of Ignorance. In R. Peels & M. Blaauw (Eds.), *The Epistemic Dimensions of Ignorance*. (pp. 160-177). Cambridge: Cambridge University Press.
- Fogelin, R. (1985). The Logic of Deep Disagreements. *Informal Logic* 7: 1-8.
- Gerken, M. (2019). Pragmatic Encroachment and the Challenge from Epistemic Injustice. *Philosophers' Imprint* 19(15).
- Ghorashi, H. (2014). Racism and the 'Ungrateful Other' in the Netherlands. In P. Essed & I. Hoving (Eds.), *Dutch Racism* (pp. 101-117). Amsterdam/New York: Rodopi.
- Goldman, A. I. (2001), Experts: Which Ones Should You Trust? *Philosophy and Phenomenological Research*, 63: 85-110.
- Grillo, T. & Wildman, S. M. (1991). Obscuring the importance of race: The implication of making comparisons between racism and sexism (or other-isms). *Duke Law Journal*, 1991(2), 397-412.
- Harding, S. Rethinking Standpoint Epistemology: What is Strong Objectivity? In L. Alcoff & E. Potter (Eds.), *Feminist Epistemologies*(pp.) New York/London: Routledge.
- Hilhorst, S. & Hermes, J. (2016). 'We have given up so much': Passion and denial in the Dutch Zwarte Piet (Black Pete) controversy. *European Journal of Cultural Studies*, 19(3), 218-233.
- Hill Collins, P. (2003). Some group matters: Intersectionality, situated standpoints, and Black feminist thought. In T. L. Lott & J. P. Pittman

- (Eds.), *A Companion to African-American Philosophy* (pp. 205-229). Oxford: Blackwell.
- Hondius, D. G. (2012). Black Dutch Voices: Reports from a country that leaves racism unchallenged. In P. Essed & I. Hoving (Eds.), *Dutch Racism* (pp.). Amsterdam/New York: Rodopi
- Hugenberg, K. & Wilson, J.P. (2013). Faces are Central to Social Cognition. In D.E. Carlston (Ed.), *Oxford library of psychology. The Oxford Handbook of social cognition* (pp. 167-193). New York: Oxford University Press.
- Huijnk, W. & Gijsberts, M. & Dagevos, J. (2014) Jaarrapport integratie 2013. *Den Haag: Sociaal en Cultureel Planbureau*.
- Kappel, K. (2012). The Problem of Deep Disagreement. *Discipline Filosofiche* 22 (2), 7-25.
- Kelly, T. (2005). The Epistemic Significance of Disagreement. In: J. Hawthorne (Ed.), *Oxford Studies in Epistemology 1* (pp. 167-197). Oxford: Oxford University Press.
- Kidd, I. J., Medina, J., & Pohlhaus Jr, G. (2017). *The Routledge Handbook of Epistemic Injustice*. New York/London: Routledge.
- Lackey, J. (2008). Learning From Words: Testimony as a Source of Knowledge. Oxford: Oxford University Press.
- Lackey, J. (2018). Credibility and the Distribution of Epistemic Goods. In McCain Kevin (ed.), *Believing in Accordance with the Evidence*, pp. 145 – 168. New York: Springer.
- Leerkes, A. & Martinez, R. & Groeneveld, P. (2018) Minority paradoxes. Ethnic and generational differences in self-reported offending and official crime statistics. *British Journal of Criminology*, 59(1), 166-187.
- Lynch, M. P.(2010). Epistemic Circularity and Epistemic Incommensurability. In A. Haddock & A. Miller & D. Pritchard (Eds.) *Social Epistemology* (pp. 262-277). Oxford: Oxford University Press.
- Lynch, Michael P. (2016). After the Spade Turns: Disagreement, First Principles and Epistemic Contractarianism. *International Journal for the Study of Skepticism*, 6 (2-3):248-259.
- Matheson, J. (2018). Deep Disagreements and Rational Resolution. *Topoi*.
- McKinnon, R. (2016). Epistemic Injustice. *Philosophy Compass*, 11: 437–446.
- Medina, J. (2016). Ignorance and Racial Insensitivity. In R. Peels & M. Blaauw (Eds.), *The Epistemic Dimensions of Ignorance*. (pp. 178-201). Cambridge: Cambridge University Press.
- Medina, J. (2017). Varieties of hermeneutical injustice. In I.J. Kidd & J. Medina & G. Pohlhaus, Jr. (Eds.), *The Routledge Handbook of Epistemic Injustice* (pp.). New York: Routledge.
- Mills, C. (2007). White Ignorance. In S. Sullivan & N. Tuana (Eds.), *Race and Epistemologies of Ignorance*, (pp. 11-38). New York: State University of New York Press.
- Pohlhaus, G. (2012). Relational knowing and epistemic injustice: Toward a theory of willful hermeneutical ignorance. *Hypatia*, 27(4), 715-735.
- Ranalli (In press). What is Deep Disagreement? *Topoi*
- Rasit, E. & Tielbeke, J. (2018, March 28). Rachid is ook gewoon een nette jongen. Retrieved from: <https://www.groene.nl/artikel/rachid-is-ook-gewoon-een-nette-jongen>

- Rodenberg, J. & Wagenaar, P. (2016). Essentializing 'Black Pete': competing narratives surrounding the Sinterklaas tradition in the Netherlands. *International Journal of Heritage Studies*, 22(9), 716-728.
- Todorov, A. & Pakrashi, M. & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, 27, 813-833.
- Van den Berg, C. & Blommaert, L. & Bijleveld, C. en Ruiter, S. (2017). Veroordeeld tot (g)een baan: hoe delict- en persoonlijkheidskenmerken arbeidsmarktkansen voor veroordeelden beïnvloeden. *Tijdschrift voor Criminologie*, 59 (1-2), 113-135.
- Van Vliet L. & Kester, J. (2018). *Onderzoek: Uiterlijk van Zwarte Piet*. Opiniepeiling Rapport Eenvandaag. Retrieved from: https://eenvandaag.avrotros.nl/fileadmin/user_upload/PDF/Rapport_age_zwarte_piet_def.pdf
- Wekker, G. (2016). *White Innocence. Paradoxes of Colonialism and Race*. Durham: Duke University Press.
- Witte, R. (2010). *Al eeuwenlang een gastvrij volk. Racistisch geweld en overheidsreacties (1950-2009)*. Amsterdam: Askant.

Argumentative Patterns of Right-Wing Populism: The Example of the “Alternative for Germany”

DAVID LANIUS

Karlsruhe Institute of Technology (KIT)

david.lanius@kit.edu

Populism has become one of the most intensely discussed topics in both public debate and academic research. So far there has been no systematic argumentation theoretic analysis of populism, however. This paper is intended to provide first steps towards such an analysis by giving a full argumentation theoretic reconstruction of the political manifesto of the German right-wing populist party “Alternative for Germany” (AfD). This allows to draw preliminary conclusions about the AfD’s argumentative strategy as exemplary for right-wing populism.

KEYWORDS: argument reconstruction, argument analysis, discourse analysis, populism, right-wing populism, post-truth, populist argumentation, argumentative strategy

1. INTRODUCTION

Populism has become one of the most intensely discussed topics in both the public debate and academic research. Many scholars are investigating, trying to explain and theorizing about the new rise of – especially right-wing – populism in Western societies.¹ However, there has been no systematic argumentation theoretic analysis of right-wing populism so far. This is astonishing, since both scholars and practitioners frequently refer to “right-wing populist argumentation” and “right-wing populist argumentative strategies” in their political assessments.

This paper is intended to provide first steps toward an analysis of such “right-wing populist argumentative strategies” by giving a full argumentation theoretic reconstruction of the political manifesto for the federal election in 2017 by the German right-wing populist party

¹ See Mudde and Rovira Kaltwasser (2017) for an introduction, Pappas (2016) for an overview, Albertazzi and McDonnell (2008) for an anthology, and Rooduijn (2019) for a recent state of the field overview.

“Alternative for Germany” (AfD).² In this manifesto, as in the one for the European elections in 2018 and in their party platform from 2016, the AfD claims to identify the “real problems” of Germany that (supposedly) go unnoticed by the public debate. According to the AfD, downfall, chaos, and disaster for “the people” are imminent.

The argumentative strategy in both manifestos and the party platform seems the following: The AfD takes justified and unjustified fears in the population, nourishes them with doomsday scenarios, and then presents itself as the only savior in time of existential need. This pattern can also consistently be found in the political arguments of the AfD (and arguably most other right-wing populists).³ In this paper I will elaborate on that pattern and call it the “core argument of populism.”

It is the central argument in the AfD’s party programs; most other arguments merely provide support to its premises.⁴ Its reconstruction draws on the argumentation theoretic framework of Betz and Brun (2016), which is able to bring comparably high clarity to the subject of the analysis due to its level of detail and focus on inferential relations within the argumentation (see also Betz 2010). It requires the interpreter to reconstruct the arguments in question in their most plausible and coherent way (Brun and Hirsch Hadorn 2018).

The first step consists in reconstructing the conclusions and premises of the arguments and the inferential relations between them and other arguments in the argumentation. This first step is conducted without directly evaluating the argumentation. Note, however, that no reconstruction can be entirely objective; it is an interpretive act and thus guided by a certain point of view. It is one way (of several possible ways) to read the AfD’s party programs. The self-imposed demand is, however, that the reconstruction will also be acceptable to the authors (i.e., the political leaders of the AfD) due to its focus on plausibility and coherence.

Only in a second step will the findings then be evaluated. There are three key findings. First, the AfD’s argumentation is impressively consistent. Second, it relies on verifiably false premises for its arguments. Third, the AfD’s pattern of argumentation is based on a dichotomy of doom by “them” and salvation by “us.”

² This paper has immensely benefitted from helpful comments and suggestions by Gregor Betz, Georg Brun, Romy Jaster and Kathrin Kazmaier.

³ See Quent (2019). See also Adorno (2019).

⁴ This reconstruction builds on Lanius (2017), which is an analysis of the AfD’s manifesto for the federal election 2017 (in German).

2. THE CORE ARGUMENT OF POPULISM

The basic idea of the core argument of populism is that the populists have to come to power because only they can save society from imminent doom. Let's examine it in standard form:

- (1) **[Will of the People]:** Society can only be saved, if the will of the people is realized.
- (2) **[Voice of the People]:** Only if the populists come to power, the will of the people will be realized.
- (3) **[Salvation]:** Society can only be saved, if the populists come to power.
- (4) **[Doom]:** Society is doomed and must be saved.
- (5) **[Power]:** The populists must come to power.

Figure 1 shows how it looks as an argument map:

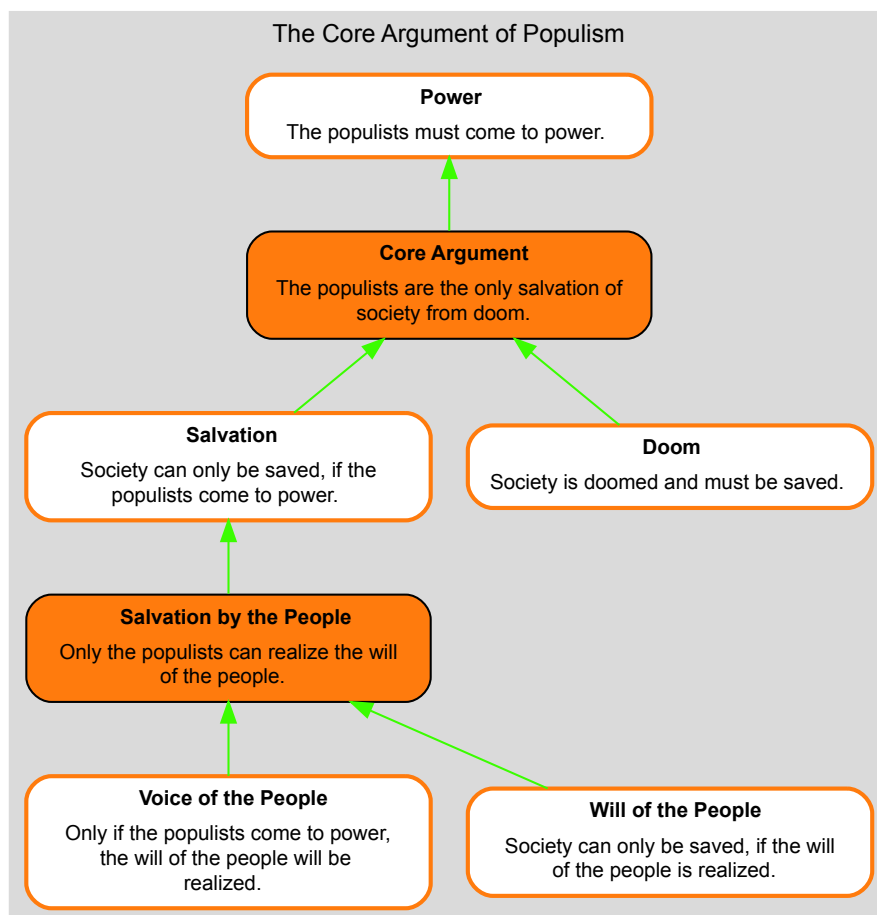


Figure 1 – The Core Argument of Populism

The argument consists of two sub-arguments. The first sub-argument consists of two premises and one conclusion. The first premise [Doom] states that society is doomed and must be saved. The second premise [Salvation] states that society can only be saved if the populists come to power. The conclusion [Power] states that the populists must come to power and follows logically from [Doom] and [Salvation].

Donald Trump's campaign slogans, for example, were based almost explicitly on this argument. His most important campaign slogan "Make America great again" presupposes that America is no longer great. Embedded in the context of his rhetoric, many of Trump's assertions and arguments during his campaign and presidency can only be understood as assuming that America is facing doom and must be saved.

Another important slogan of Trump is "Only I can fix it!" It presupposes that *it* needs fixing. What "it" exactly refers to is not entirely clear. Presumably, however, "it" refers to the United States itself or its government. Trump's slogan is thus almost synonymous with the second premise [Salvation]. The conclusion [Power] is implicit, as it is usual with natural language arguments: Donald Trump must come to and stay in power.

Both premises can also be found in many statements by AfD politicians, but also in the manifestos for the German federal election in 2017 and the European election in 2019. In the former it says: "The rule of law, especially the separation of powers, must be restored and the state must once again be able to guarantee its core tasks."⁵ This presupposes that the rule of law no longer exists in Germany and that the state no longer guarantees its core tasks. In the words (of the party platform from 2016): "We could not and did not want to stand idly by the violation of law and order, the destruction of the rule of law, and the irresponsible political action against the principles of economic reason." This presupposes that law and order are not being upheld, the rule of law is being destroyed and Germany is economically failing. The AfD assumes quite clearly in both passages that Germany is in one form or another on the brink of doom and in need of salvation.⁶

Implicitly, premise [Salvation] is even part of the AfD's name. It claims to be *the* alternative for Germany: the only party that is not part of the "political class whose primary interest is its power, its status, and its success at the polls," as it says in the manifesto from 2017. Only the AfD can save Germany.

⁵ All quotes from the AfD's political programs have been translated by the author.

⁶ As we will see, a significant part of the AfD's argumentation in its political programs is directed at showing that Germany is doomed, which seems a general strategy to justify radical action against the "establishment." Cf. Lilla (2016).

Premises [Doom] and [Salvation] are further justified. This means that new arguments are made, each supporting one of the two premises. The AfD's party platform and its manifestos, but also its (social) media messages contain such arguments. Premise [Doom] is justified by several threats to society such as immigration, Islam, and the loss of cultural identity. Across the various populist camps, these scenarios are surprisingly similar, and they demonstrably play a vital role in the AfD's political programs.

The justification for premise [Salvation] consists in the claim that the AfD and only the AfD speaks for "the people". This claim to sole representation is considered by Müller (2017) as defining feature of populism and is found almost literally in the AfD's political programs. The supporting – and by Müller's definition truly populist – argument for premise [Salvation] is the second sub-argument.⁷ It says that society can only be saved if the populists come to power because only then can the people's will be realized. Only if the AfD realizes its demands, will the "people be given the opportunity to introduce their own legislative initiatives and to pass them by referendum," as it says in the manifestos and party platform.

Only if the people become sovereign again can society be saved from imminent doom. Or in the words of the AfD from both manifestos: "We are convinced that the fundamental financial, energy, and migration crises, as well as the societal clash with Islam, cannot be managed viably by either the government or the parliament alone. This cannot and must not happen without the direct participation of the people." The AfD must hence come to power to give "the people" its voice. The premise [Will of the People] captures this fundamental assumption.

There are three unjustified premises in this argument: [Doom], [Voice of the People], and [Will of the People]. Despite being controversially debated in the political science literature, premise [Will of the People] is taken for granted. Lots of reasons are given, however, to further substantiate [Doom] and [Voice of the People].⁸

In the manifestos from 2017 and 2019, three main reasons can be found for [Voice of the people] in the form of arguments against the "Fake News", the political "Establishment" and the European Union. However, let us first look at several scenarios of doom, which are given as justifications for [Doom].

⁷ Cf. Taguieff (2006) for a more sceptical perspective on the project to define populism.

⁸ A reconstruction of the entire argumentation fully visualized with Argdown is available online as the example "The Core Argument of Populism" here: <https://argdown.org/sandbox>.

3. SCENARIOS OF DOOM

Premise [Doom] of the core argument is justified in various ways in the AfD's manifestos and party platform. The AfD puts forward (at least) nine arguments to show that society is doomed and must be saved. They all have [Doom] as their conclusions and can either be understood as inductive reasons for it or as a conjunctive deductive inference:

- **Immigration:** Immigration is a threat to society. (FE17: 5)⁹
- **Crime:** Crime and terrorism are threats to society. (FE17: 4)
- **Islam:** Islam is a threat to society. (FE17: 6)
- **Culture:** The disappearance of cultural identity is a threat to society. (FE17: 8, 9)
- **Globalization:** Globalization is a threat to society. (FE17: 3)
- **Iniquity:** Social injustice is a threat to society. (FE17: 10, 11)
- **Demography:** Demographic change is a threat to society. (FE17: 11)
- **Healthcare:** The failure of healthcare is a threat to society. (FE17: 12)
- **Innovation:** Technophobia is a threat to society. (FE17: 13)

Understood as inductive reasons, they each provide some justification to believe that society is doomed and they jointly (are supposed to) warrant the conclusion [Doom]. Being inductive reasons, not every argument must go through to do so, however. For instance, the AfD would still consider it proven that society is doomed if the argument on demographic change turned out to be unsuccessful (by its own standards). Understood as a conjunctive deductive inference, there is an intermediate argument between the nine arguments and [Doom], which contains nine premises about doom by immigration, crime, Islam, culture, globalization, iniquity, demographic change, failure of healthcare, technophobia, and a tenth premise stating that if all these threats exist, society is doomed and must be saved.

For the purposes of this paper it will not matter whether we interpret this argument inductively or deductively.¹⁰ Let's now examine the arguments themselves. One prominent theme runs through much of the argumentation in all political programs; namely the "threat of immigration". I will hence begin with the argument on immigration and show how it is linked to the arguments on crime, Islam, and culture (section 3.1). These arguments are not only closely connected to the argument on immigration, but also play an important role in both the AfD's election and media campaigns and other right-wing populist argumentations. The

⁹ The numbers refer to the section numbers of the AfD's manifesto for the federal election 2017 (FE17).

¹⁰ I take it to be more charitable to interpret the argument as inductive, but nothing of what follows depends on this.

remaining arguments for premise [Doom] will then be discussed collectively in the subsequent section (3.2).

3.1 Immigration, Crime, Islam, and Culture

Let's take a closer look at the argument on immigration. It is discussed in detail in Chapter 9 of the party platform (PP), Chapter 5 of the manifesto for the federal election in Germany 2017 (FE17), and Chapter 6 of the manifesto for the European election 2019 (EE19). It is strongly connected to the arguments on crime and Islam, but also bears on issues of national sovereignty, the national budget, the national healthcare system, and cultural identity. It can be reconstructed in form depicted in Figure 2.¹¹

According to the AfD, society is doomed because crime rates are skyrocketing. The link to "Ausländerkriminalität" (crimes committed by foreigners) is drawn at multiple instances in its political programs.¹² Islam is presented as a threat to the rule of law and liberal democracy.¹³ In general, immigration is considered the main threat to Germany and other European nations – due to its (alleged) destabilizing effects on national security, national economies, national healthcare systems, and cultural identities.

¹¹ Please note that Figure 2 and subsequent figures do not represent the entire arguments' reconstruction, which (apart from the structural ambiguity between an inductive and deductive inference discussed above) contains deductively valid arguments only. For the sake of clarity, uncontroversial and other less salient premises have been omitted. The numbers in brackets behind the statements indicate the section in the AfD's manifesto for the federal election in 2017, where the original arguments can be found.

¹² See PP16: 3.4, 3.7, 3.8, 9.6; FE17: 4.1, 4.6, 5.8; EE19: 6.2, 8.1.

¹³ See PP16: 7.6; FE17: 6; EE19: 8.5.

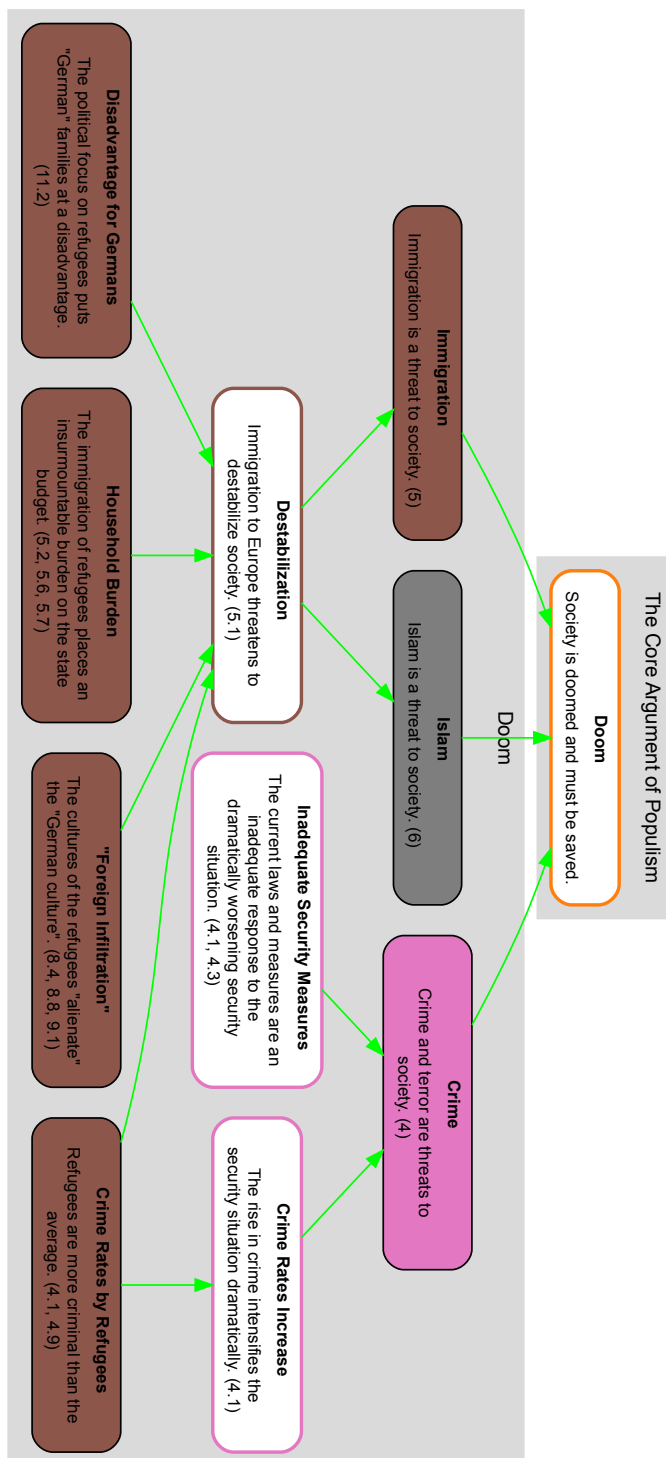


Figure 2: The Doom Arguments of Immigration, Islam, and Crime

According to the AfD's political programs, immigration brings not only criminals, terrorists, and "foreign cultures," but also people who burden the state budget and welfare system. These problems are multiplied because immigrants are reproducing faster than "German families" such that soon there won't be a country recognizable as "our Germany" anymore.¹⁴ This argumentation perfectly ties in with the narrative of the conspiracy theory of the "Great Replacement" – which is increasingly adhered to and propagated by right-wing extremists, masterminds, and high-level politicians, but also by many right-leaning voters.¹⁵

Although interwoven with the previous arguments on crime, Islam, and immigration, the argumentative thread on cultural identity is complex and merits a closer analysis. The loss of cultural identity is supposedly driven by a number of factors. The arguments' key thesis is that the disappearance of cultural identity is a threat to society. Traditionally, right-wing populists and extremists have focused on cultural hegemony and a return to cultural values that are considered in decline.¹⁶

While immigration is considered one major threat to "Germany's cultural identity," there are other independent threats, which are identified in the AfD's political programs. In particular, the AfD fears that the "traditional family" is losing its function. According to the AfD, this is a problem both because it is itself an important value and also because it multiplies, as mentioned, all the other problems due its (alleged) effect to demographic change.

Most room is made for two other (sub-)arguments, however: that "genderism" and "multiculturalism" – coming from within society – threaten "our way of life." "Gender-mainstreaming" destroys, so the AfD's manifesto, the "traditional family values" and "natural gender roles" in families. Furthermore, "gender ideology" is assumed "constitutionally invalid."¹⁷ The argumentation can be reconstructed as depicted in Figure 3.

¹⁴ This is most clearly expressed in FE17: 5.1.

¹⁵ See Betz (2018) or Bergmann (2018).

¹⁶ This has also been the German Nationalsocialists' strategy. Nowadays, the strategists of the AfD (and other right-wing populist parties) draw on Gramsci (2014) for this approach, however. Cf. Kailitz (2004).

¹⁷ See PP16: 7, 8; FE17: 7.7, 9; EE19: 8, 12.

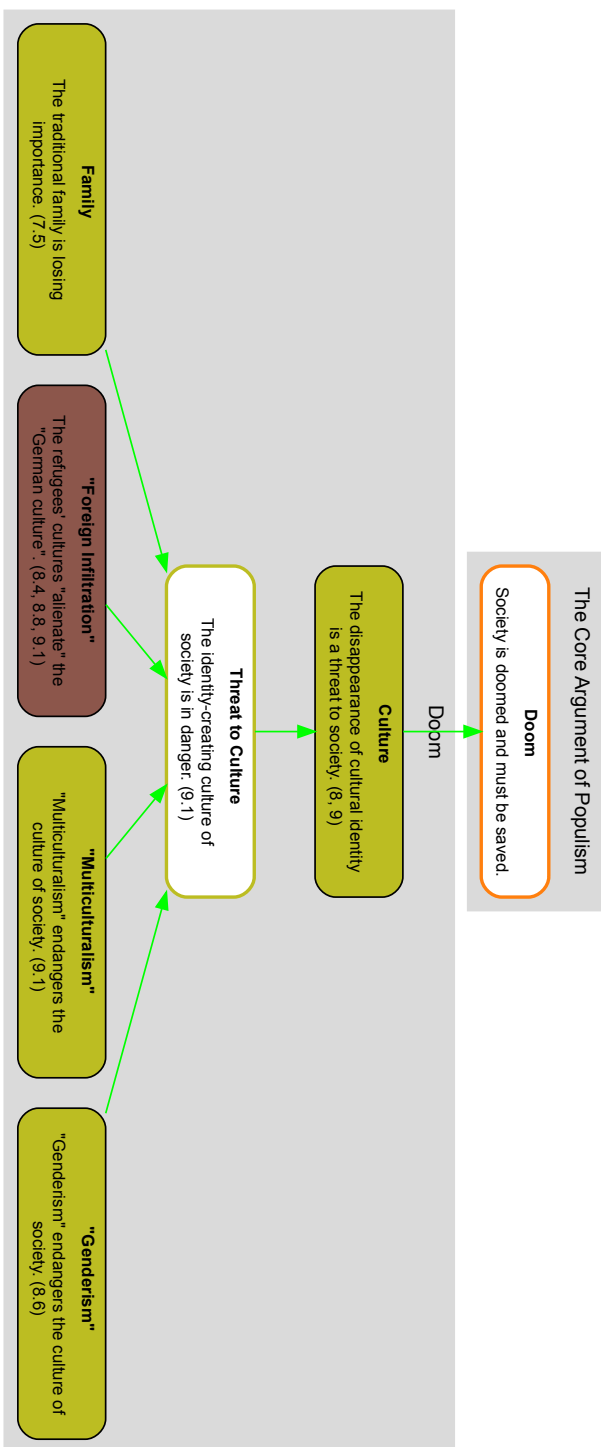


Figure 3: The Doom Argument of Culture

3.2 More Doomsday Scenarios

Various other scenarios of doom identify past achievements that have been lost, according to the AfD, in today's Germany. The AfD wants to reinstate the principles of liberty, democracy, and the rule of law (EE19:2.1/2.2), national sovereignty with respect to banking and finance (EE19:5.5), and to immigration and asylum policy (EE19:6.1), the freedom of research and teaching, and the scientific level of excellence (EE19:12.1), the neutrality of education (EE19:12.4), Germany's ability to defend its borders (EE19:3.2.2), internal security (EE19:8.1), and the competitiveness of German companies (EE19:5.3/13.4.3).¹⁸ The AfD wants to put the state into service of "the people" *again*.¹⁹ This implies – as the slogans of Donald Trump – that the mentioned states of affairs are not there anymore: Just as the United States used to be great, but is not great anymore; Germany used to be a democracy, but is not a democracy anymore.

Moreover, the threats identified to "us" and "the people" show how the AfD employs "us versus them"-schemes.²⁰ Much ink has been spilled on the "us"-part, i.e., the concept of *the people*, identifying it as a constructed concept for a fictitious entity – a culturally homogenous ethnic group (cf. Anderson 2006). Also, the "them"-part has been studied extensively. It usually instantiates a friend-enemy-thinking, which according to Schmitt (2002) governs "all true politics".

Particularly noteworthy is that the "us versus them"-thinking is not only used in relation to threats by "them out there" (the immigrants and Muslims), but also in relation to "them up there". "Those up there" are the reason why, according to the AfD, "the voice of the people" is not heard and the AfD cannot save society. And this is also explicitly shown in the AfD's political programs by specifically three more arguments supporting premise [Voice of the People] in the core argument.

4. THE ONLY SALVATION

The AfD stages itself as the savior from the threats to society discussed in the previous section. Only the AfD can save it because only they are uncorrupted and have not lost connection to "the people". The argumentation is straight-forward and can be found in the manifestos from 2017 and 2019 and party platform from 2016. It can also be found in most

¹⁸ Virtually the same points are made in the party platform from 2016 and the manifesto from 2017 in slightly different order and with slightly different focus.

¹⁹ Cf. Kämper (2017), who analyzes the AfD's party platform with respect to the use of the word "again" specifically.

²⁰ See Greene (2014) or Haidt (2013) for the disruption between moral groups.

other right-wing populist election programs, campaign programs, and political strategy papers. Let us now examine the three key ingredients in detail.

4.1 The “Establishment”

“We the people” against “them up there” – this is the classical topos of populism and it is not missing in the AfD’s argumentation. Also, Donald Trump frequently used the slogan “drain the swamp,” declared to “make our government honest once again,” and railed against the “establishment.” Classically, anti-elitism has been considered a defining feature of populism.²¹

The “establishment” is seen as the “enemy of the people”. The AfD’s manifestos and party platform contain arguments of this kind:

- (1) **[Political Alternative]:** Apart from the “establishment” only the populists can come to power.
- (2) **[Lost Connection]:** The “establishment” does not want to and cannot realize the popular will.
- (3) **[Realization]:** The will of the people will only be realized if a party comes to power that wants to and can realize the will of the people.

- (4) **[Voice of the People]:** Only if the populists come to power, the will of the people will be realized.

The following argument supports premise [Lost Connection] of the previous argument:

- (1) **[Corruption and Inability]:** The “establishment” is corrupt and incompetent.
- (2) **[Will and Ability]:** If the “establishment” is corrupt and incompetent, it cannot and will not realize the will of the people.

- (3) **[Lost Connection]:** The “establishment” does not want to and cannot realize the popular will.

While premise [Lost Connection] says that the “establishment” does not properly connect to the people – as opposed to the AfD, premise [Political Alternative] entails that only the populists are a realistic power that is not part of the “establishment.” The AfD demands that power be given back to the “sovereign citizenship of the Federal Republic of Germany” and that the “people (...) be sovereign again” (FE17: 1.3).

²¹ Cf. Barr (2009).

The AfD then criticizes (in FE17: 1.4) the government: Since the “established parties” (“Altparteien”) are corrupt and incompetent, they cannot and do not want to realize “the will of the people”. Premise [Corruption and Inability] can be found almost literally in the manifesto, where it says that the “omnipotence of the parties and their exploitation of the state endangers our democracy” (FE17: 1.7). It further argues that the current state of the party system be precarious because “numerous laws have allowed the separation of powers in Germany to erode over the years and have led to an exuberant state power” (FE17: 1.5).

The AfD must therefore come to power to give its voice to “the people.” The corruption and inability of the “established parties” prevent that Germany’s problems be solved. Only “the people” can do this and only the AfD is willing and able to facilitate “the people” doing it. The argument is visualized in Figure 4.

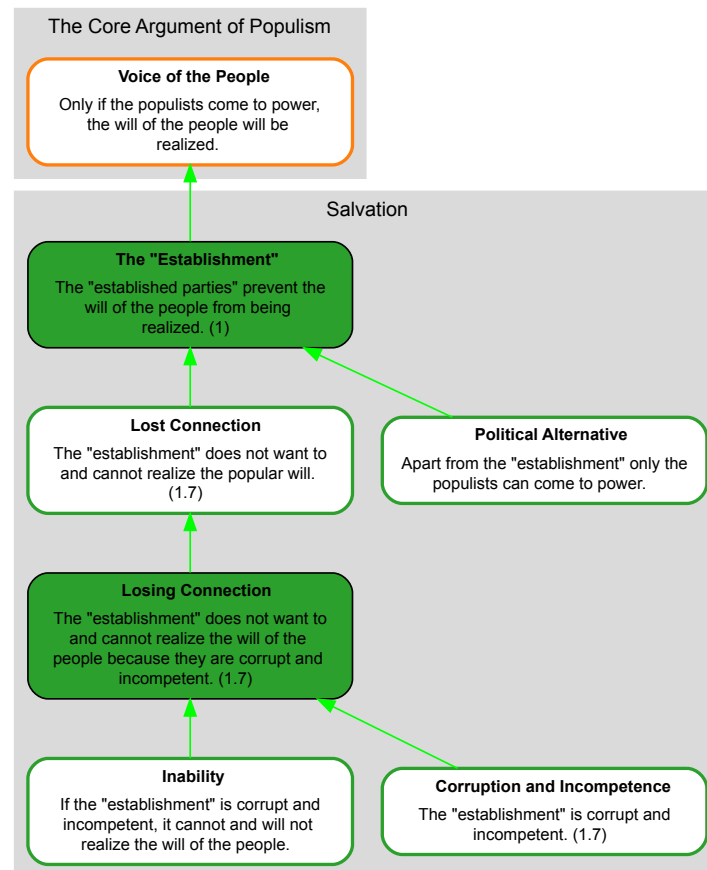


Figure 4: The Salvation Argument on the “Establishment”

4.2 The “Fake News”

Crucial to the AfD’s argumentation is also the alleged role of the traditional media, which is depicted as another “enemy of the people”. The argumentation contains the infamous allegations of “political correctness” and manipulation and “co-ordination” of the traditional media by the state.

The argument can be reconstructed in the following way:

- (1) **["Fake News" and Censorship]:** Freedom of expression will not be subject to any restriction or censorship unless the "fake news" is abolished.
- (2) **[Abolition of the “Fake News”]:** The “fake news” will only be abolished if the populists come to power.
- (3) **[Will of the People and Censorship]:** The will of the people will only be realized if freedom of expression is not subject to any restrictions or censorship.

- (4) **[Voice of the People]:** Only if the populists come to power, the will of the people will be realized.

The following argument supports premise [“Fake News” and Censorship] of the previous argument:

- (1) **["Political Correctness"]:** If the “fake news” prevails, the parties continue to use the instrument of "political correctness" and bring the press in line.
- (2) **[Opinion Formation]:** If the parties continue to use the instrument of "political correctness" and bring the press in line, the freedom of expression will be limited and censored.

- (3) **["Fake News" and Censorship]:** Freedom of expression will not be subject to any restriction or censorship unless the "fake news" is abolished.

The central premise of the argument is [“Fake News” and Censorship]. The manifesto (FE17: 1.7) says: "The omnipotence of the established parties is also cause (...) of the freedom-limiting ‘political correctness’ and the dictate of opinion in all public discourses." Premise [Opinion Formation] picks up another classic topos of populism. In the AfD’s words from the manifesto (FE17: 9.2) it reads: "We reject decisively ‘politically correct’ language requirements because they (...) restrict the freedom of expression." Premise [“Political Correctness”] is merely conceptual. It can be understood as a partial definition of “fake news.” In contrast to the more controversial (if not outright problematic) premise [Opinion Formation], it is only implicitly found in the manifesto. The argumentation is visualized in Figure 5.

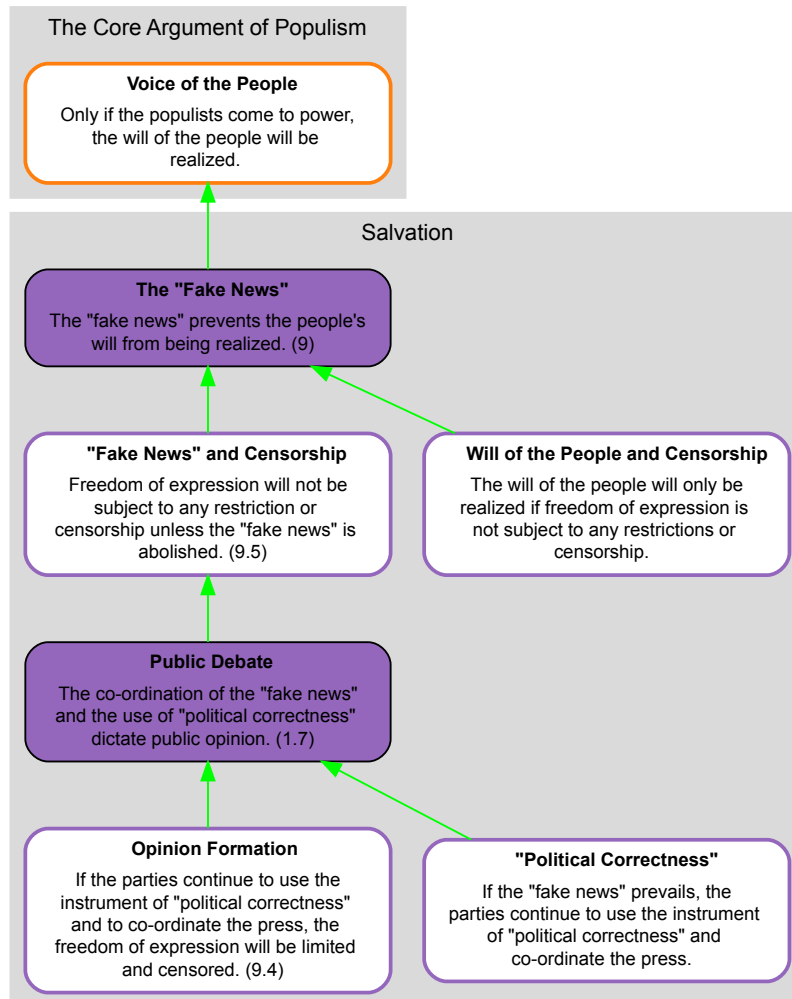


Figure 5: The Salvation Argument on the "Fake News"

In sum, the argument says that the traditional media prevent us from effectively finding solutions to the threats to society that only the AfD has identified due to its ability to recognize "the people's will" and because only the AfD is valiant enough to speak out against the "opinion dictate" ("Meinungsdiktatur") established by the "system." This is why the threats to society are not mentioned in the public and, for this reason, many people are not aware of them.

Based on this assumption can the AfD maintain the plausibility of most other premises in its political programs. This is a crucial argument to make sense of much of its argumentation. It is required to prevent contradictions between the AfD's positions and what other parties, the media, and most citizens in Germany think and say. It makes the

argumentation coherent because it is a *prima facie* good reason to doubt criticism from outside and inside the party.

4.3 The European Union

The last thread in the argumentation supporting premise [Voice of the People] is based on resentments toward the European Union. It was highly relevant to the European elections in 2019, and the AfD – but also many other right-wing populists such as Matteo Salvini’s Lega Nord or Viktor Orbán’s Fidesz – used it to convince people to vote for them.

This argument is very elaborate in the AfD’s manifestos. It can be reconstructed as follows:

- (1) **[No People's Will in EU]:** The will of the people will only be realized if we leave the EU.
- (2) **[Exit from EU]:** Only the populists have the political goal of leaving the EU.

- (3) **[Voice of the People]:** Only if the populists come to power, the will of the people will be realized.

The following argument supports premise [No People's Will in EU] of the previous argument:

- (1) **[No Nationality in EU]:** If we stay within the EU, there is no national statehood.
- (2) **[No Nationality, No Sovereignty]:** Without national statehood, there is no sovereignty of the people.
- (3) **[No Sovereignty, No Democracy]:** Without sovereignty of the people, there is no democracy.
- (4) **[No Democracy, No People's Will]:** The will of the people will only be realized if democracy is restored.

- (5) **[No People's Will in EU]:** The will of the people will only be realized if we leave the EU.

Premise [No Nationality in EU] and [No Nationality, No Sovereignty] invoke the EU as an enemy of the nation state and popular sovereignty. The AfD says (FE17: 1.1): “The treaties of Schengen, Maastricht and Lisbon illegally intervened in the inviolable popular sovereignty. A state that abandons its border control and thus sovereignty over its territory dissolves. It loses its statehood.”

Premise [No Sovereignty, No Democracy] reads in the AfD’s own words: “Only in national states can people’s sovereignty be lived, the mother and the heart of democracy.” Premise [No Democracy, No People’s Will] is implicit in the title of Chapter 1 of the manifesto for the federal election in 2017: “Restoring Democracy in Germany.” The entire

chapter argues for the claim that democracy must be restored in Germany because otherwise society is doomed.

The argumentation can be reconstructed as depicted in Figure 6.

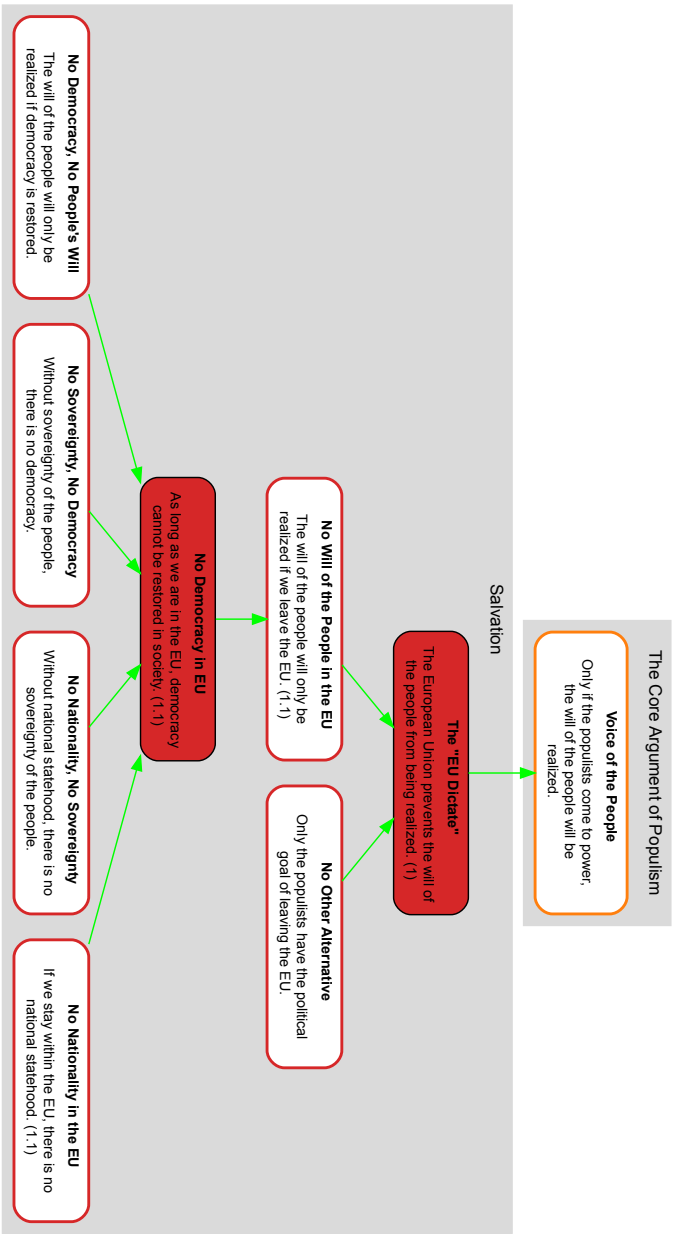


Figure 6: The Salvation Argument on the EU

It is striking how clear the argumentative structure is even on the original texts' surface. In both manifestos and the party platform, the AfD gives

the very same arguments aiming to show that democracy is not possible without popular sovereignty, which in turn is not possible without national statehood, which finally is not possible within the EU. That is why, according to the AfD, Germany (and for that matter any country within the EU) is not a democracy. Due to the “establishment” and “fake news” (who are in cahoots with the “swamp in Brussels”), only the AfD can restore democracy by disbanding the EU or carrying out a “DEXIT.”

5. CONCLUSION

The AfD’s party platform and manifestos do not lack argumentation. Instead, these political programs operate with false or simplistic claims that emotionally appeal to their supporters and at the same time enable stringent argumentation. As a result, it is relatively easy to reconstruct deductively valid arguments where all premises that had to be added are comparatively uncontroversial. In contrast, most explicit premises are verifiably false, misleading, or morally problematic. In the few cases where added premises are more controversial, they implicitly follow from other statements in the AfD’s programs and it can justifiably be assumed that AfD politicians, if confronted with this reconstruction, would accept them. This entails that the argumentation is easily comprehensible and can be reconstructed in a logically valid form without questionable explanations.

In terms of content, it is remarkable that the central theme of “threat by immigration” runs through most of the arguments. It is responsible for problems in the national health system, social injustice, globalization, demographic change, internal security and cultural identity. The EU, the “establishment” and the “Fake News” are the three major obstacles to solving these problems.

The reconstruction allows us to draw a few plausible (even though preliminary) conclusions about the AfD’s argumentative strategy. It seems to consist of three steps. First, the AfD specifically uses emotions such as fear, anger and indignation. These emotions – whether well-founded or not – are addressed in many premises of the argumentation. They bring attention and support by (potential) voters. The fact that many of the premises are verifiably false may (despite what one might think) be ultimately advantageous from the AfD’s point of view, since this facilitates a simple narrative with logically clear arguments. People often seek reasons for their positions rather than adapting them according to the reasons they find. By relying on this, the AfD may also be able to exploit our tribal nature, by which we adhere to the positions of our social group irrespective of their overall plausibility. As a group it may even be rational to adopt false beliefs when it improves its argumentative stance in society by binding it together as a group. A similar effect seems to be

achieved by employing "us versus them"-schemes. Immigration and Islam play a vital role in the debate and populists benefit strongly from antagonizing against immigrants and Muslims. The arguments are designed to exploit people's preexisting opinions – no matter the facts.

In doing so, secondly, the AfD heats up the emotions addressed in the premises. It establishes doomsday scenarios, which rely on prejudices and already existing enemy images. The most detailed and effective way to do this is by means of "us versus them"-schemes (in particular by means of invoking the threat of the "refugee crisis" and "the Islam," but also the enemy from within, i.e., the Greens and left-wingers who are responsible for "fake news," "genderism," "multiculturalism," and the general decay of traditional values).

The imminent doom invoked by this can then be used as justification to, thirdly, present oneself as the only salvation – as the last "Alternative for Germany". While other parties and political agents are portrayed as corrupt, incapable, part of the "EU tyranny", the "fake news," or the "establishment", the AfD can claim to be the "voice of the people" – the only party that can and will save "the people" from this existential threat (which is unacknowledged by everyone else).

In a nutshell, the AfD uses the mobilizing power of fear, anger and indignation to construct a community of "the people," which is threatened by immigrants and Muslims, and can only be saved if the AfD prevails against the "Fake News" and "establishment." The analysis thus suggests a particular interpretation of the AfD's argumentative strategy. Further argumentation theoretically informed research is needed, however, into the argumentative strategies of both other populist and democratic political agents. It would be desirable to compare these findings to alternative reconstructions of the AfD's argumentation to allow stronger conclusions about the strategies employed.

REFERENCES

- Adorno, T. W. (2019). *Aspekte des neuen Rechtsradikalismus. Ein Vortrag* (5th ed.). Berlin: Suhrkamp.
- Albertazzi, D. & McDonnell, D. (2008). *Twenty-first century populism. The spectre of Western European democracy*. New York: Palgrave Macmillan.
- Anderson, B. (2006). *Imagined Communities. Reflections on the Origin and Spread of Nationalism*. London, New York: Verso.
- Barr, R. R. (2009). Populists, outsiders and anti-establishment politics. *Party Politics*, 15, (1), 29–48.
- Bergmann, E. (2018). *Conspiracy & Populism. The Politics of Misinformation*. Cham: Springer International Publishing.
- Betz, G. (2010). *Theorie dialektischer Strukturen* (Philosophische Abhandlungen). Frankfurt am Main: Klostermann.

- Betz, G. & Brun, G. (2016). Analysing Practical Argumentation. In S. O. Hansson, & G. Hirsch Hadorn (Eds.), *The Argumentative Turn in Policy Analysis. Reasoning about Uncertainty*, 39–77. Dordrecht: Springer.
- Betz, H.-G. (2018). The Radical Right and Populism. In J. Rydgren (Ed.), *The Oxford Handbook of the Radical Right*, 86–104. Oxford: Oxford University Press.
- Brun, G. & Hirsch Hadorn, G. (2018). *Textanalyse in den Wissenschaften. Inhalte und Argumente analysieren und verstehen* (3rd ed.). Zürich: vdf Hochschulverlag AG an der ETH Zürich.
- Gramsci, A. (2014). *A Great and Terrible World. The Pre-Prison Letters (1908-1926)*. London: Lawrence & Wishart.
- Greene, J. D. (2014). *Moral tribes. Emotion, reason, and the gap between us and them*. New York: Penguin Books.
- Haidt, J. (2013). *The Righteous Mind. Why Good People Are Divided by Politics and Religion*. London: Penguin.
- Herman, A. (1998). *Propheten des Niedergangs. Der Endzeitmythos im westlichen Denken*. Berlin: Propyläen.
- Kailitz, S. (2004). *Politischer Extremismus in der Bundesrepublik Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kämper, H. (2017). Das Grundsatzprogramm der AfD und seine historischen Parallelen. Eine Perspektive der Politolinguistik. *Sprachreport*, 33, 1–21.
- Laclau, E. (2007). *On Populist Reason*. London: Verso.
- Lanius, D. (2017). Wie argumentieren Rechtspopulisten? Eine Argumentationsanalyse am Beispiel des AfD-Wahlprogramms. *ITZ Diskussionspapiere*, 4, 1–35. doi:10.5445/IR/1000074060
- Lilla, M. (2016). *The shipwrecked mind. On political reaction* (New York review books). New York: New York Review Books.
- Moffitt, B. & Tormey, S. (2014). Rethinking Populism. Politics, Mediatization and Political Style. *Political Studies*, 62, (2), 381–397.
- Mudde, C. & Rovira Kaltwasser, C. (2017). *Populism. A Very Short Introduction*. Oxford: Oxford University Press.
- Müller, J.-W. (2017). *What is Populism?* London: Penguin Random House.
- Pappas, T. S. (2016). Modern Populism. Research Advances, Conceptual and Methodological Pitfalls, and the Minimal Definition. *Oxford Research Encyclopedia of Politics*, 1–27.
- Rooduijn, M. (2019). State of the field: How to study populism and adjacent topics? A plea for both more and less focus. *European Journal of Political Research*, 58, (1), 362–372. doi:10.1111/1475-6765.12314
- Schmitt, C. (2002). *Der Begriff des Politischen. Text von 1932 mit einem Vorwort und drei Corollarien* (7th ed.). Berlin: Duncker und Humblot.
- Taguieff, P.-A. (1995). Political science confronts populism. From a conceptual mirage to a real problem. *Telos*, 103, 9–43.
- Quent, M. (2019). *Deutschland rechts außen: Wie die Rechten nach der Macht greifen und wie wir sie stoppen können*. München: Piper.

A Decision Tree for Annotating Argumentation Scheme Corpora

JOHN LAWRENCE

Centre for Argument Technology, University of Dundee, UK
j.lawrence@dundee.ac.uk

JACKY VISSER

Centre for Argument Technology, University of Dundee, UK
j.visser@dundee.ac.uk

DOUGLAS WALTON

*Centre for Research in Reasoning, Argumentation and Rhetoric, the
University of Windsor, Canada*
dwalton@uwindsor.ca

CHRIS REED

Centre for Argument Technology, University of Dundee, UK
c.a.reed@dundee.ac.uk

Quantitative approaches, necessary for e.g. computational-linguistic methods such as argument mining, require large annotated corpora of argumentative discourse. Publicly available corpora of argumentation schemes often only cover a small selection of example schemes and suffer from low inter-annotator agreement. To address this, we present a heuristic decision tree for the classification of Walton's top-level taxonomy of 60 schemes. An annotation study on 505 arguments resulted in a 97% classification covering 38 schemes (Cohen's κ 0.723).

KEYWORDS: annotation, argument analysis, argumentation schemes, corpora, decision tree, software

1. INTRODUCTION

Data about argumentative practice both informs descriptive approaches to argument and provides a testing ground for normative models. This data can come from the qualitative appraisal of selected examples, but

quantitative approaches, while labour intensive, are gaining traction, motivated by the rise of computational-linguistic methods such as argument mining. Such quantitative approaches require large corpora of argumentative discourse annotated using theories of argumentation.

Argumentation schemes capture the passage of (typically presumptive) inference from a set of premises to a conclusion representing stereotypical patterns of human reasoning. As such, they form a historical descendant of the topics of Aristotle (Aristotle, 1958) and, much like Aristotle’s topics, play a valuable role in both the construction and evaluation of arguments. Various attempts have been made to identify and classify schemes and though these sets of schemes overlap, both their granularity and comprehensiveness vary greatly. As a result, annotated corpora of argumentation schemes tend to contain a selection of examples from only one scheme set, with those based on Douglas Walton’s typology (Walton, 1996) being the most common.

Despite several proposals to systematise the Walton scheme set by imposing some ordering principle on the typology, to our knowledge, no exhaustive and systematic account currently exists. This absence is reflected in the publicly available argumentation scheme corpora, all of which suffer from a combination of either low inter-annotator agreement, or lack of exhaustive coverage, with, in many cases, only those examples that clearly fit a particular pattern annotated.

In the current paper, we present a decision tree for the classification of Walton’s scheme set. Whilst intended primarily as a guide for annotators, the decision tree captures a detailed systematisation of the scheme set, with each of the top-level branches representing divisions into general categories (for example, arguments based on character, or on opinion), before breaking these down further by following a path of simple questions until a definitive scheme classification is reached. To ensure a comprehensive coverage, the decision tree is based on Walton, Reed and Macagno’s 2008 book *Argumentation Schemes*, which describes over 60 schemes.

In order to test the applicability of the decision tree to challenging, real-world data, an annotation study was carried out to classify all occurrences of inference relations in an existing argumentative analysis of the first US presidential election debate between Hillary Clinton and Donald Trump. The annotation resulted in substantial inter-annotator agreement. These results suggest that the application of the decision tree to argumentation scheme annotation constitutes a significant improvement to both reliability and breadth of coverage when compared to previous scheme annotation work.

The rest of the paper is structured as follows. In Section 2, we discuss relevant existing annotations of argumentation schemes. In Section 3, we introduce the decision tree heuristic for annotating

argumentation schemes. In Section 4, we describe and evaluate the annotation study. In Sections 5, we discuss ways of further improving the annotation of argumentation schemes, by considering scheme clusters and a systematisation of the Walton scheme set. In Section 6, we conclude the paper.

2. ANNOTATING ARGUMENTATION SCHEMES

The annotation of argumentation schemes comprises the classification of the inferential relations between premises and conclusions of arguments in accordance with a particular typology. While we start from Walton's typology, alternative approaches are also employed for scheme identification: (Green, 2015) presents ten custom argumentation schemes for genetics research articles, (Musi, Ghosh, & Muresan, 2016) explore annotation guidelines on the basis of the Argumentum Model of Topics (Rigotti & Greco, 2019), and (Visser, Lawrence, Wagemans, & Reed, 2019) annotate argumentation schemes on the basis of the Periodic Table of Arguments (Wagemans, 2016).

Existing annotations on the basis of Walton's typology tend to use a restricted set of scheme types, and struggle to obtain replicable results. For example, (Duschl, 2007) initially adopts a selection of nine argumentation schemes described by (Walton, 1996), for his annotation of transcribed middle-school student interviews about science fair projects. Later, however, he collapses several schemes into four more general classes no longer directly related to particular scheme types. This deviation from Walton's typology appears to be motivated by the need to improve annotation agreement. The validation of the annotation method does not account for chance agreement, by only providing percentage-agreement scores (instead of resorting to, e.g., a κ or α metric). Out of a total of 17 texts, the inter-annotator agreement is reported on two as 90% and 84%, without any further detailing of the sampling method.

Similarly, (Song, Heilman, Beigman Klebanov, & Deane, 2014) base their annotation on a modification of Walton's typology, settling on a restricted set of three more general schemes: policy, causal, and sample – resulting in Cohen's κ scores for inter-annotator agreement ranging from 0.364 to 0.848. (Anthony & Kim, 2015) employ a bespoke set of nine coding labels modified from the categories used by (Duschl, 2007) and nine schemes described in a textbook by (Walton, 2006). They do not measure any inter-annotator agreement, opting for a fully open collaborative annotation without any testing of the reliability of the methods.

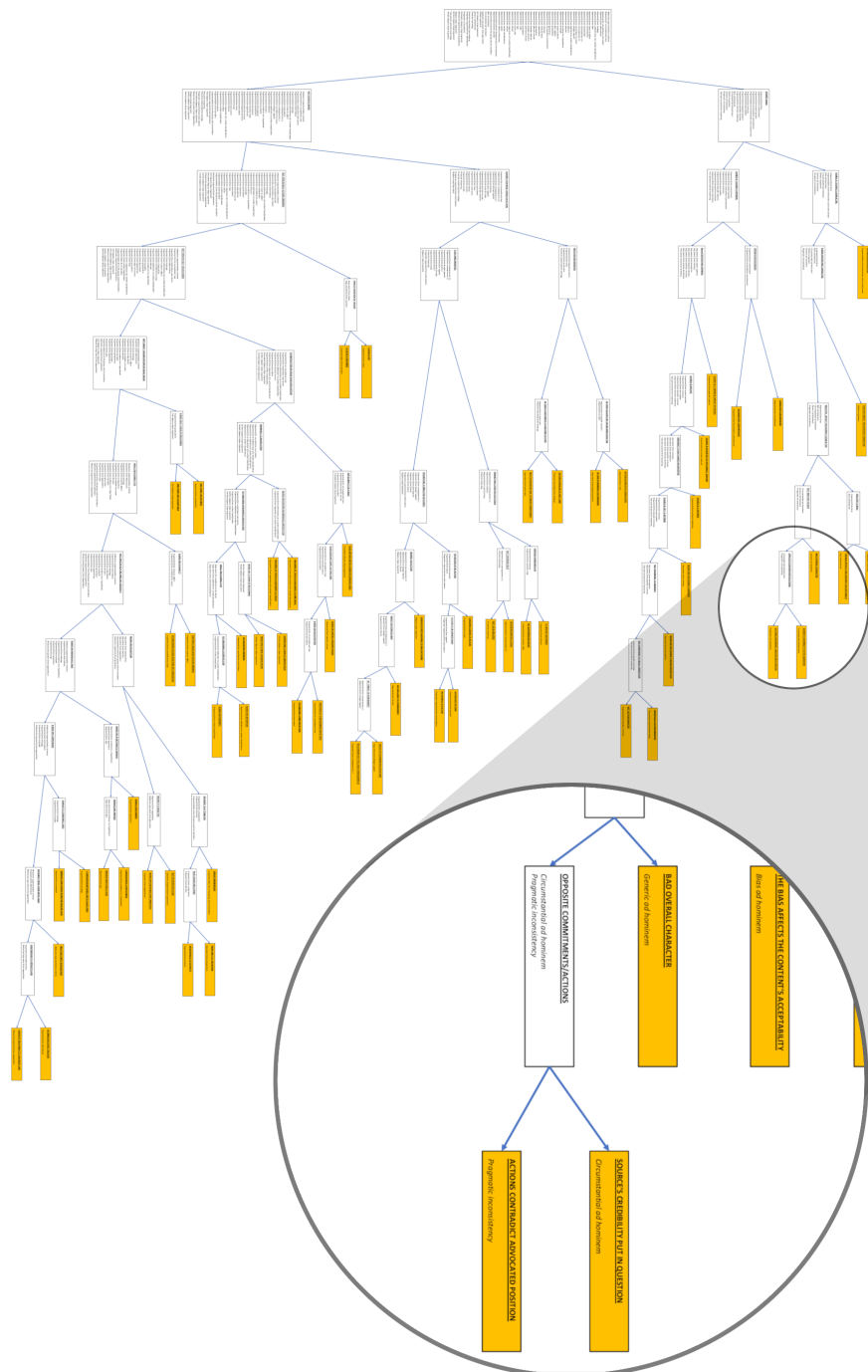


Figure 1 – Decision tree for argumentation scheme annotation

3. ARGUMENTATION SCHEME DECISION TREE

To facilitate the process of annotating the main 60 argumentation schemes described by Walton, Reed and Macagno in their 2008 book *Argumentation Schemes*, we developed a classification decision tree: an indicative heuristic for the annotators, to intuitively support their analytical task. The decision tree interprets the book's Chapter 9 (Walton et al., 2008, pp. 308–346) *A User's Compendium of Schemes* as the starting point of annotation guidelines. The main principle guiding the annotation is the clustering of argumentation schemes on the basis of intuitively clear features recognisable for annotators, resulting in the decision tree of Figure 1. The decision tree constitutes a dichotomous identification tree that leads the analyst through a series of disjunctive choices based on the distinctive features of a 'species' of argumentation scheme to the particular type. Starting from the distinction between source-based and other arguments, each further choice in the tree leads to either a particular argumentation scheme or to a further distinction.

In annotating Example (1), an analyst using the tree follows a sequence of numbered characteristics to identify the argument as an instance of *practical reasoning from analogy*: Argument does not depend on a source's opinion or character; Conclusion is about a course of action; Argument hinges on another motivation for the action [other than its outcome]; Course of action is compared to a similar or alternative action; Action is directly compared to another.

(1) Hillary Clinton: And we finally need to pass a prohibition on anyone who's on the terrorist watch list from being able to buy a gun in our country. If you're too dangerous to fly, you are too dangerous to buy a gun.

Figure 1 (available online at: <http://arg.tech/~john/scheme-tree.png>) visualises the decision procedure, with each leaf representing an argumentation scheme label, and all internal nodes showing clusters of schemes that share particular characteristic properties. For each binary decision point, the tree branches into two, thus leading the annotator from the full set of schemes, through their binary choices, to one (and only one) leaf – i.e. an argumentation scheme classification.

4. ANNOTATION STUDY

In order to test the applicability of the decision tree to challenging, real-world data, an annotation study was carried out to classify all occurrences of inference relations in US2016G1tv (Visser et al., 2019), an existing argumentative analysis of the first election debate (26

September 2016, Hempstead, NY) between Hillary Clinton and Donald Trump. The US2016G1tv corpus (stored in AIFdb (Lawrence et al., 2012), and available online at corpora.aifdb.org/US2016G1tv) is annotated on the basis of Inference Anchoring Theory (IAT) (Reed & Budzynska, 2011), resulting in an Argument Interchange Format (AIF) (Chesñevar et al., 2006) compliant corpus.

Two annotators used the argumentation scheme decision tree to classify 55% of the 505 inferential relations within the corpus; for example classifying Example (1) as an instance of the argumentation scheme *practical reasoning from analogy*. The two annotations resulted in an overlapping sample of 10.2% of the corpus annotated by both annotators. For these annotations a Cohen's κ (Cohen, 1960) of 0.723 was achieved; well within the category of 'substantial agreement' (Landis & Koch, 1977). Some classes of argumentation scheme turned out to be particularly difficult to distinguish; e.g., Example (2) was classified by one annotator as *practical reasoning*, related to promoting goals, and by the other as *argument from values*, related to promoting values.

(2) Hilary Clinton: What I have proposed would be paid for by raising taxes on the wealthy [...] I think it's time that the wealthy and corporations paid their fair share to support this country.

The results of the annotation in accordance with Walton's classification of argumentation schemes are collected in the US2016G1tvWALTON corpus – available online at corpora.aifdb.org/US2016G1tvWALTON. Of the 505 inferences in the original US2016G1tv corpus, a total of 491 are annotated with one of the 60 argumentation scheme types in Walton's classification, leaving only 14 as unclassified *default inference*. The most common scheme, by some margin, is *argument from example*. The *argument from expert opinion* scheme, an often used example, is remarkably rare with only three occurrences. Full results of the annotation and the corpus are discussed by Visser et al. (2018).

5. PRINCIPLES OF ARGUMENTATION SCHEME CLASSIFICATION

The annotation study on the US2016G1tv corpus using the decision tree resulted in substantial inter-annotator agreement. The argumentation schemes in the decision tree are organised according to their distinctive features allowing annotators to distinguish between them. To further improve the decision tree, we aim to take into account the guiding principles underlying the Waltonian taxonomy of argumentation schemes, and the possible clustering of schemes on that basis.

The classification of argumentation schemes should not be regarded as a completed structure, but as a work in progress that is continually being subject to readjustment and refinement as the concepts defining the schemes are formulated in a more precise way and applied to new examples. We will explain the research procedure of improving a classification system of schemes as a process of continuing adjustment between collecting data, sharpening criteria that enable coders to identify a scheme, and used to refine the typology to assist the continuing collection of data.

The 2008 classification system (Walton et al., 2008, pp. 349-350) divided schemes into three general categories, reasoning, source-based arguments and applying rules to cases. Under reasoning five subcategories were distinguished: deductive reasoning, inductive reasoning, practical reasoning, and abductive reasoning. Under the general heading of source-based arguments, four schemes were listed: arguments from position to know, arguments from commitment, arguments attacking personal credibility, and arguments from popular acceptance. The third general category was called applying rules to cases. It had four subcategories: arguments based on cases, defeasible rule-based arguments, verbal classification arguments and chained arguments connecting rules in cases. Each of these second-level types of schemes contained categories at a finer level of granularity. These third level schemes include many of the schemes that are so highly familiar to researchers on argumentation. For example, the third category under source-based arguments contains the following three schemes: argument from allegation of bias, poisoning the well by alleging group bias and *ad hominem* arguments.

The annotators in our annotation study made use of chapter 9, the *user's compendium of schemes* (Walton et al., 2008, pp. 308-346), as guidelines to build a classification decision tree intended to be used as an annotation heuristic. In this heuristic, the top level branches represent divisions and the general categories, while the lower branches break these categories down further by following a path of binary questions. As each question is answered the user is directed down the tree until a definitive scheme classification is arrived at. However, in chapter 10 there was given a proposed classification system for argumentation schemes (Walton et al., 2008, pp. 349-350). One might wonder what the relationship is between this early classification system and the classification decision tree presently offered as an annotation heuristic. One might also wonder whether the 2008 classification system has changed over the ten year interval in the continuing research on schemes classification systems. Finally one might wonder about the current state of this research. This section is designed to answer those questions.

5.1 Clusters of argumentation schemes

It is important to be aware, as stated (Walton et al., 2008, p. 348) that because of the difficulty of defining the concepts that any classification system of schemes has to be based on (such concepts as knowledge, causation, threat, and so forth), any attempt to classify schemes faces conceptual difficulties in adequately defining the contested concepts used at the top levels of the tree structure. For this reason readers were warned that the 2008 system of classifying schemes was to be regarded as a provisional hypothesis that should be subject to improvement as further empirical and analytical work on schemes classification continues. This warning is especially important now, because in the ten year interval the explosion of research on argument mining has raised many fine-grained questions about how particular groups of schemes should be fitted together into the larger picture of any general classification system.

Some subsequent work (Walton & Macagno, 2016) presented a survey of the literature on scheme classification, as well as outlining how the 2008 system needs to be modified in order to accommodate current research in artificial intelligence and computational linguistics on argument mining. In the 2016 paper, it was shown how the procedure of developing and using classification systems can only move forward by combining two approaches. One of these is a top-down approach that begins with concepts formulated at a high level of abstraction, then moves to particular types of schemes that fit under these general categories, and then finally moves to schemes representing the types of arguments we are already so widely familiar with. But at the same time, as research on argument mining continues, it is also necessary to have a bottom-up approach (Walton, 2012) that begins with real examples of arguments at the ground level of cases that distinguish in a very particular way between subtypes of a given scheme. What happens at this bottom-up level is that so-called clusters of schemes are fitted together into larger groups, and then these groups have to be fitted into more general classifications of schemes.

To get a general idea of how clusters of schemes fit into an encompassing schemes classification system, it is useful to examine a graph structure representing a typical example showing how a scheme classification system can be represented visually as a graph. Such a graph is shown in Figure 2. The familiar types of arguments associated with schemes, such as “argument from expert opinion”, are shown in the rectangles with rounded corners. Other categories useful for classifying schemes, such as “source-dependent arguments” are shown in the rectangles with sharp corners.

Two examples of clusters of schemes are indicated by the darkened borders of the rectangles and the arrows in the components of the cluster. The cluster displayed on the right depicts the various kinds of arguments that come under the general category of practical reasoning. Practical reasoning is a distinctive type of argument in its own right, and has its own scheme, but it also has subspecies such as instrumental practical reasoning and value-based practical reasoning. The cluster of schemes under the general heading of *ad hominem* arguments is displayed at the left of Figure 2. This cluster is shown as incomplete. Under the general heading of *ad hominem* arguments some schemes are shown such as the direct *ad hominem* argument type, sometimes called the abusive *ad hominem* in the logic textbooks, and the circumstantial *ad hominem* type. However, as is well known in the literature, there are many other types of *ad hominem* arguments that are not shown here (Walton et al., 2008, p. 352). All that is shown is an elliptical node at the bottom left indicating that there are other *ad hominem* arguments that need to be classified within this cluster. This particular graph is not meant to represent Walton's classification system in a finished or comprehensive form. It is merely an example meant to show what clusters of schemes look like and how the clusters can fit into a more comprehensive classification system.

Note that the graph in Figure 2 is meant to be only a fragment of a larger graph which could include other categories of kinds of arguments that are not defeasible, such as deductive *modus ponens*. Further note that the partial classification system is also incomplete at the bottom level. For example, some types of *ad hominem* arguments are classified at the left of the graph, but the elliptical node, other *ad hominem* arguments, indicates that further sub classifications are possible. For example, in (Walton et al., 2008, p. 352) a graph structure is visually presented that displays seven particular types of *ad hominem* arguments, including the poisoning the well type, the guilt by association type, the *tu quoque* type and the circumstantial *ad hominem* type, subsumed under the more general categories, such as argument from inconsistent commitment and the ethotic or personal type of *ad hominem* argument which is a direct attack on the arguer's character in order to discredit his or her argument.

As shown in Figure 2, argument from precedent combines with the basic slippery slope type of argument to produce a species of slippery slope argument called the precedent type of slippery slope argument. To explain how this works, let us look at the scheme for the basic slippery slope argument.

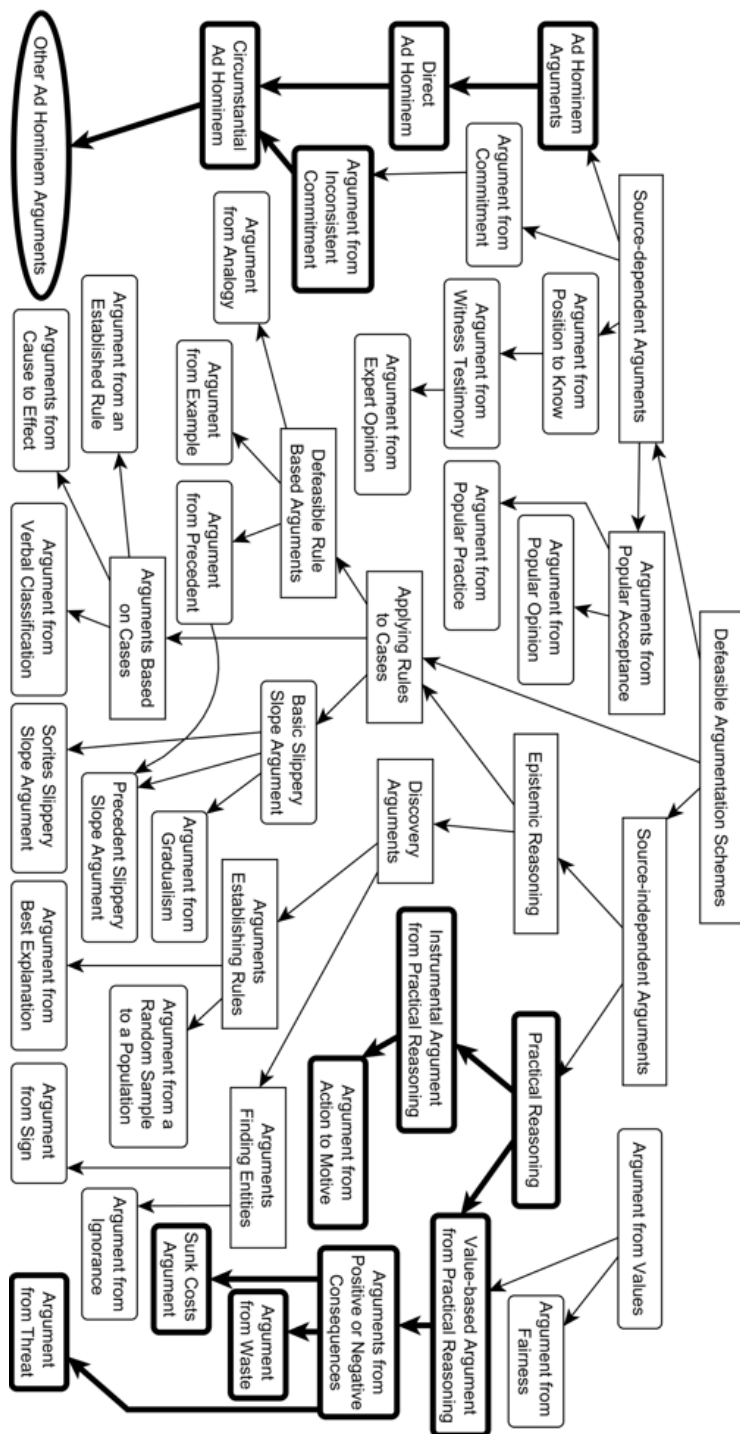


Figure 2 – Graph representing a fragment of a classification system

Initial Premise: An agent α is considering carrying out an action A_0 .

Sequential Premise: Carrying out A_0 would lead to A_1 , which would in turn lead to carrying out A_2 , and so forth, through a sequence $A_2, \dots, A_x, \dots, A_y, \dots, A_n$.

Indeterminacy Premise: There is a sequence $A_0, A_1, A_2, \dots, A_x, \dots, A_y, \dots, A_n$ that contains a subsequence A_x, \dots, A_y called the gray zone where x and y are indeterminate points.

Control Premise: α has control over whether to stop carrying out the actions in the sequence until α reaches some indeterminate point in the gray zone A_x, \dots, A_y .

Loss of Control Premise: Once α reaches the indeterminate point in the gray zone A_x, \dots, A_y , α will lose control and will be compelled to keep carrying out actions until she reaches A_n .

Catastrophic Outcome Premise: A_n is a catastrophic outcome that should be avoided if possible.

Conclusion: A_0 should not be brought about (Walton, 2015, p. 288).

There are various types of slippery slope argument that can be built by extending the basic scheme, and one of these, the precedent type of slippery slope argument, which generates a sequence whereby one case is a precedent for a second one, and the second one is a precedent for a third one, and so forth. As shown in Figure 2, the precedent slippery slope argument combines argument from precedent with the basic slippery slope type of argument. This means that the precedent slippery slope argument is formed as a cluster from other types of arguments, as shown in Figure 3.

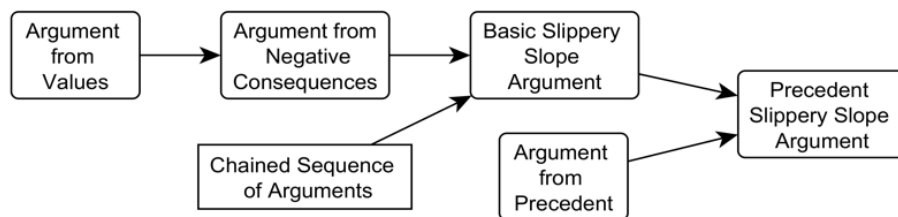


Figure 3 – A slippery slope cluster

Being aware of how this cluster of arguments is formed is helpful for enabling annotators to distinguish between a precedent slippery slope argument and a run-of-the-mill argument from precedent that should not be classified as a slippery slope argument.

5.2 Identification conditions of argumentation schemes

A central practical problem inherent in existing corpus-linguistic work on argumentation schemes is that the annotators lack enough specific guidance on how to decide whether an argument found in a real natural language text can properly be said to fit a particular scheme or not. An early study which used schemes to classify kinds of arguments put forward by candidates in a provincial election in Canada (Hansen & Walton, 2013) classified 256 arguments using 14 schemes and a category called “none of the above”. A group of six annotators, two of them experts in argumentation theory, collected arguments found in newspaper articles reporting arguments commenting on issues being debated in the campaign. The difficulty they encountered was that the four non-expert annotators, as well as the two experts in some instances, found it difficult to classify arguments in some instances because of the open texture of key terms used in the schemes. For example, annotators found it difficult to differentiate between a circumstantial *ad hominem* argument and an argument from inconsistency, a species of argument from commitment. Both kinds of arguments allege that an opposed arguer has put forward an argument, or part of an argument, that conflicts with a prior commitment of the arguer. But only the circumstantial *ad hominem* type of argument uses the commitment to derive a secondary inference to the conclusion that the arguer has exhibited some ethical defect of character, indicated by the use of a keyword, such as ‘hypocrite’ or ‘liar’.

The solution to this problem recommended in (Walton, 2012) was to devise a set of so-called identification conditions that can be used to offer annotators additional guidance on whether a particular scheme fits a particular case or not. There were 24 of these identification conditions formulated by Walton (2012, pp. 49-56). A current project is to refine these conditions to make them more precise and easier to use. To give the reader an idea of what these kinds of conditions are like, here are six of the reformulated ones.

(IC1) Argument from Inconsistent Commitments: (1) There has to be evidence from the way a has put A forward as a claim (assertion) in a dialogue exchange to indicate that a is committed to A, and (2) there has to be evidence from the way a has put A forward as a claim (assertion) in a dialogue exchange or the circumstances of a case to indicate that a is committed to not-A. (4) The conclusion is drawn on the basis of (1) and (2) that a is committed to (A and not-A).

- (IC2) Direct *Ad Hominem* Argument: there has to be (1) not only an attack on the arguer's ethical character (ethos), but (2) this attack has to be used to discredit the arguer's credibility (personal trustworthiness as a source), (3) in order to try to defeat his argument.
- (IC3) Circumstantial *Ad Hominem* Argument: there has to be (1) an attack on the arguer's ethical character, but (2) this attack has to be based on an alleged inconsistency among the arguer's commitments (3) which has to be used to discredit the arguer's credibility (personal trustworthiness as a source), and (4) the premises (1), (2) and (3) have to be put forward to try to defeat his argument.
- (IC4) Argument from Values: (1) The audience to whom the argument is addressed is thought by the arguer to hold a positive (or negative) value with respect to a proposition, and (2) appeal to this value is used by the arguer as a means of supporting (or attacking) the commitment of the audience to some goal or policy he advocates.
- (IC5) Argument from Positive Consequences: (1) A proposal in favour of carrying out action A is put forward, (2) pro A and con A arguments are being considered, (3) the claim is made that A, if carried out, will have positive consequences, (4) use of the term 'positive' means that the action is claimed to have positive value for the audience the argument is directed to, and (5) on this basis it is claimed that the audience should support the proposal to carry out A.
- (IC6) Argument from Negative Consequences: (1) A proposal against carrying out action A is put forward, (2) pro A and con A arguments are being considered, (3) the claim is made that A, if carried out, will have negative consequences, (4) use of the term 'negative' means that the action is claimed to have negative value for the audience the argument is directed to, and (5) on this basis it is claimed that the audience should decline to support the proposal to carry out A.

The other identification conditions have the same general format, except that some of them are more complex forms of argument that contain simpler forms of argument, such as the seven types represented above. One scheme can be shown to contain another scheme, using the identification conditions, and the structure of this relationship can be visually displayed as a graph. By this means, for example, a complex scheme, such as the slippery slope argument, can be shown to contain another simpler form of argument, such as argument from negative consequences, embedded within its structure.

5.3 Identification conditions applied to clusters

This vagueness and ambiguity about how to more precisely define these three types of arguments pervaded Hansen and Walton's (2013) election project because we found numerous examples of argument from inconsistent commitments, some of them arguably being *ad hominem* arguments, some arguably not. A first step toward resolving the problem is to treat the circumstantial *ad hominem* argument (CA) as being a subspecies of the wider category of argument from inconsistent commitments (IC), which is in turn a subspecies of argument from commitment (AC) and the direct *ad hominem* argument (DA) – see Figure 4.

This way it doesn't matter so much if there was a difficulty in trying to make a decision based on the text of discourse of the example on which category the argument should be placed into. If there is doubt whether the argument is really meant to be a personal attack, then we can classify it into the more general category of argument from inconsistent commitment. If there is no doubt, we can classify it into the circumstantial *ad hominem* category.

To cite another example, elements (5) and (6) of the scheme for argument from values (AV) are carried over and incorporated into the structure for argument from value-based practical reasoning (VBPR) – see Figure 5. Similarly, on the left, elements of the scheme for instrumental practical reasoning (IPR) are incorporated into the scheme for value-based practical reasoning (VBPR).

Using this approach, the structure that holds a cluster together is derived from the identification conditions for a particular scheme, showing how that scheme is related to other neighbouring schemes to form a cluster.

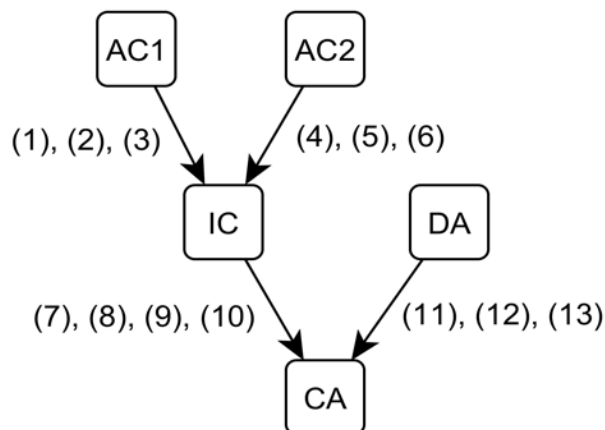


Figure 4 – Part of the *ad hominem* cluster

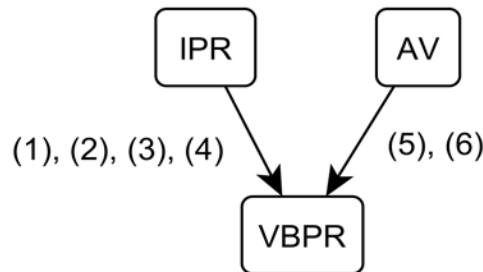


Figure 5 – The value-based practical reasoning cluster

5.4 Leveraging clusters in argumentation scheme classification

Once this method is used to form several clusters of schemes, the clusters can be all put together into a larger classification graph of the kind illustrated in Figure 2. But it is not certain what will happen when this is done on a very large scale. It may be that there are connections between two different clusters in a classification graph, or even connections between several different clusters. At the state this kind of research has reached now, none of this has yet been explored. As more and more examples of arguments fitting a given structure are collected and classified, the clusters can be expected to grow in complexity. For instance, as more and more examples of *ad hominem* arguments are collected and analysed, new types of *ad hominem* arguments are likely to be discovered. Part of this line of discovery will be the formulation of identification conditions for each of these new schemes.

This procedure is circular in nature, but in this instance the circularity is not evidence that a fallacy has been committed. It shows that the activity of collecting data from natural language corpora, and using that to refine the classification system, is a defeasible but scientific way of collecting evidence for or against a hypothesis and improving it by feedback. The recursive application of the procedure improves the accuracy of the formulation of the schemes.

It is shown how the production of an evolving taxonomy that takes increasing sophistication of sub-schemes into account through the use of identification conditions. Essentially the procedure consists in the refinement and evolution of the taxonomy as it is tested against the data by being continuously applied to real examples of naturally occurring arguments.

At the higher levels, the identification conditions can be used to sharpen the general concepts, providing precise and definitions of these terms. By this means, both tools, the identification conditions and the clusters they generate, can be applied to real examples in order to

improve a given classification system for schemes, making it both more precise and more applicable to identifying types of arguments found in discourse.

6. CONCLUSION

Theory-driven applications of computational models of argument, and empirically oriented work alike, rely on data about the actual use of argumentation in practice. The availability of large, reliable, and representative datasets of argumentation scheme usage is essential both to the empirical study of such schemes, and to the development of automated classifiers and argument mining techniques (Budzynska & Villata, 2017). In this paper, we present a decision tree heuristic for annotating argumentation schemes. The decision tree supports annotation which is both comprehensive in the range of schemes it covers, and reliable in the results obtained. Finally, we have considered future directions for improving the decision tree, taking into account the guiding principles underlying the Waltonian taxonomy of argumentation schemes, and the possible clustering of schemes on that basis.

ACKNOWLEDGEMENTS: This research was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) in the United Kingdom under grant EP/N014871/1, and in part by Social Sciences and Humanities Research Council of Canada Insight Grant 435-2012-0104.

REFERENCES

- Anthony, R., & Kim, M. (2015). Challenges and remedies for identifying and classifying argumentation schemes. *Argumentation*, 29(1), 81–113.
- Aristotle. (1958). *Topics* (W. D. Ross, Trans.). Oxford University Press.
- Budzynska, K., & Villata, S. (2017). Processing argumentation in natural language texts. In P. Baroni, D. Gabbay, M. Giacomin, & L. van der Torre (Eds.), *Handbook of formal argumentation*. College Publications.
- Chesñevar, C., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., ... others (2006). Towards an argument interchange format. *The Knowledge Engineering Review*, 21(04), 293–316.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Duschl, R. A. (2007). Quality argumentation and epistemic criteria. In *Argumentation in science education* (pp. 159–175). Springer.

- Gordon, T. F., Prakken, H., & Walton, D. (2007). The carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10), 875–896.
- Green, N. (2015, June). Identifying argumentation schemes in genetics research articles. In *Proceedings of the 2nd workshop on argumentation mining* (pp. 12–21). Denver, CO: ACL.
- Hansen, H. V., & Walton, D. (2013). Argument kinds and argument roles in the Ontario provincial election. *Journal of Argumentation in Context*, 2(2), 226–258.
- Janier, M., Lawrence, J., & Reed, C. (2014). OVA+: An argument analysis interface. In S. Parsons, N. Oren, C. Reed, & F. Cerutti (Eds.), *Proceedings of the fifth international conference on computational models of argument (comma 2014)* (pp. 463–464). Pitlochry: IOS Press.
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 3, 159–174.
- Lawrence, J., Bex, F., Reed, C., & Snaith, M. (2012). AIFdb: Infrastructure for the argument web. In *Proceedings of the fourth international conference on computational models of argument (comma 2012)* (pp. 515–516). Vienna: IOS Press.
- Musi, E., Ghosh, D., & Muresan, S. (2016, August). Towards feasible guidelines for the annotation of argument schemes. In *Proceedings of the 3rd workshop on argumentation mining*. Berlin: ACL.
- Reed, C., & Budzynska, K. (2011). How dialogues create arguments. In F. H. van Eemeren, B. Garssen, D. Godden, & G. Mitchell (Eds.), *Proceedings of the 7th conference of the international society for the study of argumentation (issa)* (pp. 1633–1645). SicSat.
- Reed, C., & Rowe, G. (2004). Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(4), 961–980.
- Rigotti, E., & Greco, S. (2019). *Inference in argumentation: A topics-based approach to argument schemes*. Springer.
- Song, Y., Heilman, M., Beigman Klebanov, B., & Deane, P. (2014). Applying argumentation schemes for essay scoring. In *Proceedings of the first workshop on argumentation mining* (pp. 69–78). Association for Computational Linguistics. doi: 10.3115/v1/W14-2110
- van Gelder, T. (2007). The rationale for rationale. *Law, probability and risk*, 6(1–4), 23–42.
- Visser, J., Konat, B., Duthie, R., Koszowy, M., Budzynska, K., & Reed, C. (2019). Argumentation in the 2016 US presidential elections: annotated corpora of television debates and social media reaction. *Language Resources and Evaluation*. doi: 10.1007/s10579-019-09446-8
- Visser, J., Lawrence, J., Wagemans, J., & Reed, C. (2018). Revisiting Computational Models of Argument Schemes: Classification, Annotation, Comparison. In S. Modgil, K. Budzynska, & J. Lawrence (Eds.), *Proceedings of the 7th international conference on computational models of argument (comma 2018)*. Warsaw, Poland: IOS Press. doi: <https://doi.org/10.3233/978-1-61499-906-5-313>
- Visser, J., Lawrence, J., Wagemans, J., & Reed, C. (2019). An annotated corpus of argument schemes in us election debates. In B. Garssen, D. Godden, G.

- Mitchell, & J. Wagemans (Eds.), *Proceedings of the 9th conference of the international society for the study of argumentation (issa)* (pp. 1101–1111). Amsterdam, The Netherlands: Sic Sat.
- Wagemans, J. H. M. (2016). Constructing a Periodic Table of Arguments. In P. Bondy & L. Benacquista (Eds.), *Argumentation, objectivity, and bias: Proceedings of the 11th international conference of the Ontario society for the study of argumentation (ossa)* (pp. 1–12). OSSA.
- Walton, D. (1996). *Argumentation schemes for presumptive reasoning*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Walton, D. (2006). *Fundamentals of critical argumentation*. Cambridge University Press.
- Walton, D. (2012). Using argumentation schemes for argument extraction: A bottom-up method. *International Journal of Cognitive Informatics and Natural Intelligence*, 6(3), 33–61.
- Walton, D. (2015). The basic slippery slope argument. *Informal Logic*, 35(3), 273–311.
- Walton, D., & Macagno, F. (2016). A classification system for argumentation schemes. *Argument and Computation*, 6(3), 219–245.
- Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. Cambridge University Press.

Philosophical Foundation of Reasonableness in Mencius's Argumentative Discourse: Based on the use of dissociation

YAN LINQIONG

*Department of Philosophy /Institute of Logic and Cognition, Sun
Yat-sen University
School of Foreign Languages, Jiangsu University
yanlq3@mail2.sysu.edu.cn*

XIONG MINGHUI

*Department of Philosophy /Institute of Logic and Cognition, Sun
Yat-sen University Guangzhou China
Institute of Reasoning, Argumentation and Communication,
Southwestern University of Finance and Economics Chengdu China
hssxmh@mail.sysu.edu.cn*

Mencius was known as “being fond of argumentation”. The philosophical foundation of reasonableness in Mencius's argumentative discourse is analysed by resorting to the pragma-dialectical model of critical discussion where dissociation appears with different argumentative functions. The analysis reveals that reasonableness is originated in goodness in human nature, which is embodied as humaneness and righteousness respectively, and which is reflected in holding to the Mean that is based on principle and allows for expediency.

KEYWORDS: Reasonableness, Mencius, Argumentative discourse, Dissociation, Humaneness, Righteousness, Holding to the Mean

1. INTRODUCTION

Mencius (372 BC - 289 BC), the “second sage” after Confucius in the school of Confucianism, was one of the reputed public intellectuals of “Hundred Schools of Thought” in the Warring States period of ancient China. He was known as “being fond of argumentation” in his times, as was recorded in Book 3B9 (See in this paper all the quotes of Mencius's

discourse in APPENDIX) of his work *Mencius*. In response, Mencius gave his reasons for argumentation. The numerous researches on the work *Mencius* also conclude that Mencius is well recognized as being good at argumentation, too. Then a question must have come to our minds: Since Mencius was known as being fond of and good at argumentation, and he seemingly had noble reasons for argumentation, what is the philosophical foundation of reasonableness in his argumentative discourse?

To elaborate the philosophical foundation of reasonableness in Mencius's argumentative discourse, examples with the argumentative technique of dissociation will be analyzed with the use of the theoretical model of critical discussion in pragma-dialectics, especially the four discussion stages at which dissociation may appear, namely, the confrontation stage (establishing the standpoint and the difference of opinion), the opening stage (clarifying the parties concerned and common grounds), the argumentation stage (putting forward argumentation and critical responses) and the concluding stage (getting a result of the discussion) (van Eemeren and Grootendorst 1992, p. 35). Section 2 will probe into the moral metaphysical foundation of reasonableness through analyzing examples of dissociation in Mencius's argumentative discourse on human nature. Following the moral metaphysical foundation of reasonableness, Section 3 will discuss examples of dissociation which may lay bare the embodiments of the metaphysical foundation, namely, humaneness and righteousness, a combination of Confucian virtue ethic and Confucian deontic ethic. Section 4 will further analyze examples of dissociation to see how Mencius put humaneness and righteousness into practice in terms of Confucian practical ethic. Section 5 will conclude the paper.

2. REASONABLENESS ORIGINATED IN CONFUCIAN MORAL METAPHYSICS – HUMAN NATURE BEING GOOD

In Mencius's times, an era of "Hundred Schools of Thought", discussion about human nature was not just heated but also indispensable, for different views about human nature led directly to different claims about moral and political life. The term "nature" or "human nature" has been acknowledged to be one of the key terms in the "common discourse" in the Warring States period (Schwartz 1985, p. 174). More specifically, views on human nature constitute the philosophical bases of the different academic schools. In Book 6A6 of the work *Mencius*, Gongduzi quoted four different views about human nature. Mencius was the first Confucian who proposed that human nature is good in Book 3A1 and argued for it with quite a few passages in the work *Mencius*. According to Mencius, human nature being good should be more precisely elaborated as

inclination to goodness that is inherent in human nature, which is also the metaphysical ground of Confucian ethics (Xiao 2004, pp. 234-235; Yang 2017, p. 87). Such metaphysically philosophical foundation of reasonableness can be expounded in the following examples using dissociation.

Dissociation is an argumentative technique used to separate an original concept into two new ones, whose prototype is considered to be the philosophical pair of “appearance-reality”, with Term I representing the “appearance” level and Term II the “reality” level, while “Term I corresponds to the apparent, to what occurs in the first instance, to what is actual, immediate, and known directly”, and “Term II provides a criterion, a norm which allows us to distinguish those aspects of term I which are of value from those which are not” (Perelman and Olbrechts-Tyteca 1969, pp. 415-416). Agnes van Rees (2009) systematically analyzed and evaluated the use of dissociation at the four discussion stages in the pragma-dialectical model of critical discussion, whose work can thus set as an example for the analysis of dissociation in Mencius’s argumentative discourse.

According to the work *Mencius*, Mencius argued mainly with Gaozi in an explicit manner on human nature, where Mencius argued that human nature is good, while Gaozi held that human nature is neither good nor non-good. For example, in Book 6A3, Gaozi explicitly claimed that that which is inborn is what is meant by “nature”, including human nature, while Mencius disagreed with him by putting as an example “white” as an abstract attribute opposed to the specific “white-colored” things. Book 6A4 further clarifies what Gaozi meant to be that which is inborn - appetite for food and sex.

Referring to Book 4B19, where Mencius pointed out on the one hand that “human beings differ from the birds and beasts (in nature)”, and on the other hand also emphasized that such difference is “slight”. This may imply that Mencius did not deny that the physiological aspect like appetite for food and sex is also part of human nature but in the meantime indicated that it is just not the total of human nature.

Apart from the physiological part, the rest part in human nature that differentiates human beings from birds and beasts is indicated in Book 3A1 and Book 6A6. Book 3A1 mentions for the first time that Mencius held that human nature is good, and in Book 6A6, Gongduzi, one of Mencius’s disciples, quoted the then prevailing views of human nature and asked Mencius straightforwardly what Mencius meant when saying human nature is good. In other words, Mencius dissociated human nature into the physiological part like appetite for food and sex (the “appearance” level) and the moral part of being good that distinguishes from birds and beasts (the “reality” level). So, the dissociation of human

nature here is used at the confrontation stage (**bolded**) of the discussion about human nature.

Mencius's dissociating human nature into the physiological part (the "appearance" level) (Book 6A3, Book 6A4) and the moral part (the "reality" level) (Book 4B19, Book 3A1 and Book 6A6) paves the way for expounding and argumentation on all his philosophical, ethical and political ideas. He especially highlighted the moral part in human nature, that is, the inclination to goodness that is inherent in human nature, which functions as a moral metaphysical source for reasonableness in his argumentative discourse. Such a moral metaphysical conception of reasonableness can explain the origin of humaneness (*Ren*), the core virtue put forward by Confucius, and of righteousness (*Yi*), the core deontic concept that was expanded by Mencius on the basis of the concept humaneness, while humaneness and righteousness are the internal embodiments of the moral goodness in human nature, as is illustrated in Book 4B19 quoted above. In Book 4B19, after pointing out that there is but slight difference between human beings and the birds and beasts, Mencius offered the example of the noble person Shun to show that Shun's noble actions came from his following humaneness and righteousness inherent in him. Here Mencius dissociated "doing humaneness and righteousness" into "following humaneness and righteousness inherent in him" (the "reality" level) and "just performing acts of humaneness and righteousness" (the "appearance" level). This dissociation belongs to the argumentation stage (underlined) of the discussion about human nature.

3. REASONABLENESS EMBODIED AS HUMANENESS AND RIGHTEOUSNESS – A COMBINATION OF CONFUCIAN VIRTUE ETHIC AND CONFUCIAN DEONTIC ETHIC

Section 2 elaborates the moral metaphysical source of reasonableness in Confucianism, that is, the inclination to goodness that is inherent in human nature, which was put forward and expounded by Mencius. Such a moral metaphysical perspective on reasonableness in Mencius's argumentative discourse is embodied as humaneness, which includes the feeling of pity and commiseration and the feeling of approving and disapproving, and righteousness, which includes the feeling of shame and dislike and the feeling of respectfulness and reverence (See Book 2A6, Book 6A6 and Book 7A15), a combination of the Confucian virtue ethic and the Confucian deontic ethic. The following examples of dissociation in Mencius's kingcraft politics will be quoted to illustrate this point.

For example, in Book 2A3, according to a king's motives in performing humaneness, Mencius first differentiated a hegemon and a true king by dissociating "the act of performing humaneness" into

pretending to be humane by force (the “appearance” level) and practicing humaneness out of Virtue (the “reality” level). This dissociation functions as setting a common ground between Mencius and his potential audience (including the kings and dukes with whom he talked, like King Hui of Liang, King Xuan of Qi, Duke Wen of Teng and so on) by definition. So, this dissociation belongs to the opening stage (*italicized*) of the discussion about kingcraft.

Mencius further dissociated “people’s submission” into “people’s submission under force” (the “appearance” level) and “people’s submission out of Virtue” (the “reality” level) (Book 2A3). By quoting the example of the seventy disciples all submitting to Confucius and the ode taken from *Book of Songs*¹, we can see that what Mencius wanted to highlight is the “reality” levels in the two dissociations, thus forming his implicit standpoint about kingcraft: when a true king practices humaneness out of Virtue, people will submit to him sincerely. So, this second dissociation belongs to the argumentation stage (underlined).

Book 1B8 is another example with the use of dissociation that centers about humaneness and righteousness and is related with kingcraft. Between King Xuan of Qi and Mencius, after with the common acknowledgements of the previous ministers like Tang and Ji Fa (later King Wu of Zhou) banishing or assaulting the previous kings like King Jie of Xia and King Zhou of Shang, King Xuan of Qi raised the question whether it is allowed for a minister to slay a ruler. The doubt in King Xuan of Qi indicates the difference of opinion between him and Mencius about the issue. In reply, Mencius first explicated by definition the natures of those rulers who offend against humaneness and righteousness, as brigands, outlaws or outcasts. He then dissociated “the act of killing a ruler” into “slaying a ruler” (the “appearance” level) and “punishing an outcast who offended against humaneness and righteousness” (the “reality” level) (Book 1B8). Such a dissociation appears at the argumentation stage (underlined) of the discussion about how to view the act of killing rulers.

Mencius’s argumentation by dissociation in Book 1B8 shows that he treated as reasonable the killing of those rulers who were determined as outcasts, which is the reality level of the dissociation. Such conception of reasonableness is founded on the criteria of whether a ruler goes for or against humaneness and righteousness. So, in this discussion, Mencius’s implicit standpoint is that rulers who offended against humaneness and righteousness deserved to be slayed and overthrown by

¹ *Book of Songs*, also called *Classic of Poetry* or *Odes* (*Shi Jing*), is one of the five Classics of ancient Chinese literature. It is said to have been compiled by Confucius and is the oldest existing collection of Chinese poetry.

their ministers. Hence this idea is also part of Mencius's kingcraft politics, only narrated in a negative manner.

Book 1B3 begins with the question posed by King Xuan of Qi to Mencius about the means of dealing with relations with neighboring states. In reply, Mencius emphasized the importance of a king to be humane in such diplomatic affairs, which is also part of Mencius's thought of kingcraft and which seemingly earned King Xuan of Qi's agreement according to his exclamation "How great are these words". Then King Xuan of Qi confessed his failing of being fond of valor, which implies the difference of opinion between him and Mencius about whether a king being fond of valor can be a humane king. Mencius first took the concession of King Xuan of Qi's being fond of valor, but then differentiated the valor of an ordinary man in confronting just one person and the valor of King Wen of Zhou and King Wu of Zhou in confronting evil rules and bringing peace to all the people in the world. So, here Mencius dissociated "valor" into "small valor" in terms of an ordinary man confronting just one person for the sake of his personal benefit (the "appearance" level) and "big valor" in terms of confronting evil rulers for the benefit of all the people in the world (the "reality" level) (Book 1B3). The dissociation here belongs to the argumentation stage (underlined) of the discussion about the standpoint explicitly expressed in the concluding stage (**CAPITALIZED**): that King Xuan of Qi's being fond of valor should bring peace to all the people in the world. Mencius's standpoint in terms of valor indicates he favored the "big valor" at the reality level.

Conversations between King Xuan of Qi and Mencius continued in Book 1B5, where King Xuan of Qi confessed another two failings - being fond of wealth and being fond of women, and Mencius again adopted the argumentative technique of dissociation. In Book 1B5, following King Xuan of Qi's question about whether to demolish the Hall of Light, Mencius once again drew King Xuan of Qi's attention to his thought of kingcraft by connecting the Hall of Light with true kingly government. With the same people-oriented ideas expressed in the dissociation of valor, Mencius dissociated "one's fondness of wealth" into "enjoying one's fondness of wealth by oneself" (the "appearance" level) and "sharing one's fondness of wealth with the people" (the "reality" level), and dissociated "one's being fond of women" into "enjoying one's fondness of women by oneself" (the "appearance" level) and "sharing one's fondness of women with the people" (the "reality" level) (Book 1B5). The two dissociations here belong to the argumentation stages (underlined) of the two discussions about the same topic, namely, how to enforce true kingly government or how to become a true king. The corresponding two standpoints are explicitly stated at the concluding stages (**CAPITALIZED**): that if a king shares his fondness of wealth with the people, he can become a true king, and that if a king shares his fondness of women with

the people, he can become a true king. The explicit standpoints put forward by Mencius indicate again his preference for the “reality” levels of the two dissociations.

4. REASONABLENESS REFLECTED IN HOLDING TO THE MEAN THAT IS BASED ON PRINCIPLE AND ALLOWS FOR EXPEDIENCY - A PERSPECTIVE OF CONFUCIAN PRACTICAL ETHICS

Discussions in Section 3 indicate that Mencius regarded humaneness and righteousness as the guiding principle for his kingcraft politics. In the meantime, Mencius made some concessions in convincing his target audience, like King Xuan of Qi, to adopt his kingcraft claim by holding to the principle of humaneness and righteousness. For example, in Book 1B3 and Book 1B5, Mencius did not deny King Xuan of Qi’s fondness of valor, wealth and women. Instead, Mencius dissociated the said fondness into “fondness of valor, wealth and women by oneself” (the “appearance” level) and “fondness of valor, wealth and women together with the people” (the “reality” level), and encouraged King Xuan of Qi to extend his personal fondness to his people. Such concessions set a common ground between Mencius and King Xuan of Qi for Mencius’s subsequent argumentation on his claim of kingcraft, as is analyzed in Section 3. They also imply that in argumentation for his kingcraft claim that is founded on the principle of humaneness and righteousness, Mencius took the strategy of conciliation - adopting the other party’s arguments for defending one’s own standpoint (van Eemeren 2009, p. 13; 2010, p. 165) by means of expediency, while adopting expediency serves in the end the purpose of holding to the principle of humaneness and righteousness. This idea can be summarized as holding to the Mean that is based on principle and allows for expediency (Ding 2004, p. 192; Xu 2004, pp. 589-590, 593-594), which is the conception of reasonableness reflected in Confucian practical ethics. Examples will be analyzed below to elaborate the reflection of reasonableness in Mencius’s argumentative discourse.

Mencius expressed his preference for holding to the Mean several times in his work *Mencius*, for example, in Book 4B20, Book 7A26, Book 7A41 and Book 7B37.

In Book 4B20, Mencius quoted Tang, King Wen of Zhou, King Wu of Zhou, and the Duke of Zhou, who all “held fast to the Mean”, while these quoted persons are all moral models admired by Mencius and other Confucians. In Book 7B37, When Wan Zhang, one of Mencius’s disciples, asked Mencius why Confucius, being in the state of Chen, was still thinking about the mad scholars of the state of Lu, Mencius pointed out that what Confucius really preferred was those scholars who “followed the middle way”. “Holding fast to the Mean” and “following the middle way” are synonyms, meaning not going to extremes.

However, according to Mencius, one's holding to the Mean does not mean that one can do without holding to the principle. In Book 7A41, Gongsun Chou, another one of Mencius's disciples, acknowledged on the one hand that the Way or Dao that Mencius claimed and tried to promote, namely, the moral idealist principle of humaneness and righteousness, is very much lofty and beautiful, but on the other hand pointed out that it is much too difficult for ordinary people to attain it. So, Gongsun Chou asked his Master Mencius why not make the Way or Dao more easily attained by ordinary people in their daily life, implying that the Way or Dao can be more attainable with expediencies considered. Gongsun Chou's question indicates the difference of opinion between him and Mencius about whether the Way or Dao (the principle) can be compromised to be more attainable for ordinary people. In response, by resorting to examples of the great artisan and the proficient archer Yi, Mencius argued implicitly that the Way or Dao (the principle) should not be compromised to adapt to others (holding to the principle) but instead should be held fast to by positioning oneself "at the center of the Way" (or holding to the mean). So, here Mencius coordinated holding to the principle with holding to the mean.

Although Mencius argued that one should hold to the principle while holding to the mean, he did not neglect the changing circumstances where holding to the principle may confront in practice. In 7A26, Mencius first offered Yangzi's and Mozi's examples of choosing extremes - egoism and impartial care respectively. Then he gave Zimo's example of holding to the Mean. Compared with the two extremes, Mencius took a positive attitude to holding to the Mean, but immediately added that in holding to the Mean exigencies (or expediencies) should be allowed for; otherwise, holding to the Mean would resemble holding to one point (or extreme).

Mencius's idea of holding to the Mean that is based on principle but allows for expediency can be expounded with examples using the argumentative technique of dissociation. For example, in Book 1A1, which is well known as Mencius's debate on righteousness and profit in Chinese history, Mencius came to the kingdom of Wei to meet its ruler King Hui of Liang. In their first conversation, King Hui of Liang asked Mencius straightforward whether Mencius brought with him some means to profit the kingdom of Wei. From Mencius's rhetorical question "Why must the king speak of profit?", we can see that the difference of opinion between Mencius and King Hui of Liang is whether a king should speak of profit (the confrontation stage, **bolded**). In reply to King Hui of Liang's question, Mencius stated that he had only humaneness and righteousness with him, and then listed by reasoning the consequences of speaking of profit and the consequences of putting profit-pursuing before righteousness pursuit, from the king to his officers till the gentlemen and the common people (the argumentation stage,

underlined). At the end of this conversation, Mencius restated his standpoint that a king should only speak of humaneness and righteousness instead of profit (the concluding stage, CAPITALIZED).

The expressly formulated standpoint in Book 1A1 seems to indicate that Mencius is against a king's pursuit of profit. However, looking more closely at Book 1A1, we may notice that what Mencius really disapproved of is the pursuit of profiting just "our state", "our house" or "myself", which are all privately-cantered. In Book 1B3 and Book 1B5 (quoted in Section 3), Mencius did not deny King Xuan of Qi's personal fondness of valor, wealth and women, but argued that King Xuan of Qi should extend his fondness of valor, wealth and women to all the people in the world. So, from Book 1B3 and Book 1B5, we can see that Mencius approved of the pursuit of profit for the sake of the people. Now connecting Book 1A1 with Book 1B3 and Book 1B5, Mencius dissociated "the pursuit of profit" into "the pursuit of self-centered profit" (the "appearance" level) (Book 1A1, Book 1B3, Book 1B5) and "the pursuit of people-centered profit" (the "reality" level) (Book 1B3, Book 1B5). He also highlighted the importance of holding to the principle of humaneness and righteousness in the pursuit of profit (Book 1A1) but in the meantime acknowledged the necessity of allowing for expediency (Book 1B3 and Book 1B5) (Wang 2018, p. 472). The debate on the relationship between (humaneness and) righteousness and profit is an outstanding reflection of Mencius's idea of holding to the Mean based on the principle of humaneness and righteousness and allowing for expediency, which is also the core of Confucian practical ethics.

Holding to the Mean based on principle and with expediency is also reflected in Mencius's view on the war. For example, in Book 1A6, in answering King Xiang of Liang's consecutive questions about how to settle and unite an empire, Mencius stated his claims (CAPITALIZED) explicitly that an empire can be settled through unity, that a king who is not fond of killing people can unite a settled empire, and that people will return voluntarily to a king who is not fond of killing people. Here Mencius expressed his disapproval of wars because people would be killed and of the kings who were fond of killing people in his time. He also promoted the humaneness-centered kingcraft - humane governance in not being fond of killing people (Chen 2018, pp. 40-41).

In Book 4A14, Mencius first quoted Confucius's negative attitude to the example of Ran Qiu for his enriching a ruler who was not practicing humane governance. Then Mencius associated such lack of humaneness with those men bent on making war. Claiming making wars as "leading the earth to devour human flesh", Mencius stated his opposition to wars so vehemently that he suggested sentencing the severest punishment to those who are skilled in war. Mencius's strong opposition to war is again because people are killed in wars, which is against the principle of

humaneness and righteousness as well as the human-oriented and people-oriented thoughts rooted in the principle (Chen 2018, p. 41).

Book 1A6 and Book 4A14 seem to suggest that Mencius was against wars and against rulers making wars because people would be killed and the principle of humaneness and righteousness was not practiced. However, in Book 1A5, when King Hui of Liang asked Mencius how to make revenge for those who had died for the kingdom of Wei in wars with the kingdoms of Qi, Qin and Chu, Mencius suggested that if King Hui of Liang practice humane governance and pursue profit for his people, then he could easily defeat those other rulers who did not practice humane governance and did not pursue profit for their people. This implies that Mencius approved of the wars made by the humane rulers and for the sake of the people's profit.

Book 1B11 further confirmed Mencius's positive attitude to wars made by rulers like Tang who practiced humane governance and acted for the people's profit (the opening stage, *italicized*). At the same time, Mencius on the one hand approved of King Xuan of Qi's making a war against the kingdom of Yan as a punishment of the ruler of Yan who oppressed its people and did not practice humane governance. On the other hand, Mencius pointed out the acts of Qi after conquering Yan were against the wills of the people and as a result against the principle of humane governance (the argumentation stage, underlined). At the end of the conversation, Mencius suggested that King Xuan of Qi stop his inhumane acts in the kingdom of Yan and consult with the people of Yan in order to stop an attack from the other states (the concluding stage, CAPITALIZED). Mencius's two-fold analysis on the state of Qi's making a war against the inhumane ruler of Yan and on Qi's stopping an attack by other states because of its inhumane acts after conquering Yan shows his consideration of expediency according to the principle of humaneness and righteousness in state governance and in terms of wars, which is essentially founded upon the differentiation between self-centered pursuit of profit and people-oriented pursuit of profit.

Now combining Book 1A5, Book 1A6, Book 1B11 and Book 4A14, we can see that Mencius dissociated "wars" into "unjust wars made by rulers who do not practice humane governance but aim for enlarging private profit" (the "appearance" level) (Book 1A6, Book 4A14) and "just wars made by rulers who practice humane governance and enlarge the people's profit" (the "reality" level) (Book 1A5, Book 1B11) (Chen 2018, p. 98). So, Mencius's view on wars also reflects his conception of reasonableness in holding to the Mean based on the principle but with expediency. In other words, Mencius did not totally negate wars and the necessity of making wars, but just indicated that pursuing the profit for the people under the guidance of the principle of humaneness and

righteousness should become the starting point of deciding on making a war.

5. CONCLUSION

The Confucian philosophy is a moral philosophy (Yang 2017, p. 124). Following but also developing from Confucius's core moral concept - humaneness (*Ren*), Mencius extended humaneness into humaneness and righteousness (*Ren Yi*), a combination of Confucian virtue ethic and Confucian deontic ethic. He further traced humaneness and righteousness back to a moral metaphysical basis - human nature being good, which can be more precisely elaborated as inclination to goodness that is inherent in human nature. Unlike the pursuit of pure knowledge as a philosophical interest in ancient Greece, ancient Chinese philosophers pursued to put knowledge into practice. Just as Mencius explained to his disciple Gongduzi why he argued, the ultimate purpose of Mencius's developing Confucian humaneness into humaneness and righteousness and tracing their source of moral metaphysics is to influence the realistic politics in his times by seeking a sound philosophical foundation for his advocacy of kingcraft as humane governance and to benefit the massive ordinary people. Such a philosophical foundation can be summarized as the conception of reasonableness in the moral idealist perspective. The extensive use of dissociation in the work *Mencius* helps lay bare what was counted as reasonable by Mencius in his argumentative discourse.

ACKNOWLEDGEMENTS: The study is supported by International Program for PhD Candidates, Sun Yat-sen University, and funded by Jiangsu Province Projects for Philosophy and Social Sciences in Higher Educational Institutions (Grant No. 2017SJB1076) , Argumentation Studies in Ancient China (Grant No. 17GZGX23) , and China National Social Sciences Foundation Project (Grant No. 19AZX017).

REFERENCES

- Chen, L (陈来). (2018). *Interpretation of Book King Hui of Liang* (《梁惠王篇》). In Chen Lai & Wang Zhimin (Eds.), (陈来、王志民主编) *Interpretations of the Seven Books of Mencius* (《孟子》七篇解读) (pp. 1-116). Shandong: Qilu Publishing House (山东: 齐鲁书社).
- Ding, Wei (丁为祥). (2004). Tradition: Specific and General - On the methodology of interpreting ancient books and records and discussing with Mr. Liu Qingping and Mr. Mu Nanke (传统: 具体而又普遍——论典籍诠释

- 的方法兼与刘清平、穆南柯先生商榷). In Guo Qiyong (Eds.) (郭齐勇主编), *Collection of contentions on Confucian ethics: Centering around "kins' covering up for each other"* (《儒家伦理争鸣集——以“亲亲互隐”为中心》) (pp. 188-201). Wuhan: Hubei Education Press (湖北教育出版社).
- van Eemeren, F. H., and Grootendorst, R. (1992). *Argumentation, communication, and fallacies: A pragma-dialectical perspective*. New Jersey: Lawrence Erlbaum Associates.
- van Eemeren, F. H. (2018). *Argumentation theory: A pragma-dialectical perspective*. Springer International Publishing AG.
- Li, C (李存山). (2018). *Interpretation of Book Jin Xin* (《尽心篇》). In Chen Lai & Wang Zhimin (Eds.) (陈来、王志民主编), *Interpretations of the Seven Books of Mencius* (《孟子》七篇解读) (pp. 829-945). Shandong: Qilu Publishing House (山东: 齐鲁书社).
- Mencius. (2004). *Mencius*. Trans. by D. C. Lau. London: Penguin Books Ltd. First published in 1970.
- Mencius. (2009). *Mencius*. Trans. by Irene Bloom. Edited and with an introduction by Philip J. Ivanhoe. New York: Columbia University Press.
- Perelman, Ch. & Olbrechts-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation*. Trans. by John Wilkinson and Purcell Weaver. Notre Dame / London: University of Notre Dame Press.
- van Rees, A. (2009). *Dissociation in argumentative discussions: A pragma-dialectical perspective*. Springer.
- Schwartz, B. I. (1985). *The world of thought in ancient China*. Cambridge, Massachusetts and London: The Belknap Press of Harvard University Press.
- Wang, Z (王中江). (2018). *Interpretation of Book Li Lou* (《离娄篇》). In Chen Lai & Wang Zhimin (Eds.) (陈来、王志民主编), *Interpretations of the Seven Books of Mencius* (《孟子》七篇解读) (pp. 421-572). Shandong: Qilu Publishing House (山东: 齐鲁书社).
- Xu, J (徐嘉). (2004). On the Confucian moral model of coordinating between principle and expediency. In Guo Qiyong (Eds.) (郭齐勇主编), *Collection of contentions on Confucian ethics: Centering around "kins' covering up for each other"* (《儒家伦理争鸣集——以“亲亲互隐”为中心》) (pp. 588-600). Wuhan: Hubei Education Press (湖北教育出版社).
- Yang, H (杨海文). (2017). *I am good at nourishing my vast flowing Qi: The world of Mencius* (《我善养吾浩然之气：孟子的世界》). Jinan (济南): Qilu Publishing House (齐鲁书社).
- Xiao, J (萧建华). (2004). On characteristics of Confucian ethics: Also discussing with Mr. Liu Qingping. In Guo Qiyong (Eds.) (郭齐勇主编), *Collection of contentions on Confucian ethics: Centering around "kins' covering up for each other"* (《儒家伦理争鸣集——以“亲亲互隐”为中心》) (pp. 230-240). Wuhan (武汉): Hubei Education Press (湖北教育出版社).

APPENDIX

The appendix has been omitted for brevity and is available upon request from the authors.

Strategic manoeuvring with the construction “We (all) know that X” in parliamentary debates

MATTHIJS LOOIJ
Leiden University
m.looij@hum.leidenuniv.nl

BERTIJN VAN DER STEENHOVEN
Leiden University
g.van.der.steenhoven@umail.leidenuniv.nl

The construction [we] – [perspective-indicating verb] – [that] – [argument] can be used by discussants to strategically introduce an argument. Through the combination of inclusive “we”, serving as an identity cue, and a perspective-indicating verb that expresses certainty, the argument is empathically presented as a common starting point. In this paper, we formulate three soundness conditions for strategic manoeuvring with this construction.

KEYWORDS: elocutio, presentational devices, parliamentary debate, strategic manoeuvring, starting points, pronoun “we”, verbs

1. INTRODUCTION

On 14 February 2019, the Spanish website The Corner published an interview with Ramón Jáuregui, Member of European Parliament. Asked if ‘unanimity is a structural disfunction in the Union’, Jáuregui answered that it can needlessly delay adequate policymaking in a time of quick developments. ‘Unanimity is the thorn in the side [of the European Union]’, he said. This quote was used as the header of the article on The Corner (see Figure 1). Interestingly, the editor added the formula “We all know that” at the beginning of the quote: “*We all know that* unanimity is the thorn in the side of the European Union” (our italics). By inserting “We all know that”, the editor suggests that Jáuregui presented his position is unanimously supported.

The above example shows that the grammatical construction “We (all) know that” can be added to an argument. In this paper we assess how the construction can be used by politicians as a rhetorical device.

“We All Know That Unanimity Is The Thorn In The Side Of The European Union”

TOPICS: Brexit EU Venezuela Eurosceptics European Commission European Elections

European Parliament Meseberg



The Spanish MEP Ramón Jáuregui.

POSTED BY: ANA FUENTES 14TH FEBRUARY 2019

Ana Fuentes (Strasbourg) | Where should the EU look in the future? What are the priorities? At a time of rapid change, protectionism and nationalist populism, the European Parliament has approved a document of minimums called **The Future of Europe**. As inevitably happens in such plural institutions, it is neither binding nor completely satisfies anyone, but sets out the challenges the still 28 members have to confront together if the European project is not to be diluted. We discuss it with Ramón Jáuregui, socialist MEP and rapporteur of the text.

Figure 1. Fragment of the article on TheCorner.eu (see Fuentes 2019).

According to the literature, the pronoun “we” creates group identity (Chilton & Schäffner 1997: 217, 2002: 30; Bazzanella 2002: 249). Perspective-indicating verbs such as “know” imply a high degree of certainty (Haeseryn et al. 1997: 1156-1158; Verhagen 2005: 100; van Leeuwen 2015: 155-157). In combination, the pronoun “we” and perspective-indicating verbs such as “know” are used to emphatically present an argument as a common starting point. Especially in politics, this might be strategic, since politicians are expected to represent a bigger whole: parties, coalitions, ‘the people’.

As far as we know, the strategic potential of the construction “We (all) know that X” in politics has not yet been discussed in the literature. From the perspective of rhetorical theory, a lot of research has been done

into the use of the pronoun “we” in the political domain, but research from an argumentation-theoretical perspective is lacking. Research into the use of what we here call “perspective-indicating verbs” (e.g. “know”, “believe”, “suppose”) is rather scarce, let alone the even more specific use of these verbs in combination with the pronoun “we”. Therefore, we carried out exploratory research into the strategic use of the grammatical construction “We (all) know that X”.

The central question of this paper is: *How can politicians manoeuvre strategically with the grammatical construction “We (all) know that X” in parliamentary debates?* To answer this question, we use the extended pragma-dialectical theory of argumentation, in which argumentation is analysed from both a rhetorical and a dialectical perspective, as a theoretical framework. It is assumed that discussants primarily have a rhetorical aim: they are out to win the discussion. On the other hand, they try to reach the dialectical aim of resolving the difference of opinion on the merits. Discussants will manoeuvre strategically between these two goals and, ideally, they try to be both reasonable and effective. In strategic manoeuvring three aspects are involved: (1) discussants make a selection from the topical potential, (2) they adapt their moves to the audience’s preferences, and (3) they exploit presentational devices. Strategic manoeuvring with grammatical constructions is an instance of exploitation of presentational devices (in classical rhetorical terms: *lexis* or *elocutio*).

The paper is organised as follows: first we will provide an argumentative characterisation of parliamentary debates (section 2). Then, we will turn to the strategic potential of the grammatical construction “We (all) know that X” to introduce an argument. It will be shown under which conditions strategic manoeuvring with the construction derails (section 3). The next section provides a demonstration of reasonable and derailed strategic manoeuvres in parliamentary debates (section 4). The final section provides a discussion and conclusion of the research as well as recommendations for future research (section 5).

2. STRATEGIC MANOEUVRING IN PARLIAMENTARY DEBATES

According to pragma-dialectics, argumentative discourse should be analysed in its institutional context. To this end, van Eemeren (2010) introduced the concept of argumentative activity types: culturally established communicative practices that have become more or less conventionalised and that are to a certain degree institutionalised. Activity types can be characterised argumentatively by describing the empirical counterparts of the four stages of a critical discussion: (1) the activity type’s initial situation, (2) its starting points, (3) the

argumentative means available in the activity type, and (4) its possible outcomes (van Eemeren 2010: 146, 152-158).

In the first discussion stage, the confrontation stage, it becomes clear what the difference of opinion is about and what type of difference it is (van Eemeren 2018: 36). In parliamentary debates, the initial situation is a disagreement on a policy issue, that is put on the table by either a Member of Parliament (MP) or the cabinet (van Eemeren & Garssen 2010: 31). Usually, the difference of opinion is mixed (different parties take opposite standpoints), but it could occur that a dispute is non-mixed (only one party has put forward a standpoint, which is questioned by another party).

Politicians speaking in parliament have a complex audience: the political opponent, who is the official antagonist, and the listening, reading or watching audience: the voting public (Tonnard 2011: 22; van Haaften 2017:181). The primary addressee of parliamentary debates are the voters: a non-interactive and heterogeneous audience, which may consist of (a mix of) supporters, opponents and neutral bystanders (van Eemeren & Garssen 2010: 24; Tonnard 2011: 23).

In the opening stage, the procedural and material starting points for the discussion are established (van Eemeren 2018: 36). Procedural starting points pertain to the procedure of the discussion (i.e. the discussion rules), while material starting points consist of propositions that the discussants can later in the discussion use in their argumentation. Usually, a parliament's Rules of Procedure, for instance the *Reglement van Orde voor de Tweede Kamer* for the Dutch parliament, contain most procedural starting points (Tonnard 2011: 25). The debate is organised as follows: participants deliver a speech with a fixed speaking time; other participants can interrupt the speaker to ask questions. Because of the heterogeneous character of the parliament, the agreement on material starting points will in many occasions only be partial and cannot be presumed without further verification. Speakers can interrupt each other to criticise the false presentation of a premise as a common starting point.

The aim of the argumentation stage is to test the acceptability of a standpoint by an exchange of arguments and criticisms (van Eemeren 2018: 37). There are no special constraints as to the argumentative means that can be employed in the argumentation stage (van Eemeren & Garssen 2010: 31). In general, politicians will put forward argumentation for or against the proposal or policy at issue, in response to expressed or anticipated criticism. The arguments should be based on material starting points that are, explicitly or implicitly, agreed upon in the opening stage.

The aim of the concluding stage is to determine the outcome of the discussion (van Eemeren 2018: 37). Although plenary debates are

always officially and explicitly closed by the President of the House, there is no real concluding stage: differences of opinion are not concluded by way of an intersubjective agreement on the outcome of the debate (van Eemeren & Garssen 2010: 31). Instead, the dispute is settled by a majority of votes.

The construction we discuss in this paper is a presentational device to strategically introduce an argument. The focus will therefore be on strategic manoeuvring in the argumentation stage of parliamentary debates.

3. THE STRATEGIC POTENTIAL OF “WE (ALL) KNOW THAT X”

According to van Eemeren (2010: 45), the dialectical aim of the argumentation stage regarding protagonists is “to achieve clarity concerning the protagonist’s argumentation in defence of the standpoints at issue”. On the other hand, the rhetorical aim for protagonists in the argumentation stage is “to establish argumentation that constitutes an optimal defence of the standpoints at issue”. In other words, in the argumentation stage of a parliamentary debate politicians have to formulate their arguments in such a way that they are maximally convincing to both their colleagues and the public, without becoming unreasonable.

The addition of “We (all) know that” to an argument can be used as a strategic manoeuvre in the argumentation stage. Since politicians are expected to represent a bigger whole, it might be effective to emphasise the fact that an argument is recognised by others. The construction should be regarded as a matter of wording (*lexis* or *elocutio*): it calls the addressee’s attention to the fact that a particular starting point is agreed upon in the opening stage; it does not affect the propositional content of the argument itself and it could, in principle, be left out without consequences for the reconstruction of the argumentation.

The construction “We (all) know that X” consists of two parts: the first part, “we (all)” expresses the shared character of an argument. Apart from “we (all)”, other formulations could be used as well, such as “all of us”, or other appeals to a bigger whole, such as “everyone”, “the whole world”, and “every reasonable person”.

The second part of the construction consists of a perspective-indicating verb and expresses certainty (“know”, “observe”). We will discuss the rhetorical potential of the separate parts of the construction in section 3.1 and 3.2. Then, we will turn to the combination of the two parts.

3.1 *The strategic potential of the pronoun “we”*

By using the personal pronoun “we”, a speaker creates a mental space in which he positions himself and his audience (Chilton & Schäffner 1997: 217). The pronoun “we” functions as an identity cue: it conceptualises group identity (Bazzanella 2002: 249; Chilton & Schäffner 2002: 30). By using “we”, a speaker distinguishes insiders and outsiders: people that belong to the group and people that do not belong to that group (Pennycook 1994: 175; Wilson 1990: 71). Therefore, the use of “we” is considered effective: it creates a bond between the speaker and the people he includes (Moberg & Eriksson 2013: 320). Because of the creation of this bond, the personal pronoun “we” can be used as a strategic means to emphatically present something as a common starting point.

It is not always clear who is included and who is not: the meaning of the personal pronoun “we” is often vague (Proctor & I-Wen Su 2011: 3253; Moberg & Eriksson 2013: 320). To solve vagueness, contextual information about the situation in which it is used is required (de Fina 1995: 400; Bull & Fetzer 2006: 11). Moberg & Eriksson (2013: 331) point out that the vagueness can be used strategically, for instance to disguise differences between parties within a coalition.

Following from the fact that politicians address a complex of audiences in parliamentary debates (as described in section 2), “we” can refer to many different groups. When a politician is faced with an opponent’s criticism regarding the false presentation of a premise as a common starting point, he can strategically make use of the vagueness of “we”, for instance by stating that he just meant to include his own party instead of the parliament as a whole.

3.2 *The strategic potential of perspective-indicating verbs*

The second part of the construction consists of a perspective-indicating verb. As a whole, the construction “We (all) [verb] that” forms a so-called matrix clause that provides a perspective on the so-called complement clause (i.e. “X”). According to van Leeuwen (2015: 155-157), depending on the nature of the matrix clause different degrees of certainty can be ascribed to the content of the complement clause. Compare the following matrix clauses:

- (1) *I think that* there will be a heatwave next week.
 - (2) *I am sure that* there will be a heatwave next week.
 - (3) *It is a fact that* there will be a heatwave next week.
- (van Leeuwen 2015: 156, our translation)

The first sentence, van Leeuwen argues, leaves much more room for discussion than the second and the third sentence, since it expresses more uncertainty. Thus, discussants can manoeuvre strategically by choosing a keen formulation of the matrix clause.

The construction “We (all) [verb] that X” leaves a discussant room to opt for different verbs that express different degrees of certainty (Haeseryn et al. 1997: 1156-1158; Verhagen 2005: 100). Compare for instance the following sentences:

- (4) We (all) *know* that there will be a heatwave next week.
- (5) We (all) *expect* that there will be a heatwave next week.
- (6) We (all) *are certain* that there will be a heatwave next week.

In the above sentences different degrees of certainty are ascribed to the propositional content of the complement clause. Compared to (4), the assertion in (5) is weaker: expecting something to happen is less sure than knowing something will happen. The formulation in (6) however might appear even stronger than (4): to be certain is stronger than to know.

3.3 Soundness conditions

The construction “We (all) know that X” works through a combination of inclusion, expressed by the pronoun “we” and certainty, expressed by a verb such as “know”. Together they form a powerful rhetorical device that gives strength to an argument. However, in some cases strategic manoeuvring with the construction derails. In this paragraph we formulate the soundness conditions for the use of the construction “We (all) know that X”.

First, the inclusion that has been enabled by using the pronoun “we” should be correct, that is: what is ascribed to a particular group should apply to this group. Indeed, politicians may speak personally, but often they represent others. If they speak in the name of others, they should always stick to what they have in common with the group they represent. If the group that is included does not agree upon the starting point that is presented as mutually shared, the politician violates the Starting Point Rule by falsely presenting a premise as a common starting point (van Eemeren 2018: 60).

Second, it should be clear who is meant by “we”. Due to the inherent vagueness of “we”, a politician might escape from his responsibility (e.g. “I did not mean to include the whole coalition, only my own party”). According to the Language Use Rule, a discussant’s contributions should be clear; if not, he commits the fallacy of misusing unclearness (van Eemeren 2018: 61).

Third, the certainty that is expressed by the verb should be in accordance with the actual amount of certainty among those who are included. A politician may falsely ascribe certainty to an utterance by using the wrong verb (e.g. “know” instead of “expect”).

4. STRATEGICALLY INTRODUCING ARGUMENTS WITH “WE (ALL) KNOW THAT” IN PARLIAMENTARY DEBATES

In this section, we will analyse examples of strategic manoeuvring with “We (all) know that X” that are either sound or derailed, based on the soundness conditions formulated in section 3. All examples are taken from the official transcripts of the so-called “General Debates” (*Algemene politieke beschouwingen*), an annual debate in the Dutch parliament to discuss the government’s plans for the next year.¹ These debates gain a lot of attention from the media. Therefore, it could be expected that politicians will do whatever they can to persuade their primary addressee: the voters.

We will start by analysing two examples of sound strategic manoeuvres; then we will discuss a couple of examples of derailed strategic manoeuvres, due to a violation of one of the soundness conditions.

4.1 Sound strategic manoeuvres

A strategic manoeuvre with the construction “We (all) know that X” is sound if it meets the criteria formulated in section 3: the use of “we” should be correct and clear, and the perspective-indicating verb should not be too strong. This is the case in example (7):

- (7) Mr. **Segers** (*ChristianUnion*): I want to continue about the point Mr. Wilders referred to, namely attacks, national security and the fight against terror. The Prime Minister started rightly referring to what happened in New York yesterday. *We know that* an attack can also take place in the Netherlands. That’s a danger, a harm we need to get ourselves armed against. We need to use all means necessary to make a stand against this. (General Debate 2017)

¹ The transcripts are retrieved from zoek.officielebekendmakingen.nl. For all examples holds: our translation (from Dutch to English) and our italics/underlining.

Segers' argumentation can be reconstructed as follows:

- 1 We need to use all means necessary to make a stand against terrorism, *because*
- 1.1 *We know that* an attack can also take place in the Netherlands.

In principle, Segers could just have said that “an attack can also take place in the Netherlands”. Instead, he manoeuvres strategically by adding “we know that” to introduce his argument. He presents the possibility of an attack taking place in the Netherlands as factual knowledge by using the verb “know” (*weten*). By using the personal pronoun “we”, he presents it as a fact that is common knowledge among the audience (i.e. as a material starting point). Since it is indeed a common starting point that, in principle, an attack could take place in the Netherlands, Segers is both reasonable and effective: his strategic manoeuvre is sound.

Excerpt (8) shows an example of a sound strategic manoeuvre with another variant of the construction:

- (8) Mr. **Asscher** (*Labour Party*): I have a short question in response to Mr. Dijkhoff's contribution about healthcare. [...] There are serious concerns among community nurses about the 100 million of budget cuts due to the Outline Accord, as shown by the Netherlands Bureau for Economic Policy Analysis. *We all know that* the total costs for healthcare are rising. Something needs to be done about that. (General Debate 2017)

Instead of “We know that”, Asscher adds “We *all* know that” to his argument to make an even stronger claim. His argumentation can be reconstructed as follows:

- 1 Something needs to be done about the healthcare costs, *because*
- 1.1 *We all know that* the total costs for healthcare are rising.

As in example (7), “We all know that” could have been left out, without changing the propositional content of the argument. Asscher manoeuvres strategically by presenting the rise of total healthcare costs as a fact that is generally known. “We all” most likely refers to everyone among the audience (voters as well as other MPs). Since the costs for healthcare have been an issue in several debates, it can be assumed that this is indeed a common (material) starting point. Thus, Asscher is both effective and reasonable.

4.2 Derailed strategic manoeuvres

4.2.1 False inclusion

The first soundness condition concerns the correct use of “we”: a speaker should not include others that do not share his starting point(s). This condition is violated in example (9):

- (9) Mr. **Wilders** (*Freedom Party*): It also seems that SP-voters do not want more asylum seekers to come to The Netherlands. *We all know that* they aren’t sad people without food and water. Even if there were such people, then indeed, they could better get this in the region than here. (General Debate 2015)

Wilders’ standpoint is left implicit in this fragment, but it could be reconstructed as follows:

- (1) (We should not allow more asylum seekers to come to The Netherlands), *because*
1.1 *We all know that* they aren’t sad people without food and water.

Instead of just claiming that “they aren’t sad people without food and water”, Wilders introduces his argument with the construction “We all know that” to manoeuvre strategically. He presents the propositional content of his argument as a fact (due to the verb “know”) that is shared among the public (due to the pronoun “we” combined with the word “all”). By being as persuasive as possible (Wilders’ rhetorical aim), Wilders loses sight of his dialectical aim: the starting point he uses as a premise in his argumentation is not shared by many other MPs and a large part of the public. By falsely presenting this premise as an accepted starting point, Wilders violates the pragma-dialectical Starting Point Rule.²

Another instance of a violation of the first soundness condition can be found in example (10):

- (10) Prime Minister **Rutte**: *We know of course that* the policy concerning the Wmo (Social Support Act) is transferred from the national government to the local authorities, since local authorities are better able to deliver tailored work via the Wmo and Special Assistance. (General Debate 2015)

² It could be argued that the use of the verb “know” is also wrong here (and that Wilders, thus, violates the third soundness condition as well). However, Wilders claims that he really *knows* that these people are not sad. Therefore, we analysed it as a violation of the first soundness condition, rather than the third.

At first sight, the construction “We know of course that” seems to introduce the factual starting point that the policy concerning the Wmo is transferred from the national government to the local authorities. Since this is indeed the case and it can be expected to be known by the MPs, the use of “We know of course that” is perfectly sound for this part. However, the construction also relates to the second part of Rutte’s statement, in which he presents his opinion (i.e. local authorities are better able to deliver tailored work) as a common starting point. Apart from the pronoun “we” and the verb “know”, the adjunct “of course” contributes to effectivity: by introducing his argument with this construction, Rutte makes it seem that it is an accepted starting point.³ Moreover, the combination this construction with both a known fact and Rutte’s opinion makes it very strategic. Nevertheless, Rutte violates the Starting Point Rule by falsely presenting something as an accepted starting point.

4.2.2 The inherent vagueness of “we”

The second soundness condition relates to a clear use of “we”: a speaker should not misuse its inherent vagueness. A violation of the condition is shown in example (11):

- (11) Mr. **Roemer** (*Socialist Party*): It is a pity that the Prime Minister rejects this plan [a national bank for investments in small and medium sized enterprises, ML&BvdS] in advance. We know the system from the past. *We know that* it works. We know that currently the banks cannot do this.

Prime Minister **Rutte**: I have discussed it with the Minister for Economic Affairs and in the first place I assure Mr. Roemer that we will comment on his suggestion when we will send the evaluation about the BMKB to the House. Being very honest, I repeat that our basic attitude is not positive. (General Debate 2014)

Roemer’s argumentation can be reconstructed as follows:

- 1 The Prime Minister should reconsider this plan [a national bank for investments in small and medium sized enterprises], *because*
 - 1.1a We know the system from the past, *and*
 - 1.1b *We know that* it works, *and*
 - 1.1c We know that currently the banks cannot do this.

³ Rutte’s standpoint is left out of this excerpt.

Roemer uses the construction “We know that” to introduce all three of his coordinative arguments.⁴ The use of this construction in argument 1.1b is particularly interesting, since Roemer uses the construction to present his opinion about the plan as an accepted starting point: the verb “know” expresses certainty, while the pronoun “we” expresses inclusion. As becomes clear from Prime Minister Rutte’s response (underlined), Roemer’s premise 1.1b is not a common starting point: Rutte disagrees with Roemer about the effectiveness of his plan. Nevertheless, since he does not specify who he wants to include, Roemer could, in principle, say that Rutte might disagree, but that his party still knows that the plan works. In other words, there are several interpretations of the group that “we” is referring to; in some interpretations the inclusion is correct, while in others it is not. Thus, Roemer strategically uses the inherent vagueness of “we” to present his argument as persuasive as possible. Nevertheless, because of this vagueness, this could count as a violation of the Language Use Rule.

4.2.3 Different degrees of certainty

Strategic manoeuvres can also derail because the perspective-indicating verb expresses too much certainty (soundness condition 3). The following excerpt is taken from the General Debate of 2017. In that year, one topic that was severely criticised by the opposition was the coalition’s proposal to abolish dividend tax, good for 1.4 billion euros of the government’s incomes:

- (12) Mr. **Klaver** (*Green Party*): The entire coalition agreement and all measures are always checked on effectiveness. [...] *We know* about this measure pertaining to the dividend tax *that* it does not contribute to the increase of employment opportunities. [...] I cannot understand such a choice, especially not in the light of the other desires we have for our society, for instance better salaries in education or in healthcare.

President of the House: You made your point.

Mr. **Dijkhoff** (*People’s Party for Freedom and Democracy*): Mr. Klaver turns the fact that one organisation says it is not possible to predict or calculate the effect, into the fact that the measure has no effect. Those are two different things. (General Debate 2017)

⁴ For 1.1a, he uses “we know” instead of “we know that”, but the rhetorical function is the same.

According to Klaver, the abolishment of the dividend tax cannot be justified: it is not effective, yet it concerns a lot of money. His standpoint and argumentation can be reconstructed as follows:

- 1 The dividend tax should not be abolished, *because*
- 1.1 *We know that* the abolishment of the tax is not effective.

In policymaking, considerations about the effectivity can be a strong argument against a policy proposal. In the presentation of his argument, Klaver suggests that it is evident that the abolishment of the dividend tax is not effective: it is something “we know”. According to Dijkhoff this is not in accordance with the by then available insights. It is rather unsure what the effects of the measure will be. However, given the possible negative effects of non-action, he states, the measure is worth a gamble.

In the above example, Klaver’s use of the verb “know” suggests too much certainty about the (in)effectiveness of the abolishment of the dividend tax, which is pointed at by Mr. Dijkhoff (underlined). The strategic manoeuvre derailed: based on the available insights, the inclusion of the House in his point of view is not justified. Klaver falsely presents his argument as a common starting point, thereby violating the Starting Point Rule.

5. DISCUSSION AND CONCLUSION

In this paper we assessed how politicians can manoeuvre strategically with the construction “We (all) know that X” in parliamentary debates. The construction can be a powerful rhetorical device to emphasise the fact that the content of their argument is recognised by others. Especially in politics this might be strategic, since politicians are expected to represent a bigger whole.

The rhetorical potential of the construction is due to the combination of inclusion by using “we” (all) and certainty by using perspective-indicating verbs such as “know”.

Strategic manoeuvring with the construction “We (all) know that X” derails when a discussant violates the soundness conditions related to this construction. First, the inclusion should be justified, that is: if a speaker tends to represent others, he should stick to what he has in common with this group. Second, it has to be clear who is included by the use of “we”. Third, the amount of certainty expressed by the verb (e.g. “know”) should be in accordance with the view of those who are included.

In this research we focused on the political domain. In future research other domains of communication could be explored. Experimental research could indicate if, and if yes, how the addition of (different variants of) the construction “We (all) know that X” affects

ordinary language user's perception of an argument. Lastly, we focused on verbs that express certainty. Future research could also focus on the strategic use of verbs that express uncertainty, e.g. to avoid responsibility, such as "do not exclude" in "We do not exclude (the possibility) that X".

REFERENCES

- Bazzanella, C. (2002). The significance of context in comprehension: The 'We case'. *Foundations of Science*, 7(3), 239-254.
- Bull, P., & Fetzer, A. (2006). 'Who Are We and Who Are You?' The Strategic Use of Forms of Address in Political Interviews. *Text & Talk*, 26(1), 3-37.
- Chilton, P.A., & Schäffner, C. (1997). Discourse and Politics. In T. van Dijk (Ed.), *Discourse as Social Interaction. Discourse Studies: A Multidisciplinary Introduction*. Vol. 2 (pp. 206-230). London: SAGE.
- Chilton, P.A., & Schäffner, C. (2002). *Politics as Text and Talk: Analytic Approaches to Political Discourse*. Amsterdam: John Benjamins.
- De Fina, A. (1995). Pronominal choice, identity, and solidarity in political discourse. *Text*, 15(3), 379-410.
- Eemeren, F.H. van. (2010). *Strategic Maneuvering in Argumentative Discourse: Extending the Pragma-dialectical Theory of Argumentation*. Amsterdam: John Benjamins.
- Eemeren, F.H. van, & Garssen, B. (2010). In Varietate Concordia – United in Diversity: European Parliamentary Debate as an Argumentative Activity Type. *Controversia*, 7(1), 19-37.
- Eemeren, F.H. van. (2018). *Argumentation theory: A pragma-dialectical perspective*. Cham: Springer.
- Fuentes, A. (2019, 14 February). "We All Know That Unanimity Is The Thorn In The Side Of The European Union". In The Corner. Accessible online: <http://thecorner.eu/news-europe/we-all-know-that-unanimity-is-the-thorn-in-the-side-of-the-european-union/78320/>.
- Haaften, T. van. (2017). Strategic manoeuvring with presentational choices in Dutch parliamentary debate. In F.H. van Eemeren & Wu P. (Eds.), *Contextualizing Pragma-Dialectics* (pp. 177-192). Amsterdam: John Benjamins.
- Haeseryn, W., Romijn, K., Geerts, G., Rooij, J. de, & Toorn, M.C. van den. (1997). *Algemene Nederlandse Spraakkunst* (2nd revised edition). Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn.
- Leeuwen, M. van. (2015). *Stijl en politiek: Een taalkundig-stilistische benadering van Nederlandse parlementaire toespraken* [Style and politics: A linguistic-stylistic approach to Dutch parliamentary speeches] (PhD-dissertation Leiden University). Utrecht: LOT.
- Moberg, U., & Eriksson, G. (2013). Managing ideological differences in joint political press conferences: A study of the pragmatic use of the personal pronoun 'we'. *Journal of Language and Politics*, 12(3), 315-334.
- Pennycook, A. (1994). The Politics of Pronouns. *ELT Journal* Volume, 48(2), 173-178.

- Proctor, K., & I-Wen Su, L. (2011). The 1st person plural in political discourse: American politicians in interviews and in a debate. *Journal of pragmatics*, 43, 3251-3266.
- Tonnard, Y.M. (2011). Getting an issue on the table: A pragma-dialectical study of presentational choices in confrontational strategic maneuvering in Dutch parliamentary debate (PhD-dissertation University of Amsterdam). [S.l: S.n.].
- Verhagen, A. (2005). *Constructions of Intersubjectivity: Discourse, Syntax and Cognition*. Oxford: Oxford University Press.
- Wilson, J. (1990). *Politically Speaking*. Oxford: Blackwell.

Argument, justification and inquiry: a pragmatist approach

FEDERICO E. LÓPEZ

*Philosophical Research Center, School of Humanities and Education
Sciences - National University of La Plata*
flopez@fahce.unlp.edu.ar

This paper aims to contribute to the understanding of argumentation in inquiry by means of the discussion of two theses inspired by the pragmatist philosopher John Dewey, namely: 1. In order to properly understand the logic of inquiry, it is necessary to overcome the prevailing justificationist and retrospectivist bias in current studies of reasoning and argumentation; 2. The concept of epistemic practice, and the idea that logic in a broad sense is a reconstructive study of epistemic practices, can serve as a promising axis in the organization of the field.

KEYWORDS: argumentation, epistemic practices, inquiry, John Dewey, justification, problem-solving

1. INTRODUCTION

In the last years, important and valuable attempts have been made to understand the relationships between argumentation and inquiry and to expand informal logic and argumentation theory to include argumentation in inquiry as one of its objects of analysis. (Bailin & Battersby, 2016; Blair, 201; Battersby & Bailin, 2018; Pratt Scott, 2010; Wohlrapp 2014) Such latter attempts respond, by and large, to the need to overcome a practical or pedagogical problem, a problem to which current studies of argumentation seem not to offer a completely satisfactory answer. When teaching informal logic, argumentation theory or critical thinking at different educational levels, one can have the feeling that what it is taught is more useful for defending any given point of view, or convincing others, than to generate in students the disposition of forming a point of view, of reaching a conclusion on the basis of reflective thinking, which would involve gathering information or evidence and thus construct a judgement, a well-founded opinion. Besides, this need for a critical process of opinion formation—not

committed in advance to an already formed point of view, nor structured by it—is also considerably valuable in a context in which the democratic election of authorities is strongly influenced by communication strategies linked to some political and economic interests, which are now more than ever effective by virtue of the development of information technologies and the dissemination of fake news through social networks and mass media.

This paper aims to contribute to such developments by means of the discussion of two theses. These theses—inspired by the aforementioned work—constitute a hopefully clearer and more precise reformulation of some ideas of the pragmatist philosopher John Dewey. Thinking with and beyond Dewey, I propose to consider the following two ideas: 1) In order to properly understand the logic of inquiry, it is necessary to overcome the prevailing justificationist and retrospectivist bias in current studies of reasoning and argumentation; 2) The concept of epistemic practice, and the idea that logic in a broad sense, in a sense that includes informal logic and argumentation theory, is a reconstructive study of epistemic practices, can serve as a promising axis in the organization of the field. These are broad and ambitious theses whose establishment would indeed require more extensive work than the one I can present on this occasion, a work that is in its early stages. Thus, I will present these ideas as hypotheses and I will try, in this paper, to briefly explain their meaning and their scope.

2. GOING BEYOND RETROSPECTIVISM

One of the aims of this first section is to contribute to a recovery of the philosophy of logic of John Dewey. He is one of the most prominent representatives of the tradition of classical pragmatism, a tradition in which a broad and deep study of logical issues was developed. Indeed, the contributions of Ch. S. Peirce and C. I. Lewis, both belonging to this tradition, have been incorporated into the logic and their names already have a prominent place in the history of the discipline. This is not what happened with Dewey, whose thinking has been fundamentally shaped by his own work on logic, a work that has been developed for more than 40 years. As early as 1916, he published a book entitled *Essays in Experimental Logic*, a title that seems now more plausible than then, and, in 1938, he published another book, entitled *Logic: the theory of inquiry*, in which he criticizes the philosophical foundations of the formal approach to logic. Despite that, Dewey is not usually considered as a logician and his work is not even considered by historical studies of logic and argumentation theory. While it is true that what led Dewey to deal with logic was a much broader philosophical and even political motivation, he did develop an interesting philosophy of logic that had a

remarkable influence, for example, in the “critical thinking” tradition, and even in an relevant author as Stephen Toulmin. As I have pointed out elsewhere, (López, 2012) it can be argued that the thinking of the latter has incorporated some important ideas of pragmatism in general and of Dewey in particular. Considering the seeming rejection of Dewey by Toulmin in *The Uses of Argument*, it is worth remembering the following late recognition:

Let me begin with a word of gratitude to John Dewey. His book, *Essays in Experimental Logic*, was regarded with some contempt by my colleagues in Britain. But its great merit is to show, long before the rest of us, how reasoning enters not only into technical life, but also into everyday life; so that how we express ourselves and -more important- the activities within which we speak and act set the stage within which judgments of soundness and acceptability can alone be made (Toulmin, 2004, p. 111).

However, there is one aspect of Dewey’s thinking that has not been incorporated, even by Toulmin, and which is crucial in his understanding of logic. According to Dewey, logic as a science or discipline is the study of the process in which we solve some problems by using reflective thinking. More precisely, reflective thinking or inquiry—Dewey uses both expressions—is the process by which we solve problems using intelligence and, as a consequence of which, we reach warranted assertions. One implication of this is that knowledge or conclusions and arguments cannot be understood unless the process of constructing them is considered, because in his opinion the only logical reason we can have to hold a belief is that it is the result of some well conducted inquiry. Of course, we usually maintain beliefs for other reasons as bare impulse, imposition or tradition, but logic, as a normative-descriptive or reconstructive study of argumentation or reasoning should not be developed as a modelling of those situations.

We face here an attempt to analyse conclusions or knowledge as products by considering the process by which they are stated. This attempt was contrary to the prevailing idea that when speaking of knowledge and conclusions, process and product must be kept apart. In order to fully understand this, it is convenient to bear in mind something that Larry Laudan has pointed out. In his opinion:

An event of major significance occurred in the course of 19th-century philosophy of science. The task of articulating a logic of scientific discovery and concept formation -a task which had been at the core of epistemology since Aristotle’s *Posterior Analytics*- was abandoned. In its place was put the

very different job of formulating a logic of post hoc theory evaluation, a logic which did not concern itself with how concepts were generated or how theories were first formulated. This transformation marks one of the central watersheds in the history of philosophical thought, a fundamental cleavage between two very different perspectives on how knowledge is to be legitimated." (1981, p. 183)

This shift was and remains so important that when a contemporary reader faces Dewey's logic, epistemology or ethics, especially if trained in the tradition of analytic philosophy, has the impression that, even if Dewey had some good ideas, he defended them in a logically flawed way. Indeed, once the first project has been abandoned, based on a dichotomous distinction between context of discovery and context of justification, embracing the idea that the very object of logic (and philosophy of science) is *justification*, Dewey's position seems completely outdated.

Now, although the distinction between contexts mentioned above has been strongly questioned in the field of philosophy of science, I believe that the idea that logic is focused on the question of justification, and its counterpart criticism, of beliefs or points of view and that it has, therefore, a certain retrospective character is still the dominant position. In this sense, the following quote from *The Uses of Argument* can be taken as indicating a program that argumentation studies would largely adopt:

[L]ogic is concerned not with the manner of our inferring, or with questions of technique: its primary business is a retrospective, justificatory one—[it is concerned] with the arguments we can put forward afterwards to make good our claim that the conclusions arrived at are acceptable, because justifiable, conclusions (Toulmin, 2003, p. 6).

Moreover, even when attempting to account for the use of arguments in research contexts, this retrospective and/or justificatory nature is also present. For example, in a paper in which he discusses the possibility of using the pragmadialectical model of critical discussion to model argumentation in investigations, Anthony Blair states that "An epistemic investigation (i.e. the phenomenon he is trying to model) begins with a question about whether some judgment is justified." (2012, p. 293). Although Blair's purpose is to model a process in which agents are not committed in advance to the belief or judgment being investigated, nor try to persuade anyone of it, he models the process as starting with the question about whether a judgement that the agent

already has is justified or not. It is noteworthy that in this context, inquiry is not so much about problems, things or phenomena in the world, but only or mainly about concepts, hypotheses or theories that somehow have been already formulated. Indeed, in his presentation of the elements of epistemic investigations, Blair not only rejects any temporal connotation of a research process that could have *moments* or *stages*, he does not consider any element of “hypothesis formulation”, but only of revision, assuming simply that the hypothesis has been formulated before, so that its formulation would not be properly an internal element within the investigation, but an external point of departure, almost an excuse or motive.

The point that I am trying to highlight is that, even when the studies of argumentation focus in inquiry, the starting point is usually the point of view or the judgment whose merits or justification is to be established. Thus, whether it is to persuade someone in a persuasive dialogue, to resolve a difference of opinion or to investigate whether a judgement is justified, the logical task, or the rational task, so to speak, is always retrospective: it begins with what has to be supported, with the conclusion or point of view and then it goes back in search of the foundations. In such a context, the so-called illative core of premises-conclusion is the fundamental, the ultimate core of logical analysis, and even more so the locus where all rationality resides.

In this way, and after the studies on argumentation abandoned the formalist and the deductivist bias, it seems that it continues to construct its object incorporating a justificatory and / or retrospective bias. Of course, I am not denying that many times we do justify, criticize or evaluate points of view already formulated or that this is a necessary and important activity. However, this is not necessarily so, particularly when engaged with inquiries.

In contrast to such a conception, inquiry is understood by Dewey, not as the attempt to determine whether a hypothesis is true or acceptable, or it is not. Inquiry is a kind of response of an agent to a problem. According to Dewey, it would be more accurate to say that it is a response to an indeterminate situation, inasmuch as the definition or characterization of the problem is already a stage within the inquiry, and usually something surrounded by controversy. Indeed, the definition of the problem works as a persuasive definition (Zarefsky, 2006) which favors one response over others. Consequently, something similar could be said of inquiries which start by discussing already at hand points of view or hypothesis: they assume some characterization of the problem, overlooking other possibilities and aspects of the problem.

Another interesting point to mention is that, according to Dewey, inquiry is a process that has a *temporal dimension*: nothing that

can properly be called a solution can be found as a solution from the beginning, since in that case no investigation would be necessary. The process that logic must model, according to Dewey, is the process by which a solution to a problem is obtained or produced through an investigation. Of course, not every solution of a problem properly involves investigation, because a problem can also be solved by appealing to a pre-established or standardized response mode. An even if in such a process, the evaluation of hypotheses and judgments, as well as the resolution of differences of opinion and even persuasion in the proper sense can find their place—since inquiry might be collectively carried out—they cannot be identified with it. In other words, according to this point of view, the investigation, and consequently logic itself—if it is concerned with inquiries—cannot be thought as modelling only a retrospective and justificatory process.

To conclude this section, it is interesting to point out that, at the moment of the rise of the formalist conception of logic, but without denying the value of the formal developments that he understands as a part of logic, Dewey saw the necessity of reflecting on reasoning and inference in a fully contextualized way. In his opinion, it is only in its own context that the nature, function and norms of inference and reasoning can and should be studied. And that context was, in his opinion, inquiry, not only scientific but also common-sense inquiries. Thus, logic was in his opinion an “empirical” science, in the sense that it studies an observable process, and has a normative dimension insofar as it reconstructs and reformulates the norms that permit constructing and establishing warranted solutions. In other words, logic in Dewey’s view is the study of the practice of reflective thinking understood as a problem-solving process.

3. A LOGIC OF EPISTEMIC PRACTICES

In this section, I would like to suggest the hypothesis, inspired by Dewey but certainly going beyond him, that the concept of *epistemic practice* can be a fruitful one to understand the relationships between different studies on argumentation. To fathom what I am suggesting, it is appropriate to point out that the theory of argumentation often has a unifying impulse: as many authors have pointed out (Blair, 2007, Bermejo Luque, 2009), most approaches propose a way to understand argumentation that claims to have a broad enough scope as to cover the entire phenomenon of argumentation. Thus, for example, logical, rhetorical or dialectical approaches, which are usually linked with argumentation as product, process or procedure, are intended to be a total approach to argumentation, an approach that would explain the essential aspects that the other approaches highlight. Indeed, as

Bermejo Luque points out, while, on the one hand, the tripartite distinction product-process-procedure or logical-dialectic-rhetorical, suggests the idea that there are three complementary approaches on the phenomenon of argumentation, the truth is that each theory tends to have, as mentioned, an allegedly *total scope*.

As presented in the previous section, Dewey's position seems to be another theory with a totalizing pretension: the logic is theory of inquiry, which is a process subjected to rules that begins with a problem and ends with its solution. However, the development of informal logic and the theory of argumentation in recent decades and the attempts to reconstruct the logic of argument evaluation, audience persuasion and difference of opinion solving could hardly be reduced to the model of inquiry proposed by Dewey. Moreover, although inquiry is for the pragmatist an irreducible social, communicative and collaborative phenomenon, he has not considered the basic characteristics and difficulties involved in trying to resolve differences of opinion or to persuade an audience. or, at least, he did not think that it deserved an independent study. In the same way, attempts to reduce inquiry to a critical discussion tend to ignore central aspects of it, such as the fact that they have a prospective temporal structure.

According to Dewey, logic was the study of a practice. However, I see no reason in reducing the logic to the study of a single type of practice. Indeed, the practice of solving a real problem through research, for example, not knowing who to vote for in the next election, seems to me a different practice than convincing someone to vote for a given candidate or the to resolve a difference of opinion about who is the best candidate. Perhaps other examples could be taken under consideration, but the one mentioned aims at suggesting that these are different practices that do not have to be analyzed with or reduced to a single model.

Thus, it could be suggested that logic, in the broad sense I have used, can be understood as the study of epistemic *practices*, not in the singular. Generally speaking, epistemic practices can be understood as different ways in which knowledge is produced, communicated or legitimated. By "knowledge" I refer both to scientific knowledge, to common sense, and to public opinion in general, regardless of its specific topic, so that we can encompass what is traditionally understood as knowledge about facts and moral knowledge. For example, inquiry, conceptual analysis, the evaluation of arguments, the attempt to convince someone about a point of view or the attempt to resolve a difference of opinion could be considered as different epistemic practices. Of course, these are not completely divergent practices, but rather practices that share common aspects. Probably, inquiry is the broadest, in the sense that it might include other types of

practices. However, it is not clear that all of them can be reduced to one of these models, or to some more general and abstract enough model to cover them all, not at least without significant losses.

4. CONCLUSION

In this paper I have argued that current studies of reasoning and argument have a justificatory and retrospective bias which needs to be overcome in order to properly account for the use of argument within inquiries. Inspired by John Dewey's pragmatist logic, it has been claimed that inquiry is a prospective and constructive process, with a temporal structure in which the definition of the problem and the formation of hypothesis must be considered as essential stages. On the other hand, it has been suggested that the concept of epistemic practices can be a useful tool to understand the relations between several approaches in argumentation theory that, more than describing different dimension of one and the same practice, seem to be describing diverse kinds of epistemic practices.

To conclude, it can be pointed out that, as Toulmin maintained regarding his own logic, a logic of epistemic practices must be a *comparative affair*. However, it can be added according to what has been stated, that the main or first distinction is not between fields, but between kinds of epistemic practices. We can search for the common features but with an eye in avoiding the practice of reducing one model to another. But, paraphrasing Toulmin, it can be said that we must learn to tolerate in comparative logic a state of affairs long-taken-for-granted in comparative anatomy: A man, a monkey, a pig, or a porcupine—not to mention a frog, a herring, a thrush and a coelacanth—each will be found to have their own anatomical structure. If we ask about that structure, about its elements and its criteria of normality, we must ask these questions within the limits of a given practice, and avoid, as it was, putting clothes on an ape to make it look as a human being, or adding quills to a pig to make it look as a porcupine.

REFERENCES

- Blair, J. A. (2007). Investigations and the critical discussion model. In F. H. van Eemeren, J. A. Blair, C. A. Willard and B. Garssen (Eds.). *Proceedings of the Sixth Conference of the International Society for the Study of Argumentation*, pp. 153-158. Amsterdam: SicSat
- Blair, A. (2012) *Groundwork in the Theory of Argumentation: Selected Papers of J. Anthony Blair*. Dordrecht: Springer.
- Bailin, S. & Battersby, M. (2016) *Reason in the Balance. An Inquiry Approach to Critical Thinking*. Indianapolis: Hackett Publishing Company.

- Battersby, M., & Bailin, S. (2018) *Inquiry: A New Paradigm for Critical Thinking*. Windsor: University of Windsor.
- Bermejo Luque, L. (2009). The Aristotelian distinction between Logic, Dialectic and Rhetoric, and its role within Argumentation Theory. *Cogency*, 1(2). 27-48.
- Dewey, J. (1916). *Essays in Experimental Logic*. New York: Dover Publications.
- Dewey, J. (1938). *Logic: The Theory of Inquiry*. New York: Henry Holt
- López, F. (2017). Las huellas pragmatistas en *Los usos de la argumentación*. *Cogency*, 4(1), 25-70.
- Pratt, Scott L. (2010). *Logic: Inquiry, Argument, and Order*. New Jersey: Wiley.
- Toulmin, S. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.
- Toulmin, S. (2004). Reasoning in Theory and Practice. *Informal Logic* 24 (2). 111-114.
- Wohlrapp, H. R. (2014). *The Concept of Argument. A Philosophical Foundation*. Dordrecht: Springer.
- Zarefsky, D. (2006). Strategic Maneuvering through Persuasive Definitions: Implications for Dialectic and Rhetoric. *Argumentation* 20 (4). 399-416.

Dissent as a voice of support:
The influence of argument types in providing support
to those who have experienced racial discrimination

UTTARA MANOHAR
University of Wisconsin-La Crosse, USA
umanohar@uwlax.edu

SUSAN L. KLINE
The Ohio State University, USA
kline.48@osu.edu

Racial microaggressions are regularly experienced by people of color. We contend that effective support includes arguments that affirm recipients' social identity and refute claims about the threatened social group. We found that high quality social identity arguments had more valid reasoning and produced more social identity esteem than low quality arguments. High quality social identity arguments were also associated with more causal reattribution that linked to higher argument evaluations and social identity esteem.

KEYWORDS: argument quality, racial microaggressions, social identity, social support, speech acts

1. INTRODUCTION

Experiences of racial discrimination can have negative consequences for the well-being of people of color (Pascoe & Smart Richman, 2009; Paradies et al., 2015). Unfortunately, the role of argument for disarming the effects of racial discrimination has not been extensively examined by argument scholars, despite argumentation theory's potential for producing social consensus and change. Fortunately, the social support literature has documented a dissent and restorative function for individuals who have experienced forms of racial discrimination (e.g., Davis & Afifi, 2019; Mossakowski & Zhang, 2014). Thus, the purpose of our study is to examine the role of argument for providing support to those who have experienced racial discrimination.

2. RACIAL MICROAGGRESSIONS, ARGUMENT, AND SUPPORT

In the United States racial microaggressions are defined by Sue and his colleagues as “brief and commonplace daily verbal, behavioral, or environmental indignities, whether intentional or unintentional, that communicate hostile, derogatory, or negative racial slights and insults to the target person or group” (Sue, Bucceri, Lin, Nadal, & Torino, 2007, p. 273). Sue (2010) has identified three types of microaggressions: acts that degrade a person’s racial heritage, acts that invalidate or exclude the reality of a person of color, and acts that overtly convey racist attitudes and beliefs. From this typology three types of communicative acts can function as racial microaggressions: avoidance (acts that convey an unwillingness to engage in contact with people of color), derogation (acts that express the superiority of members of privileged social groups), and categorization (acts that convey stereotypical assumptions about a person’s race or ethnicity).

Researchers have found that these types of racial microaggressions are linked to various negative mental health outcomes such as depression, anxiety, and emotional distress (Lui & Quezada, 2019). Racial microaggressions also pose a threat to an individual’s social identity, a part of one’s self-concept tied to membership in a social group (Cohen & Garcia, 2005; Tajfel, 1981; Thoits, 2013). Social identity threats associated with racial microaggressions negatively affect an individual’s collective self-esteem and social identity worth (Luhtanen & Crocker, 1992; Thai, Lyons, Lee, & Iwasaki, 2017; Turner, 1999).

2.1 *Racial microaggressions and argument*

To address racial microaggressions we turned to the line of argumentation theory and research that has focused on the normative pragmatics of speech acts and their role in argument. These analyses have produced insights about arguments involving acts of accusing, proposing and exhorting, among others (Kauffeld, 1998; Kauffeld & Fields, 2005; Kauffeld & Innocenti, 2016). Extending this work, racial microaggressions can be considered to be conversational exercitive speech acts that fix what is permissible in a certain situation to “confer or remove the hearer’s rights or privileges” (Austin, 1962, p. 120). We contend that racial microaggressions constitute a type of speech act that expresses a permissibility to maintain the dominance of privileged racial groups over racial minority groups, with an illocutionary force that casts individuals as inferior based on their race or ethnicity (Graumann & Wintermantel, 1989). Conversational exercitives are concerned with when it is permissible or conventionally legitimate in a

particular situation for a speaker to subordinate a hearer, based on situational conditions and beliefs about the speaker's rights (McGowan, 2003).

The normative conditions of exercitives likely influence the use of particular reasoning schemes to refute a racial microaggression. For instance, affirming the social identity of the person who has experienced a racial microaggression likely involves sign schemes that infer the essence of the recipient's character from his/her actions, causal schemes that infer consequences from the recipient's actions, and quasi-logical schemes that establish particular qualities of the recipient from general features (Perelman & Olbrechts-Tyteca, 1969; Warnick & Kline, 1992). Particular configurations of these argument schemes may be warranted since particular patterns of reasoning likely influence specific attributions made about recipients.

2.2 Argument, social support, and aims of the study

The increasing evidence about the impact of racial microaggressions has prompted scholars to examine the communication competencies that can disarm racial microaggressions and their effects (Sue et al., 2019). To build upon this literature, we theorize about the role of high-quality argument to provide social support about racial microaggressions. We consider supportive communication to be a communal and coordination process in which support providers and recipients mutually constitute helpful responses to problematic situations (Goldsmith, 2004).

In the context of racial microaggressions providing social support is an important form of dissent, as speakers can offer alternative ways of reasoning about the situation than the view offered by the person issuing the microaggression. While social support researchers have mostly focused on understanding support messages in terms of their ability to address a recipient's emotional distress, support messages can also employ arguments that address a recipient's identity and beliefs. Considering the context of racial microaggressions, we analyze one type of argument relevant in this context, social identity argument.

Racial microaggressions are typically viewed as representing threats to one's social identity that are associated with reduced social identity esteem (Cohen & Garcia, 2005; Sue et al., 2019; Thai et al., 2017). Thus, one function of effective argument and support in this context is to address the social identity threat by disarming negative beliefs about one's social group that are made salient by the microaggression, as well as affirming positive attributes of the social identity group. Crocker, Major, and Steele (1998) contend that making external attributions about discrimination and emphasizing the positive

aspects of one's social identity can address the social identity threats posed by discrimination incidents.

Social support in this context, then, may require restructuring cognitive attributions about one's social group and enhancing one's collective self-worth. Using reattribution theory (Weiner, 1986), social identity argument addresses social identity threats and affirms the value of the relevant social identity by changing attributions about the cause of the microaggression incident. These attributions are based on locus of control (whether the cause is internal or external), stability (whether the cause will/will not change over time), controllability (whether the cause is/is not controllable), and globality (whether the cause affects a wide/narrow range of situations). Arguments high in social identity affirmation are expected to cast racial microaggressions as external, unstable, controllable, and specific, while arguments low in social identity affirmation cast microaggressions as internal, stable, uncontrollable, and global. These argument elements also differ to the extent that they provide relevant, informative and accurate reasoning about the microaggression, which further enables message recipients to see the argument as valid and feel that their social identity worth is enhanced. Thus, a first aim of our study is to examine if the quality of social identity arguments affects recipients' evaluations of social identity arguments and their perceived social identity esteem. Specifically, we expect that:

H1: Argument recipients perceive high quality social identity arguments compared to low quality social identity arguments higher in (a) argument evaluations and (b) social identity esteem.

A second aim of the study is to determine if the effects of social identity arguments on social identity esteem are indirectly accounted for by causal reattribution and argument evaluations. Previous research on effective social support messages have found that causal reattribution mediates the effects of high-quality support on outcomes like self-esteem (Holmstrom & Kim, 2015). Message effectiveness evaluations also mediate support message quality on message outcomes (Bodie & Burleson, 2008; Bodie et al., 2012). By applying both of these lines of work to racial microaggressions, we contend that the effects of social identity arguments are related to causal reattributions and argument evaluations that are linked in a serial form to affect individuals' level of social identity esteem. This model is illustrated in Figure 1.

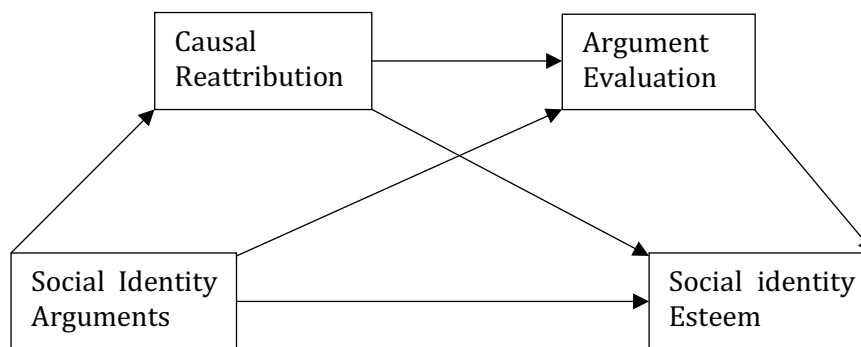


Figure 1. Serial mediation model

In the context of racial microaggressions high quality social identity arguments would enable recipients to reattribute the cause of the incident to affirm their social identity. Causal reattribution would allow support seekers to believe that racial microaggressions are caused by the perpetrator's ignorance or frame the problem as of no consequence to their social identity. Reattribution also functions as a serial mediator in that high-quality social identity arguments should engender causal reattributions that enable recipients to recognize the validity of these arguments that would produce higher social identity esteem. Overall, then, reattribution should enable high quality message arguments to influence argument evaluation and social identity esteem. Thus we expect that the effect of high-quality messages on social identity worth is mediated by reattribution and argument evaluation, with reattributions linked to higher argument evaluations that enhances social identity worth:

H2: High quality compared to low quality social identity arguments indirectly enhance recipients' social identity esteem through causal reattribution that are serially linked to argument evaluations.

3. METHOD

Participants were 256 adults (137 women, 119 men) recruited through the panel provider, Qualtrics.com, to complete an online questionnaire that included two attention filter questions. Participants were over 18 years of age, US citizens and self-identified as either Asian/Asian American ($n = 137$) or Black/African American ($n = 119$). They ranged from 18 to over 70 years of age, with 37.3% between 18-29 years, 27.8% between 30-49 years, 30.1% between 50-69 years and 5.1% who were 70 years or more.

After reporting their demographic information participants were randomly assigned to read one of three racial microaggression scenarios. Participants were cast as the target of specific discriminatory behaviors they later disclosed to a friend. They then read a message expressed by the friend whose picture was included with the message. Photographs were randomized for race and matched for gender. Participants last answered questions about the scenario and the message.

The study employed a $2 \times 3 \times 2$ between groups design that examined the quality of social identity arguments (high, low), type of racial microaggression scenarios (avoidance, derogation, categorization), and situations for each scenario type (two instances). We developed scenarios for the three types of microaggression scenarios and two situations for each scenario type; given their distinct stereotypes we created different *categorization* scenarios and situational instances for African American and Asian American participants.

After reading the microaggression scenario, participants read and evaluated one message that represented high- or low-quality levels of social identity arguments. Arguments high in social identity affirmed the recipient's social identity and reasoned that discrimination is an external problem, with the incident specific, transient and controllable. High social identity arguments used causal reasoning schemes, act-essence reasoning, and dissociation to separate the appearance of racial biases from the preferred reality of respect and achievement. Arguments low in social identity affirmation cast discrimination as an internal problem, with the incident typical and not controllable, and with behaviors generalizable to the participant's ethnicity. Low social identity arguments used schemes that fixed ethnicity as a causal factor in discriminatory behaviors. The messages and scenarios were pretested, revised, and were of similar length (100-112 words), with examples in the Appendix.

Following the message arguments participants assessed the realism of the scenario and completed measures of general argument effectiveness, social identity worth, and the mediator of reattribution. Participants completed all items on 7-point Likert scales (1= strongly disagree; 7= strongly agree). The *realism* of the discrimination scenarios was measured with three items each that have been used in previous research (e.g., "This scenario is believable"; Feng & Burleson, 2008). Responses were averaged to form measures of scenario realism ($\alpha = .91$, $M = 4.88$, $SD = .994$). One ANOVA with scenario (3), argument quality (2) and situation (2) detected no significant differences for scenario realism across situational factors (at $p < .05$).

Participants next assessed the effectiveness and outcome of the support message and the proposed mediator. *General argument evaluation* was measured with four items that assessed argument quality (e.g., the message would reassure me that my perspective on the incident is correct; “the message validated my feelings about the situation”). Averaged responses formed the measure ($\alpha = .83$, $M = 4.72$, $SD = 1.51$). *Social identity esteem* was measured with four averaged items developed for the study that measured the extent to which participants perceived that the message reaffirmed their social identity and their belief that others respected and valued their race or ethnicity (e.g., “This message would make me feel proud about belonging to my racial/ethnic group”; “This message would make me feel that others respect my race/ethnicity” $\alpha = .95$, $M = 4.84$, $SD = 1.52$). Finally, participants rated the *reattribution* facilitated by the message with three averaged items adapted from Holmstrom and Kim (2012; e.g., “This message made me think that this incident can’t affect me”; $\alpha = .78$, $M = 4.16$, $SD = 1.44$). A measurement model specified the two mediators (Reattribution, Argument Evaluation) and assessed its fit with a confirmatory factor analysis using a comparative fit index (CFI) greater than or equal to .90, and a root mean square error of approximation (RMSEA) less than or equal to .06 (Holbert & Stephenson, 2008). Model fit was acceptable, $\chi^2(11) = 24.9$, $p < .01$, CFI = .994, RMSEA = .040 (90% CI = .019, .060), after correlating the error terms of two reappraisal and argument evaluation items.

Our analyses initially utilized analyses of variance to test if high social identity arguments were evaluated differently on general argument effectiveness and enhanced social identity esteem than low social identity arguments (H1). A second set of analyses examined if causal reattribution mediated social identity arguments and argument evaluations, and that also mediated social identity arguments on social identity esteem (H2). Mediation analyses were conducted using ordinary least square path analysis with the PROCESS macro developed by Hayes (2018). Model 6 was utilized with bias corrected bootstrap confidence intervals for indirect effects based on 5000 bootstrap samples. Unstandardized coefficients and standard errors are reported.

4. RESULTS

H1 predicted that social identity affirmation arguments are positively related to argument effectiveness and social identity esteem. H1 was confirmed, as high social identity arguments ($M = 4.97$, $SD = 1.37$) compared to low social identity arguments ($M = 4.48$, $SD = 1.37$) were evaluated as having more effective arguments, $F(1, 244) = 5.175$, $p < .05$, *partial* $\eta^2 = .021$, and higher social identity esteem, $F(1, 244) =$

14.332, $p < .001$, $\text{partial } \eta^2 = .055$; $M = 5.28$, $SD = 1.40$, than low social identity arguments ($M = 4.45$, $SD = 1.53$). There were no other significant effects for scenario types, situation instances or social identity arguments in these two analyses.

H2 predicted that the quality of social identity arguments used to address racial microaggressions would enhance recipients' social identity esteem through causal reattribution and argument evaluations. As expected, causal reattribution was positively correlated with both argument evaluations (.615) and social identity esteem (.586, each $p < .001$).

H2 was confirmed, as persons who evaluated high quality social identity arguments compared to low quality social identity arguments were more likely to engage in causal reattribution ($a_1 = 0.539$ (.17), $t = 3.02$, $p < .01$), which was serially linked to higher argument evaluations ($d_{21} = 0.634$ (.05), $t = 12.04$, $p < .001$), that were linked to higher levels of social identity esteem ($b_2 = 0.588$ (.05), $t = 11.25$, $p < .001$; $adb = 0.201$; 95% CI = 0.0691 to 0.3446; $c' = 0.42$, $t = 3.32$, $p < .01$). In addition to the serial mediation effect, the quality of social identity argument was also associated with causal reattribution ($a_1 = 0.539$ (.17), $t = 3.02$, $p < .01$); which was directly associated with social identity esteem ($b_1 = 0.212$ (.05), $t = 3.871$, $p < .001$; $ab = 0.114$; 95% CI = 0.0274 to 0.2340). However, the quality of social identity argument was not associated with argument evaluations ($a_2 = 0.150$ (.15), $t = .986$, $p = .986$); so while argument evaluations were linked to social identity esteem ($b_2 = 0.588$ (.05), $t = 11.25$, $p < .001$, the indirect effect was not significant ($ab = 0.088$; 95% CI = -0.075274 to 0.2697). While the total effect of social identity argument on social identity esteem was reduced ($c_{ps} = .541$, $t = 4.48$, $p < .001$; $c'_{ps} = .276$, $t = 3.32$, $p < .01$), the direct effect remained statistically significant, which indicate that other variables related to the relationship between social identity argument and social identity esteem remain to be identified.

5. DISCUSSION

Our analysis of social identity arguments determined that high quality social identity arguments were judged to be more valid and likely to increase social identity esteem than low quality social identity arguments. We found clear evidence that differences in quality were associated with argument validity and enhanced social identity esteem with no significant variation across discrimination scenarios or situations. There was also clear evidence of a serial mediation effect for social identity arguments on social identity esteem. That is, high quality social identity arguments were more likely to invoke more causal

reattributions about the microaggression scenario, which was linked to higher argument validity and enhanced social identity esteem.

Our findings are theoretically and methodologically significant in several ways. The findings were provided by Asian-Americans and African-Americans, two groups who experience racial microaggressions. These groups provided the reasoning that when the microaggression is viewed as external to the recipient, unstable, controllable, and specific, support providers are more likely to succeed in affirming their social identities. The serial mediation findings further suggest that arguments that invite causal reattributions are important for they facilitate judgments about the validity of the arguments provided that also lead to esteem enhancement. However, whether or not it is the particular set of argument schemes or the causal attribution factors that facilitate attributions about the recipient's ethnicity is unclear and needs to be studied further.

The study's findings contribute to both argumentation theory and interpersonal communication theory. For several decades, theorists have recognized that argument theory should involve the study of argument as both product and process. Yet few theorists have focused their work on understanding argumentation in everyday interaction, describing patterns of everyday argument and analyzing ways of engaging in argument across diverse interaction tasks, modalities and relationships. The Illinois argument tradition has emphasized the need for such empirical work. This study contributes to that tradition by linking examining systematic linkages in the use of argument schemes and attributional factors as they address the aim of exercitive speech acts. The analysis of exercitive speech acts could be probed further to determine the specific felicity conditions that enact the speaker's permissibility to subordinate a hearer, which would then provide the issues or stases for arguers to employ in response and rebuttal (Kline, 1979). The findings also establish an independent role for argument evaluation as affecting argument outcomes such as enhancing a recipient's esteem. It should also be noted that the findings we obtained were consistent across multiple situations and three different types of racial microaggressions.

The findings of the study also contribute to interpersonal communication theory. While there is an important literature on interpersonal conflict with implications for the study of argument, conflict researchers have typically not incorporated argument types and patterns systematically into their research programs. Differences in argument types, evidence use and argument schemes would be valuable additions to illuminate the conflict management practices between friends, co-workers, and family members. Our findings contribute to the interpersonal conflict literature because they show that conflict

situations often call for both dissent and support, with social identity arguments designed to bring about different ways of reasoning about a situation in an effort to help the recipient targeted in the conflict situation.

As with any study future research could improve upon our design. Instead of scenarios participants could provide recalled instances of their experiences with microaggressions, or be asked to produce messages to provide support to recipients of microaggressions. Other types of arguments could be theorized and differences in their quality could be examined as ways of disarming racial microaggressions. Longitudinal designs could be employed, too.

6. CONCLUSION

Argumentation theory can advance with more pragmatic analyses of argumentation and more study about the impact of everyday argument on people's lives. In the context of racial microaggressions, effective argument includes social identity arguments that refute claims about the threatened social group and affirm the recipient's social identity. Valid arguments can make a difference for those who experience racial microaggressions, for social identity arguments that engage recipients in more causal reattribution are viewed as more valid and hence more likely to enhance recipients' self-worth and respect. Thus, social identity argument can play a role in disarming the effects of racial microaggressions.

REFERENCES

- Austin, J. L. (1962). *How to do things with words* (Cambridge, Mass.: Harvard University Press).
- Bodie, G. D., & Burleson, B. R. (2008). Explaining variation in the effects of supportive messages: A dual-process framework. *Communication Yearbook*, 32, 355-398.
- Bodie, G. D., Burleson, B. R., & Jones, S. M. (2012). Explaining the relationships among supportive message quality, evaluations, and outcomes: A dual-process approach. *Communication Monographs*, 79(1), 1-22. doi:10.1080/03637751.2011.646491
- Cohen, G. L., & Garcia, J. (2005). "I am us": Negative stereotypes as collective threats. *Journal of Personality and Social Psychology*, 89, 566-582. doi:10.1037/0022-3514.89.4.566
- Crocker, J., Major, B., & Steele, C. (1998). Social stigma. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, 4th ed.), pp. 504-553. Boston: McGraw-Hill.
- Feagin, J. R. (1991). The continuing significance of race: Anti-Black discrimination in public places. *American Sociological Review*, 56, 101-116. Retrieved from <http://www.jstor.org/stable/2095676>

- Goldsmith, D. J. (2004). *Communicating social support*. Cambridge: Cambridge University Press.
- Graumann, C. F., & Wintermantel, M. (1989). Discriminatory speech acts: A functional approach. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.), *Stereotyping and prejudice: Changing conceptions* (pp. 183–204). New York, NY: Springer.
- Hayes, A. F. (2018). *An introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. 2nd Ed. New York: The Guilford Press.
- Holbert, R. L., & Stephenson, M. T. (2008). Commentary on the uses and misuses of structural equation modeling in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 185–218). Thousand Oaks, CA, US: Sage Publications, Inc.
- Holmstrom, A. J., & Kim, S. (2015). The mediating role of cognitive reattribution and reappraisal in the esteem support process. *Communication Research*, 42, 60-86. doi:10.1177/0093650212449015
- Kauffeld, F. J. (1998). Presumption and the distribution of argumentative burdens in acts of proposing and accusing. *Argumentation*, 12, 245-266. doi:10.1023/A:1007704116379
- Kauffeld, F. J., & Fields, J. (2005). The commitment speakers undertake in giving testimony. In D. Hitchcock & D. Farr (Eds.), *The Uses of Argument: Proceedings of a conference at McMaster University, 18-21 May 2005*, (pp. 232-243). Windsor, ON: Ontario Society for the Study of Argumentation.
- Kauffeld, F. J. & Innocenti, B. (2016). Inducing a sympathetic (empathic) reception for exhortation. In P. Bondy, & L. Benacquista (Eds.). *Argumentation, objectivity, and bias: Proceedings of the 11th International Conference of the Ontario Society for the Study of Argumentation (OSSA)*, 18-21 May 2016. Windsor, ON: OSSA, pp. 1-15. <http://scholar.uwindsor.ca/ossaarchive/OSSA11/papersandcommentaries/127>
- Kline, S. L. (1979). Toward a contemporary linguistic interpretation of the concept of stasis. *Journal of the American Forensic Association*, 16, 95-103. doi:10.1080/00028533.1979.11951162
- Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, 18, 302-318. doi:10.1177/0146167292183006
- Lui, P. P., & Quezada, L. (2019). Associations between microaggression and adjustment outcomes: A meta-analytic and narrative review. *Psychological Bulletin*, 145, 45-78. doi:10.1037/bul0000172
- McGowan, M. (2003). Conversational exorcistics and the force of pornography. *Philosophy and Public Affairs*, 31, 155-189.
- Mossakowski, K. N., & Zhang, W. (2014). Does social support buffer the stress of discrimination and reduce psychological distress among Asian Americans? *Social Psychology Quarterly*, 77(3), 273-295. doi:10.1177/0022062214534271

- Paradies, Y., Ben, J., Denson, N., Elias, A., Priest, N., Pieterse, A., & Gee, G. (2015). Racism as a determinant of health: A systematic review and meta-analysis. *PLoS ONE*, 10, e0138511. doi:10.1371/journal.pone.0138511
- Pascoe, E. A., & Smart Richman, L. (2009). Perceived discrimination and health: a meta-analytic review. *Psychological Bulletin*, 135, 531-554. doi:10.1037/a0016059
- Perelman, C., & Olbrechts-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation*. Notre Dame: Univ. of Notre Dame Press.
- Sue, D. W. (2010). *Microaggressions in everyday life: Race, gender, and sexual orientation*. Hoboken, NJ: John Wiley & Sons.
- Sue, D. W., Alsaidi, S., Awad, M. N., Glaeser, E., Calle, C. Z., & Mendez, N. (2019). Disarming racial microaggressions: Microintervention strategies for targets, White allies, and bystanders. *American Psychologist*, 74(1), 128-142. doi:10.1037/amp0000296
- Sue, D. W., Bucceri, J., Lin, A. I., Nadal, K. L., & Torino, G. C. (2007). Racial microaggressions and the Asian American experience. *Cultural Diversity and Ethnic Minority Psychology*, 13(1), 72-81. doi:10.1037/1099-9809.13.1.72
- Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cambridge, England: Cambridge University Press.
- Thai, C. J., Lyons, H. Z., Lee, M. R., & Iwasaki, M. (2017). Microaggressions and self-esteem in emerging Asian American adults: The moderating role of racial socialization. *Asian American Journal of Psychology*, 8(2), 83-93. doi:10.1037/aap0000079
- Thoits, P. A. (2013). Self, identity, stress, and mental health. In C. Aneshensel, J. Phelan & A. Bierman (Eds.), *Handbook of the sociology of mental health* (2nd ed., pp. 357-377). Cham, The Netherlands: Springer.
- Turner, J. C. (1999). Some current issues in research on social identity and self-categorization theories. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity: Context, commitment, content* (pp. 6-34). Oxford, UK: Blackwell.
- Warnick, B., & Kline, S. L. (1992). The New Rhetoric's argument schemes: A rhetorical view of practical reasoning. *Argumentation and Advocacy*, 28, 1-14. doi:10.1080/00028533.1992.11951551
- Weiner, B. (1986). Attribution, emotion, and action. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (pp. 281-312). New York: Guilford.

APPENDIX

Examples of Social Identity Arguments for addressing Racial Microaggressions

High Social Identity Affirmation: You know this is not about you, right? It's not your fault that some people can't get past your ethnicity. Other people and I respect you for who you are. As a member of your ethnicity, you have so many achievements in your personal and professional life to be proud of. So, incidents like these won't harm your strong heritage and rich culture. Maybe if we explain to people that their actions can be seen as offensive, they might not

act that way again. I think things are changing because people are becoming more aware of such biases. Incidents like these might become less frequent in the future.

Low Social Identity Affirmation: Unfortunately, people tend to make so many assumptions about you because of your ethnicity. That's why some people might not sit next to you on the bus or socialize with you in other situations. I know that you can't easily succeed at work or even expect great service at a restaurant because of your ethnicity. People treat you differently because of your ethnicity all the time. But we can't change everyone's mindset and make them treat everyone equally. People will continue to behave like this because such biases don't go away quickly. I think such incidents will continue to happen in the future.

Holism of reasons and its consequences for argumentation theory

HUBERT MARRAUD

Universidad Autónoma de Madrid

hubert.marraud@uam.es

I contend that a suitable theory of argument should be reasons-based instead of inference-based. I first explore the consequences of this shift for the distinctions linked/convergent arguments and coordinate/multiple argumentation. I suggest that the former is a distinction between single reasons and many reasons, while the latter is a distinction between different kinds of reasons composition. Second I show that the holistic concepts of modifiers and conditions provide more fine-grained accounts of these argument structures.

KEYWORDS: argument structure, attenuators, coordinate vs multiple argumentation, disablers, enablers, holism, intensifiers, linked vs convergent arguments, reasons.

1

I will consider three related, though different distinctions, viz.: linked vs convergent arguments, coordinate vs multiple argumentation, and grounds vs modifiers. The first two come from the theory of argumentation while the latter comes from the theory of normative reasons (or theory of reasons, for short). It could be expected that there would be a smooth flow of information between argumentation theory and the theory of reasons, since — ignoring complexities arising from con arguments, not to be considered here— arguing is to present to someone something as a reason for something else. But in fact these fields have developed in mutual ignorance. My general aim is to connect these disciplinary perspectives for mutual benefit.

The main consequence of building up the theory of argument on the notion of reasons is that argument becomes a weighted notion since reasons are paradigmatic weighted notion. If a good argument is indeed the one that presents a good reason for its conclusion and reasons have

weight or strength, so too do arguments. Thus before concluding anything one has to compare the relative strength of the various relevant arguments. So weighing lies at the heart of any reasons based theory of argument.

The linked/convergent distinction marks the difference between different ways in which a group of premises lends support for the conclusion, the coordinate/multiple distinction has to do with various ways of combining a number of arguments with the same conclusion into a single argumentation, and finally the grounds vs modifiers distinction refers to different roles a premise can play in the constitution of a reason. To clarify these distinctions one has to explicate the related concepts of premise, argument and reason. My starting point will be the above definition of arguing:

- arguing is presenting to someone something as a reason for something else.

As a product of the act of arguing an argument will then have two minimal constituents, the “something” and the “something else” in the definition. For manifestly presenting something as a reason for something else we use conventional devices such as the disposition of statements in the text, punctuation marks, pauses and intonation schemes, argumentative connectives and operators, and explicit meta-argumentative vocabulary. We thus arrive at the canonical representation of an argument in logic: $A_1, \dots, A_n \text{ so } C$ with the variables ranging over statements. Statements put forward to offer reasons are called “premises”, while the statement for which they are offered is the claim or conclusion.

On the other hand, reason is a normative concept. The standard definition of a (pro) reason is a consideration that favors a position, taking a “position” to be an attitude towards a proposition, an action or policy recommendation, or an evaluation (Blair 2012:148). Of course the mere fact of presenting something as a reason for something else does not ensure that it actually is. Likewise inserting the connective “therefore” before the last term in a sequence of statements doesn’t make the former probatively relevant for the latter, but expresses the speaker’s commitment that they are. Francesco Sizzi in his *Dianoia Astronomica* (1611), offered the following argument:

There are seven windows in the head, two nostrils, two ears, two eyes and a mouth; so in the heavens there are two favorable stars, two unpropitious, two luminaries, and Mercury alone undecided and indifferent. From which we gather that the number of planets is necessarily seven.

Even if the number of windows in the head and the alleged features of the Sun, Earth, Moon and the outer planets is not actually a reason to believe that the number of planets is necessarily seven, there is no doubt that Sizzi was offering an argument, and consequently that these statements were premises.

Bad arguments are still arguments. I am not claiming that an argument is made of a set of premises that jointly give a reason for the conclusion. Rather using an argument amounts to creating a *prima facie* reason. A *prima facie* reason is something that appears to be a reason, but may actually not be a reason at all. What makes *A* appear as a reason in *A so C* is that the word “so” carries an assumption of inferential relevance. After all, relevance is not a property of statements but rather a presupposition of utterances (Sperber & Wilson 1995). *Prima facie* reasons should not be confused with *pro tanto* reasons: a *pro tanto* reason is a consideration that counts in favor of some position, and thus has genuine weight, but nonetheless may be outweighed by other considerations. In brief, premises are always relevant, in the sense of expressing a *prima facie* reason for the conclusion. However argument’s cogency requires that the premises offer a *pro tanto* for the conclusion.

2

There is no 1-1 correspondence between the premises of an argument, understood as separate statements, and the reasons the argument offers for its conclusion. Sizzi’s argument, for instance, has two premises but it is intended to convey a single reason. A single reason can be conveyed by one or more statements, and one reason arguments can be either one premise or many premises. How should reasons be individuated? Blair’s answer is that “A single reason is the smallest amount of information that by itself lends some measure of credence to a position.” (2012:148). Although Blair definition leaves room for complex or compound reasons, he does not mention them anymore. Otherwise most current definitions of “argument” acknowledge that there are many reasons arguments; e.g., “By ‘argument,’ we mean a claim, together with one or more sets of reasons offered by someone to support that claim” (Johnson & Blair 1994:10). Accordingly we distinguish single premise arguments and many premises arguments, on one side, and arguments offering a single reason and arguments offering two or more reasons, on the other.

This raises the question of how to distinguish single reason many premises arguments from many reasons arguments. Argumentative connectives like “moreover” or “besides” provide a

linguistic test, since they are used to combine different reasons (not just different premises).¹

- (1) This cake contains chocolate and Mary is allergic to chocolate, so she should refrain from tasting it.

In a normal context, this a single reason two premises argument; this is why the following paraphrase sounds weird:

- (2) This cake contains chocolate, besides Mary is allergic to chocolate, so she should refrain from tasting it*

By contrast “besides” can be soundly inserted into a many reasons argument like

- (3) This coat costs more than I intended to spend and it is not exactly what I was looking for; so, I will not buy it

Yielding

- (4) This coat costs more than I intended to spend, besides it is not exactly what I was looking for; so, I will not buy it

If “besides” can be taken to combine two statements into a single one (as do sentential connectives “and”, “or”, etc.), it could be thought that argument (4) is many reasons but single premise. The alternative is to see argumentative connectives as “besides”, “moreover”, “but”, etc. as introducing a relation between arguments, or better said between linguistic units insofar as, in the given context, they identify or refer to arguments. Along these lines (4) should be interpreted as

- (4a) (This coat costs more than I intended to spend, so I will not buy it) besides (This coat is not exactly what I was looking for, so I will not buy it)

Instead of being interpreted as

¹ Hitchcock points out that “There is also an interpretive difficulty in determining whether an additional supporting reason introduced by a bridging term like ‘besides’ or ‘moreover’ or ‘further’ is a new argument or merely an independently relevant part of a single argument” (2017:24). My answer is that “besides” and the like introduce a new argument or a modifier, whose insertion in the given argument produces a modified reason (see § 4 below).

(4b) (This coat costs more than I intended to spend, besides this coat is not exactly what I was looking for), so I will not buy it.

In (4a) the statements “This coat costs more than I intended to spend” and “This coat is not exactly what I was looking for” are the premises of two different, though cooriented arguments, while in (4b) they are the premises of a single argument. (4b) represents the dominant reading in informal logic, while (4a) represents the dominant reading in pragma-dialectics and French linguistics.

3

The different readings of (4) help us to understand the differences between the linked/convergent distinction and the coordinative/multiple distinction. Both (1) and (4b) are two premises arguments; however the premises lend support to the conclusion in different ways. In (1) the two premises link to each other to produce a single and simple reason, while in (4b) they do so to produce a compound reason. Hence, in (1) premises are linked while in (4b) they converge. Summing up, from the point of view of informal logic both linked and convergent arguments are many premises arguments expressing a unique reason, the difference being whether this reason is simple or compound. On this picture the complexity of an argument is determined by the complexity of the reason offered to consideration. This is consistent with most standard accounts of the linked vs convergent distinction:

When there are several premises in an argument, those premises support the conclusion together and will have to be considered together when we come to appraise the argument. In the linked pattern of support, the premises are interdependent in the way they support the conclusion; if we did not consider them together, they could provide no support at all. In the convergent pattern of support, on the other hand, one premise alone could provide some support to the conclusion, but the various premises, together, are intended to cumulate so as to offer more support (Govier 2010:54-55).

These cases should be distinguished from cases in which several reasons for the same position are offered —a situation that many informal logicians identify with multiple argumentation. When many reasons are presented for choice, intending that the chosen reason be the sole reason for accepting the conclusion, it can be alleged that no compound reason is produced, and thus that the premises don't support

the conclusion together. Freeman makes this point saying that “The multiple-coordinatively compound distinction is dialectical, whereas the linked convergent distinction is logical.” (Freeman 2011:109). For Freeman the unit of logical analysis is the individual argument, while the unit of dialectical analysis is the entire argumentation made up of several arguments. In the same line, Hitchcock holds that “From the pragma-dialectical perspective, the linked-convergent distinction is a distinction within the class of coordinatively compound argumentation. (2017:24).

When reasons are identified with single reasons, as I guess pragma-dialecticians do, the difference between (1) and (4) is that the former conveys a single reason while the second conveys several reasons. The coordinative vs multiple distinction concerns the way several different reasons can be combined:

The distinction between coordinative and multiple argumentation is therefore not that coordinative argumentation describes the relations between premises within one single argument and that multiple argumentation consists of a combination of single arguments, but that the relations between the single arguments that constitute these two types of complex argument are different. (Snoeck-Henkemans 2000:460). Here the underlying criterion of complexity is the number of reasons conveyed by an argument. In the simplest cases, in multiple argumentation several alternative reasons are offered for the same conclusion, while in coordinative argumentation these several reasons constitute a joint defense of the conclusion. Therefore it can be said that a multiple argumentation offers a disjunction of reasons, as in

(5) Either (This coat costs more than I intended to spend, so I will not buy it) or (This coat is not exactly what I was looking for, so I will not buy it)

and a coordinative argumentation offers a conjunction of reasons, as in

(6) (This coat costs more than I intended to spend, so I will not buy it) and (This coat is not exactly what I was looking for, so I will not buy it)

(5) and (6) express different commitments. The relevant differences can be explained counterfactually. Using (5) the speaker says that she would buy the coat it were either cheaper or more akin to her preconceived idea. By contrast, using (6) she says that she would only buy it were both cheaper and similar to her preconceived idea.

Normative reasons can be defined as considerations that count in favor or against some commitment. In the theory of reasons holism is the claim that contexts differ in terms of whether a certain consideration constitutes a reason at all, as well as in terms of the weight and (possibly even) polarity of the reason. Therefore, any reason –and hence any argument– has to be evaluated in the context of all relevant reasons that apply in a given situation.

To explain how reasons can vary across contexts, Bader (2016), drawing upon the work of Dancy (2004), introduces a distinction in the “necessitation base (that which explains and necessitates the reason) between the source or the ground of a reason, the conditions of the reasons and the modifiers of the reason. Bader gives two non-equivalent definitions of the source or ground of a reason: “that in virtue of which something is a reason –the source of the reason” (2016:282), “the source or ground of a reason is to be identified with the consideration that constitutes the reason” (*Op.cit.* 6). In the Toulmin model the source or ground in the first sense is similar to the warrant, while in the second it corresponds to the data. Henceforth I will reserve the term “ground of a reason” to designate the consideration that constitutes the reason, using “warrant of a reason” to designate that which makes something a reason for something else.

The conditions of the reason are that on condition of which something is a reason. There are two kinds of conditions: enablers and disablers. If the conditions are satisfied (which can consist in the presence of enablers or the absence of disablers), then the ground does constitute a reason. Otherwise, if the relevant enablers are absent or disablers are present, then it will fail to do so.

Finally modifiers of the reason are considerations that affects the weight of a reason. There are also two kind of modifiers: intensifiers and attenuators. Intensifiers are facts that make the weight of some reason greater without themselves being reasons, while attenuators are facts that make the weight of some reason weaker without themselves being reasons.

An example, built from one in Dancy (2004: 38) will give us some taste of all these concepts.

1. I promised to do it.
2. My promise was not given under duress.
3. I am able to do it.
4. Doing it would not be too costly for me.
5. So: I will do it.
6. Since: We ought to keep our promises.

1 is the ground for claim 5. If, given the argument *1 so 5* someone asks “How do you get there?”, a possible answer would be 6. Hence 6 is the warrant explaining why 1 is relevant to 5. Neither 2 nor 3 are by themselves reasons for doing the action. 2 is an enabling condition consisting in the absence of a disabler, and 3 is an enabling condition consisting in the presence of an enabler, and hence of a different type as 2. Nor is the mere fact that doing it would not be too costly for me a reason for doing that action. Even if it were costly for me doing what I promised, I would still have a reason for doing that, a reason given by my promise to do so. But 4 does make a rational difference, all the same. What it does is to intensify the reason given me by 1. Instead of two reasons, according to Bader, what we have here is one ground or source 1 and one intensifier 4.

It can be objected that the addition of warrants is at odds with particularism, a thesis that often goes hand in hand with holism in the theory of reasons. Particularism is the thesis, espoused, among others, by Dancy, that moral reasoning can dispense with moral principles at all. Normally, particularists are holists and generalists are atomists. In any case, appeal to warrants is compatible with weak holism inasmuch as warrants are not constituents of an argument on equal terms with premises and conclusion, but another way answering the question “How do you get there?”, not the only one. Crisp (2007) distinguishes two forms of holism, weak and strong. Both forms agree that a feature that is a reason in one case may be no reason at all, or an opposite reason, in another. But strong holism draws the conclusion from this holistic tenet that there can be no general principles stating the reason-giving status of any such feature, while weak holism allows that there will be such a principle for any invariably reason-giving feature and that there may be such features.

5

Bader’s holistic framework, duly adapted to the theory of argument, allows for the discovery, description and evaluation of new argument structures, besides the ones identified in Snoeck-Henkemans (2000).

When modification occurs, there are two distinct reasons. On the one hand, there is the unmodified reason and, on the other, the modified reason (Bader 2016:15). Thus, if the identity of an argument depends on the conveyed reason, in our previous example two different arguments should be distinguished; the unmodified argument:

(7) I promised to do it; so, I will do it.

And the modified argument:

(8) I promised to do it. Doing it would not be too costly for me; so I will do it.

Since intensifiers “are facts that make the weight of some reason greater without themselves being reasons” (Lord & Maguire 2016:11), some might think that (8) is stronger than (7). That would be a mistake. If we take holism seriously arguments are to be evaluated in context. In a situation in which doing the promised thing is not too costly for me, both arguments have the same force or strength; otherwise, the second argument would have an unacceptable premise, in which case strength comparison would be pointless. We have to distinguish carefully intensifiers₁, i.e. facts that strengthen a reason, from identifiers₂, premises expressing intensifiers₁. It is the fact that doing the promised would not be too costly for me what makes greater the weight of (7), not the addition of the identifier₂ “Doing it would not be too costly for me”.

The need of distinguishing intensifiers₁ from identifiers₂ is even clearer when we consider attenuators. Here is another example from Dancy (2004: 42):

1. She is in trouble and needs help.
2. It is all her own fault, and she got in this situation through trying to spite someone else.
3. But still: I will help her.

Here 1 is the ground, 3 the claim and 2 is an attenuator. Correspondingly we have two distinct arguments:

(9) She is in trouble and needs help; so, I will help her

(10) She is in trouble and needs help. It is all her own fault, and she got in this situation through trying to spite someone else; but still, I will help her

The idea that inclusion of an attenuator in (10) would make this argument weaker than (9) clashes with the intuition that in a given situation the stronger argument is the one that should guide our decision. I assume that, in a situation in which it is a fact that it she got in this situation through trying to spite someone else, the guiding argument will be (10), supposedly the weaker argument.

Grounds, conditions and modifiers can be identified with different kinds of premises. For Dancy they correspond to “three sorts of role that a relevant consideration can play”, “which just shows that there is more than one way of being relevant—more than one form of relevance, as one might say.” (2004: 42). If Dancy is right, and there are

different kinds of premises and accordingly different senses of relevance, tests for making the linked vs convergent distinction, and notably Walton's five tests (Walton 1996: 119–120, 127) must be drastically revised.

6

Although the very notion of modifier belongs to the theory of reasons, we can find some inklings of it in argumentation theory, notably in Pinto and Blair's notion of complementary premises and Vorobej's notion of supplementation.

Walton (1996:133–134) tells us that Pinto and Blair, in an unpublished work cited by Snoeck-Henkemans (1992), distinguish a third kind of argument, besides linked and convergent arguments. Pinto and Blair define a complementary argument as one in which "Some premises complete others". To illustrate what Pinto and Blair mean by "some premises completing others", Walton (*Op.cit.*, 134) cites the following argument:

1. I promised my girlfriend I'd take her to see the latest Woody Allen movie tonight.
2. She'll be really disappointed if I don't go to that movie with her.
3. I don't have any excuse for not doing so.
4. So: I guess I should take her to see that movie tonight.

Pinto and Blair reason that this is not a linked argument for in a linked argument no single premise by itself supports the conclusion, as 1 does. It is not a convergent argument for neither 2 nor 3 directly support the conclusion, but in a convergent argument each premise alone lends some support to the conclusion, that gets stronger and stronger with each additional premise. Thus, they conclude, this is a different argument pattern: a complementary argument. Notice that in the terminology of the theory of reasons, 1 is the ground or source of the reason, 2 is an intensifier and 3 is an enabler. In Blair and Pinto's words, as reported by Walton, premises 2 and 3 "complete" premise 1, increasing the amount of support. Thus modifiers and conditions are complementary premises, and a complementary argument seems to be arguments in which some modifier (or some condition) occurs.

Vorobej (1995:292) gives the following definition of supplementation for two premises arguments,:

- A premise P supplements a premise Q iff
1. P is not relevant to C,
 2. Q is relevant to C, and

3. {P,Q} offers an additional reason R in support of C, which Q alone does not provide.

Therefore the supplementation relation is a kind of asymmetric linkage. A two premises argument with a supplementary premise is neither linked nor convergent. To be convergent, every premise should be directly relevant to C, but P is not. To be linked, no proper subset of {P,Q} should be relevant to C, but {Q} is. Thus we are facing another pattern of argument. Vorobej example is:

1. All the ducks that I've seen on the pond are yellow.
2. I've seen all the ducks on the pond.
3. So: All the ducks on the pond are yellow.

Given that 2 is a supplementary premise, this is in Vorobej terminology a "hybrid argument" (1995:293). Here 1 is the ground while the supplementary premise 2 turns out to be a modifier. In general, since modifiers by themselves are not reasons, they are supplementary premises, and thus any argument offering a modified reason will count as a hybrid argument.

7

I have said little about conditions. However conditions can also be detected in argumentation theory. Disabling conditions are closely related to Toulmin's conditions of exception or rebuttal, as the following passage makes it clear.

The special force of the qualifier used in this second type of case (presumably) is directly connected with the idea of rebuttals. It registers the fact that the inference is warranted—that the claim is directly supported by the grounds-only in the absence of some particular exceptional condition, which would undercut (i.e., withdraw the authority of the warrant for) the inference. (Toulmin, Rieke & Janik 1984:96).

Let us consider an example similar to the one in Toulmin, Rieke & Janik 1984: 96-98.

1. This patient has a streptococcal infection.
2. So: this patient needs penicillin treatment.
3. Since penicillin may be safely and effectively prescribed for streptococcal infections.
4. But this patient is allergic to penicillin.

1 is the ground or source of the reason, 2 is the claim, 3 is the warrant making the ground 1 relevant to the claim 2, and 4 is a disabler or condition of rebuttal, as explicitly pointed in Toulmin, Rieke and Janik (1984). On the holist account, when the patient is allergic to penicillin, the fact that she has a streptococcal infection is no longer a reason for penicillin prescription. A Toulminian explanation is that since it is the warrant what makes the datum or ground a reason for the claim, if some particular exceptional condition withdraws the authority of the warrant, no reason remains.

Bader emphasizes that a disabler is not equivalent to an extreme attenuator. When an attenuator is applied, he argues, the unmodified reason remains, and it is only the modified reason that is "reduced to zero", whereas in the case of disablers there is no reason at all (Bader 2016:37).

Although it is more difficult to find an analogous to enabling conditions on the Toulmin model, Toulmin, Rieke & Janik describe the process through which disabling conditions give way to enabling conditions:

... if the situation develops to a point at which no standing presumption can be securely established in the first place, there will be nothing left to rebut. Instead, we shall then have to work with two alternative parallel arguments and apply one or the other of them in any particular case, depending upon which of the alternative conditions holds good: "On the one hand, if the patient is not allergic to it, penicillin may be safely and effectively prescribed for upper respiratory infections. On the other hand, in cases of penicillin sensitivity, some other broad-spectrum antibiotic should be prescribed, such as tetracycline."

That is, where the "exceptions" are not truly exceptional, we cannot present the conclusions of our arguments as being "presumably" sound, subject only to a possible rebuttal. Instead we do better to restate our warrants, explicitly, as holding good only on condition that certain specific conditions are satisfied. (Toulmin, Rieke & Janik 1984: 99).

Thus when exceptions become frequent, the above argument must be reformulated as follows, incorporating an enabling condition 4:

1. This patient has a streptococcal infection.
2. So: this patient needs penicillin treatment.
3. Since penicillin may be safely and effectively prescribed for streptococcal infections.
4. Given that this patient is not allergic to penicillin.

I have attempted to show the mutual relevance of the theory of reasons for the theory of argumentation. On one side, the theory of argumentation provides an explanation of what is to be a *prima facie* reasons: *prima facie* reasons are considerations presented as reasons in argument. I would like to suggest further that *pro tanto* reasons result from *prima facie* reasons through a process of critical discussion. Thus reasons are not something given in some *a priori* logical space, but something built up in the course of argumentative exchanges, of the game of giving and asking for reasons.

On the other side, the theory of reasons may provide the theory of argumentation with a fine-grained classification of premises and forms of relevance. Although this contribution of the theory of reasons can improve our understanding of argument structures, the concepts of intensifier, attenuator, enabler and disabler are not completely alien to argumentation theory, since they have been somehow anticipated by authors like Pinto, Blair, Vorobej or Toulmin. When argumentation theory is based in a holistic notion of reasons, argument becomes a weighted concept. The main lesson for argumentation theorists to be drawn from holism in the theory of reasons is that argument evaluation, even logical evaluation, is always context dependent.

ACKNOWLEDGEMENTS: This work was supported by FEDER/Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación, as part of the project PGC2018-095941-B-I00 Prácticas argumentativas y pragmática de las razones (Parg_Praz).

REFERENCES

- Bader, Ralph (2016). "Conditions, Modifiers and Holism". In Lord & Maguire (2016): 27-55.
- Blair, J. Anthony (2011). *Groundwork in the Theory of Argumentation*. Dordrecht Heidelberg London & New York: Springer.
- Crisp, Roger (2007). "Ethics without Reasons?". *Journal of Moral Philosophy* 4(1): 40-49.
- Dancy, Jonathan (2004). *Ethics without Principles*. Oxford: Oxford University Press.
- Freeman, James B. (2011). *Argument Structure: Representation and Theory*. New York: Springer.
- Govier, Trudy (2010). *A Practical Study of Argument*. 7th edition. Belmont, CA: Wadsworth.

- Hitchcock, David (2017). *On Reasoning and Argument: Essays in Informal Logic and on Critical Thinking*. Dordrecht Heidelberg London & New York: Springer.
- Johnson, Ralph H. (2000). *Manifest Rationality*. Mahwah, NJ: Lawrence Erlbaum.
- Johnson, Ralph H. & Blair, J. Anthony (1994). *Logical Self-Defense*. New York: McGraw Hill.
- Lord, Errol & Maguire, Barry, eds. (2016). *Weighing Reasons*. New York: Oxford University Press.
- Snoeck-Henkemans, A. Francisca (1992). *Analysing Complex Argumentation*. Amsterdam: SICSAT.
- (2000). "State-of-the-Art: The Structure of Argumentation". *Argumentation* 14: 447–473
- Sperber, Dan & Wilson, Deirdre (1995). *Relevance. Communication & Cognition*. 2nd edition. Oxford: Blackwell.
- Toulmin, Stephen E., Rieke, Richard. & Janik, Allan (1984). *An Introduction to Reasoning*. 2nd edition. New York: McMillan.
- Vorobej, Mark (1995), "Hybrid Arguments", *Informal Logic* 17, 289–296.
- Walton, Douglas N. (1996), *Argument Structure: A Pragmatic Theory*, Toronto, ON: University of Toronto Press.

“Why did not your correspondent make an honest inquiry before so writing?”: The text structure and discourse of disagreement in Irish letters to the editor

DAVIDE MAZZI

University of Modena and Reggio Emilia (Italy)
davide.mazzi@unimore.it

This paper looks at disagreement in letters to the editor published by national newspapers in late-nineteenth/early-twentieth-century Ireland. The research was conducted in the form of a qualitative analysis of argumentative discourse in letters, in many of which disagreement was expressed. Overall, findings provide evidence of recurrent discourse sequences of disagreement in news settings from a period where an Irish public opinion was to gradually shape up and legitimise ordinary people's right to dissent.

KEYWORDS: Argumentation; Disagreement; Text; Discourse; Letters; Newspapers; Public opinion; Ireland.

1. INTRODUCTION

There is much evidence that disagreement has been discussed in a variety of settings and from a wide range of perspectives. On the one hand, a number of studies have focused on disagreement in contexts characterised by ostensibly formal constraints such as politics or the judiciary. This is shown by works on the institutionalised management of disagreement in parliamentary settings (Ilie, 2010; Robles, 2011) as well as constitutional debate (Shiffman, 2002), and research comparing the conventionalized politeness strategies in English and American judicial opinions arguing against previous decisions or judges sitting on the same bench (Kurzon, 2001).

On the other hand, disagreement has been investigated in less formalised settings instantiated by public debate. Thus, for instance, Nir (2011) deals with different forms of disagreement and opposition in social networks, in order to assess the effects of supportive, mixed and oppositional discussion networks on the likelihood of political participation in the US. Furthermore, Uzelgun et al. (2015) examine data from the debate on climate change, showing the relationship between

concessive constructions and participants' attempt to present themselves as reasonable agents who comply with their dialectical obligations in an implicit critical discussion, by accepting what represents possible common ground and justifying in what respects the contradiction with the opponent is raised.

Moreover, a number of works have been devoted to the peculiarities of polemical discourse. To begin with, this has been dealt with as a form of interaction in dialogic spaces. In its capacity as a *contrario* discourse, polemical discourse displays marked dialogism. Even when the interlocutor is not physically present – as with a facebook post targeting a politician – polemical texts appear narrowly focused on one or more opponents, whose discourse they reject and discredit (Kerbrat-Orecchioni, 1980; Angenot, 1982; Garand, 1998; Amossy, 2009). In addition, polemical discourse has been examined in relation to patterns of argumentation through the media. In this regard, polemical discourse has more recently been described as verbal interaction where complex procedures are established in order to negotiate antagonistic viewpoints (Amossy & Burger, 2011; Amossy, 2011; Burger, 2011; Jacquin, 2011).

In this vein, the aim of this paper is to study the text structure and discourse of disagreement in letters to the editor published by national newspapers in the context of late nineteenth- / early twentieth-century Ireland. This is agreed by many authors to have been a key period in the development of Irish journalism for several reasons. These include the loosening of Government censorship, the greater technological expertise the Irish newspaper industry could benefit from at the turn of the century, a sharp increase in literacy rates and the circulation of newspapers and periodicals, and an increasing readership along with the gradual formation of a public sphere in its own right (Dunlop, 1911; Glandon, 1985; Legg, 1999; Morash, 2010; Rafter, 2011; O'Brien & Rafter, 2012; Steele & de Nie, 2014).

The rest of the paper is organised as follows. In Section 2, corpus design criteria are discussed, and the methodological tools are introduced: this will allow for a presentation of the dataset as well as a preliminary review of the procedure(s) through which the data were studied. Section 3 then presents the findings of the study, which are eventually discussed in the light of the relevant literature in Section 4.

2. MATERIALS AND METHODS

The study was based on a section of *Éirnews*, a small corpus of 115 news texts published between 1895 and 1905 by four national newspapers, i.e. *Belfast Newsletter*, *Irish Examiner*, *Freeman's Journal* and *Irish Independent*. The decade behind the *Éirnews* corpus was selected in

order to assemble a core collection of news texts capable of capturing the essence of contemporary press coverage about the west of Ireland, the territory that commanded most of scholarly as well as public attention (Walsh, 2008; Mazzi, 2019), between the new phase in Irish politics after Charles S. Parnell's death, and the opening stage of twentieth-century nationalism marked by the establishment of the *Irish Independent* in 1905. The corpus texts were downloaded from the official website of the Irish Newspaper Archives,¹ the world's largest and oldest online database of Irish newspapers. Overall, *Éirnews* included specimens of three news genres: news reports, editorials and letters to the editor. As far as this study is concerned, the focus was on the 33 letters in the corpus.

From a methodological point of view, the research adopted a primarily descriptive and data-driven approach. In the search for suggestive patterns, therefore, data were "not adjusted in any way to fit the predefined categories of the analyst", while "recurrent patterns and distributions" were expected "to form the basic evidence for linguistic categories" (Tognini Bonelli, 2001, p. 84). The study was carried out within the broad framework of historical discourse analysis (Brinton, 2001), which has been fruitfully applied to historical news analysis (Brownlees, 2009, 2016a). In particular, a systematic qualitative analysis was performed of any salient patterns of text and discourse structure underpinning the writers' argumentation in those letters where disagreement was voiced (Paltridge, 1996; Amossy, 2011; Burger, 2011; Brownlees, 2016b).

The rich harvest of qualitative analysis in association with news text has been reaped in more than a contribution over the past few years. Accordingly, Fürsich (2009, p. 240) argues that the method involves serious engagement with the chosen texts through genre and rhetorical approaches, and it enables one to successfully "focus on the underlying ideological and cultural assumptions of the text". In her view, the meticulous reading and contextualised interpretation of the news text is ultimately suited "to textually derive a particular media content's unique [...] conditions of production and how [...] text positions audiences in specific [...] ways" (Fürsich, 2009, p. 248).

3. RESULTS

3.1 Disagreeing in letters: a prototypical text structure

Moving on to the findings of the study, it seems appropriate to begin with a brief remark about the letters as a whole. In her study on

¹ <https://www.irishnewsarchive.com> (last accessed 1 July 2019).

polemical discourse, Kerbrat-Orecchioni (1980, p. 12) defines it as ‘disqualifying discourse’ [*un discours disqualifiant*] aiming at a ‘target’ [*une cible*]. The main targets attacked in our letters include, first and foremost, the Congested Districts Board, namely the body established in 1891 to combat poverty and alleviate congested living conditions in the west and north-west of Ireland (Breathnach, 2005, p. 11); secondly, the British Government; thirdly, the Chief Secretary for Ireland as the de facto government minister with responsibility for governing Ireland (Kee, 2000 [1972]); fourthly, other individuals such as journalists and priests.

It is worth mentioning that one type of text structure was noted with amazing regularity in association with the discourse of disagreement with the individuals or institutions mentioned above. This is schematised in Figure 1 below:

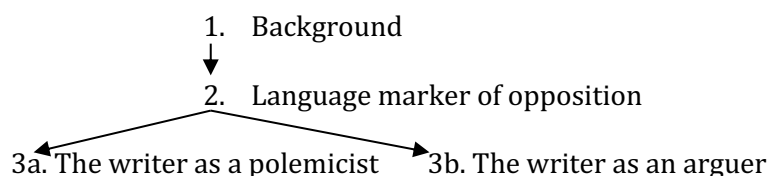


Figure 1 – Recurrent text structure in corpus letters

The first stage in the structure is of a discretionary kind. It is represented by the writers’ deliberate intention to provide readers with the necessary background, in order to let them understand the issue they are going to discuss. This may involve either a short account of the incidents that sparked the writers’ outcry, or else a brief restatement of the very words of the person or people engaged by writers in critical debate. The second stage marks the onset of disagreement, signalled by what Burger (2011) calls *marques langagiers d’opposition* [‘language markers of opposition’]. In third place, two interrelated discursive identities are forged for writers. The first is that of writers as polemicists, who level criticisms at the intended target(s) in their capacity as representatives of a dispreferred standpoint. The second is that of writers as arguers supporting a standpoint of their own while at the same time refuting the target’s own standpoint. This is again in keeping with Burger’s (2011) views on polemical disagreement, whereby he points to two ‘complex identities unfolding in interaction’ [*des identités complexes se construisant dans l’interaction*].

In order to describe what this looks like in practice, it might be useful to draw the attention on an example from the corpus. On 27 December 1900, the *Freeman’s Journal* published an article with the headline “Connemara Terra Incognita”. Essentially, the text read as a

critical overview of education in the west of Ireland, where the journalist was very dismissive of the teaching and learning standards of one school in particular. This was the National School in Lettermullen, a remote island off the coast of Co. Galway. As can be seen from the passage reported in (1) below, the school's Principal Teacher is depicted as not knowing a word of Irish. From the journalist's perspective, this can only contribute to stamping out Ireland's true national language from the territories along the western seaboard:²

- (1) But it is sad to see the attempts that are being made to stamp out the National Language in these remote places, for the largest attended school in the three islands or on the neighbouring mainland is taught by a schoolmistress who does not know a word of Irish. The children in most cases learn nothing in such schools.

As the schoolmistress referred to by the journalist, Mrs Ellen Healy predictably reacted to the article by sending a letter to the *Freeman's* editor. This was duly published on 10 January 1901 with the headline "The Terra Incognita of Connemara and a lady teacher in the West". An extract from the first part of Mrs Healy's letter is reported in (2) below:

- (2) In your issue of Thursday [...], the writer gives a very glowing account of the backward character of, among other places in the West, the island of Lettermullen. [...] Why did not your correspondent make honest inquiry before so writing?

In (2), the first two stages of the text structure in Figure 1 are clearly visible. Mrs Healy begins the letter by referring to the relevant article and summarising the key point raised by the journalist. Her disagreement then goes on the record, as it were, by means of a language marker that takes the form of a loaded question, namely "one that has a presupposition that the respondent is not committed to" (Walton, 1999, p. 381). There is no doubt that Mrs Healy's 'why?' plainly took for granted that the writer failed to 'make honest inquiry'. In the following excerpt (3), moreover, the Principal Teacher assumes the two identities numbered as 3a. and 3b. in Figure 1:

- (3) Now, I am the "schoolmistress" referred to [...] and I must stigmatise that statement as a lie – an unjust aspersion on my love of our grand old Celtic tongue – and a more than unjust, a hurtful and libellous animadversion on my

² In all numbered examples, emphasis is mine.

professional character, and one which I shall take the proper steps to meet. [...] I do speak Irish. I am competent to instruct and explain through the medium of Irish, and, so far from my pupils “learning nothing”, the contrary is the fact, as many a poor boy and girl, taught by myself and my late husband, occupy fairly respectable positions in Ireland and America through the education received here. [...] My Inspectors, too, can prove the efficiency of my work, and my respected manager can not only authenticate my statement, but bear testimony to the weary weary drudgery in such a place as this [...]

First, she takes on the role of polemicist: her counter-discourse therefore questions the credibility of the *Freeman's* journalist, portrayed as a liar whose allegations are both unfair and libellous. Secondly, she takes on the role of arguer by advancing two standpoints. The first is that she speaks Irish and can use it as a medium of instruction. The second is that her pupils actually learn a lot. Each standpoint is supported by a specific argument. The first is that Mrs Healy's ‘Inspectors’ and ‘Manager’ can prove the veracity of her statements. The second lies in the fact that several of her former pupils now occupy respectable and well-paid positions both in Ireland and overseas. Interestingly, the publication of the letter was accompanied by the *Freeman's* editor's apology for any ‘inaccuracy’ in the previous article, as reported in (4) below:

- (4) We are sure that our correspondent, in attacking the system of education in Irish-speaking districts, did not intend to do any injustice to Mrs Healy, and if his remarks were read as referring to her, we regret his inaccuracy in describing her as not knowing Irish.

3.2 *The discourse of disagreement: a broader corpus perspective*

The scope of this section will extend our appreciation of the dataset, by shifting the attention from the one example taken as a case in point earlier on to the corpus at large. Focusing on the stages of text structure where disagreement is expressed, i.e. stages 2, 3a. and 3b. in Figure 1, discernible patterns tend to emerge from the data. As we noted in the previous section, to begin with, language markers of opposition are very often phrased as questions. Some of these appear more tentative, as with the first two listed under (5) below; some others, such as the following one, are slightly more personal; others still are definitely more direct and potentially harder to dodge, as is the case for the last two:

(5) Why, I wonder, ...?

These being the facts, what becomes of...?

Now I am entitled to ask, and the Chief Secretary will be obliged to answer, by what right...?

...and what I want to know is why in goodness they didn't try to...?

In the name of common sense, and for the credit of Irish journalism, may I ask for what reason...?

What is rhetorically interesting about some of the questions in (5) is that they bolster the writer's case in that they retain a semblance of factuality – cf. 'these being the facts, what becomes of...?' – or they are endoxically rooted in public opinion – e.g., 'In the name of common sense...'. As Amossy (2009, p. 7) rightly points out, this may serve the purpose of "objectively" evoking feelings of indignation from the readership. Only at a later stage are targets presented as worth wholesale moral condemnation.

Such use of questions therefore creates the ideal ground for the voicing of disagreement unfolding as the writer acts as a polemicist and at once an arguer. As far as the discourse of writers as polemicists is concerned, the findings show that targets, particularly political ones, were discredited in three main ways. First of all, they were attacked for taking utterly unreasonable steps to reverse an undesirable state of affairs, as in (6) and (7) below. In (6), the writer found it inexplicable that so many families had been 'struck off' programmes of public works while living in grinding poverty. In (7), likewise, the Congested Districts Board are blamed for their 'attitude' towards the fishing industry, which they were allegedly leaving in an awful state of neglect:

(6) The Guardians and Local Government Board have struck off thirty families who were on the public works [...]. It is a most extraordinary and unaccountable piece of work to do so now, when the poor people stand most in need of assistance...

(7) ...and the Congested Districts Board would seem to be afraid to give any assistance to the starving industry [...] This, I think, is an attitude entirely unworthy of the Congested Districts Board...

Secondly, the authorities were targeted for doing too little, too late to relieve the west of Ireland's economic distress. Hence in (8), the belated recognition of the people's need for employment after the failure of the potato crop only led the competent boards to give 'poor Paddy' – itself a

sarcastic, derogatory way to refer to the Irish people – ‘a few potatoes in his hat’ instead of actual jobs. Thirdly, targets such the British Government were picked and identified as the cause of the evils they were themselves supposed to be fighting, as in (9). In a passage that leaves little to the imagination, the writer asserted that the Government were responsible for the pitiful state of affairs Connemara had got into. In other words, it was the State-sponsored ‘machinery of law and armed forces’ that allowed the ‘exterminators’ to prosper that brought about and perpetuated ‘the rotten economic system’ of absentee landlordism.

(8) Oh the horrid idea of giving poor Paddy a few potatoes in his hat at starvation periods. It is degrading and not reproductive of any good, and never will be. No English labourer would be satisfied so easily...

(9) The Government is responsible for this condition of affairs. It encouraged with its patronage, and assisted with its whole machinery of law and armed forces, the exterminators who devoted all their energies to bringing about the rotten economic system which exists in Connemara.

As regards writers as arguers, finally, there is much evidence that they adopted two general strategies. In first place, they frequently embarked in what Van Eemeren et al. (2007, p. 177) refer to as “criticism of causal argumentation”. In many a passage characterized by a high degree of dialogism, writers ideally respond to the target(s)’ causal argumentation with reference to the effects the proposed measures may or may not be able to produce. As is often the case, writers can be accounted for as answering critical questions such as the following:

Do the proposed measures indeed lead to the intended result?
Could the intended result be achieved by any other means?

This is apparent from passages such as (10). Here, the writer again referred to the complete failure of the potato crop in Connemara, an all-but-isolated incident in the west of Ireland:

(10) ...the failure of the potato crop has been not only partial, but complete [...] and now for the first time in Connemara history are very many of our people, shopkeepers, public institutions, etc. getting their potatoes in Dublin. [...] The outlook is gloomy in the extreme; and “the authorities”, instead of sending round their inspectors and acting on “the pinch-of-hunger policy”, would be doing more wisely if at once they gave

remunerative employment to the people and made preparations for supplying them with good seed next spring. And the opportunities for employment are very many indeed.

In (10), two strands can be identified in the writer's argumentative discourse. On the one hand, he rejected the authorities' policy in that these would fail to accomplish the desired objective. On the other hand, he suggested that they should adopt different policies on the grounds that these would in fact produce the desirable outcome of relieving poverty. The two strands reproduced below intersect each other in a passionate plea for help for the poor people of Connemara, who needed remunerative employment far more than 'inspectors' and eleemosynary relief:

1. The authorities' policy (sending round inspectors etc.) should be rejected.

1.1 The authorities' policy fails to lead to the intended result (relieving poverty).

2. Different policies (giving people employment etc.) should be adopted.

2.1. Different policies would lead to the intended result.

Besides undermining the target(s)' argumentation on the effectiveness of the proposed measures, secondly, letter writers were observed to resort to patterns of pragmatic argumentation. In particular, writers' discourse was identified as putting forward Variant I of pragmatic argumentation, through which an action is called for by virtue of the worthwhile end it is assumed to attain (Van Poppel, 2012, p. 99). This can be seen in passages such as that reported in (11). Here, the writer talks about a well-known economic activity of Ireland's coastal regions at least until the end of the nineteenth century. This was kelp making, namely the practice of collecting and air-drying seaweed in order to extract iodine, which was then sold to soap makers as well as the pharmaceutical industry (Mac an Iomaire, 2000 [1938], pp. 139-141):

(11) I would suggest [...] the substitution of some other process of preparing the air-dried seaweed for the market, for the present crude and ruinous system of burning it. It is a well recognised fact that fifty per cent of the iodine [...] becomes volatilized in the process of burning, and is entirely lost. If this 50 per cent of waste could be saved the value of the industry would thereby be considerably enhanced. I understand that a process such as I refer to

has been successfully carried out for years on some islands in the North of Scotland.

In an attempt to address the problem of iodine becoming ‘volatilized’ and going to waste in the burning process, the writer provides his own constructive feedback. As indicated in the schematisation below, he proposed that the method employed in Scotland should be implemented in Ireland, too. This is justified by the beneficial result that would deliver, that is the saving of iodine along with an increased profitability of the industry:

1. The method employed in Scotland should be adopted here, too.

1.1 The method carried out in Scotland saves iodine and makes the industry more profitable.

4. DISCUSSION AND CONCLUSIONS

At a more general level, the findings presented earlier on show that letters may have played a central role in establishing an embryonic public sphere in Ireland. A few decades after the introduction of a national system of education and facilities such as school libraries as well as reading rooms (McDowell, 1952), the figures from the 1901 census indicated that the percentage of illiterates had dropped from 47% of fifty years before to 14% (Tobin, 2018). Along with that, the abolition of stamp duty, decreasing costs of resources, tools and processes, improved transport for distribution and the intensification of population concentration through urbanisation all ensured that Irish journalism came of age at a time of momentous change in the Irish public landscape.

At the turn of the century, therefore, Irish journalists possessed traits that would set rising standards of professionalism. They “had a skill (shorthand), a professional adherence to objectivity and were aware of new developments within journalism, such as the new form of writing, the interview” (Foley, 2004, p. 381). Yet through the medium of newspapers and other periodicals, opinionated readers also had an opportunity to have their say on the issues that mattered most to them, from national politics to local government, from education to the measures needed to stimulate the economy.

From this perspective, letters to the editor are likely to have been considered by more and more people as evidence of increased civic involvement and a civil society “in which democratic practices can develop, in which an autonomous public opinion can be formed, and in which people acquire the skills that make them effective citizens in a modern polity” (Kissane, 2002, p. 113). In a period which would later be

regarded as setting the preconditions for the emergence and consolidation of Irish democracy over the first three decades of the twentieth century, letters contributed to bringing informed readers' opinions to bear to a new informational order, where they "could be seen, assessed, and, if possible, become part of a public discourse" (Morash, 2014, p. 31).

More specifically, the patterns of text and discourse structure reviewed in Section 3 appear instrumental in shaping up writers' argumentation in three main respects: first of all, in way that is inscribed in the fully legitimised space provided by the print media (Charaudeau, 2005); secondly, in a way that gives rise to antagonistic counter-discourses mobilising public dialogue and debate (Amossy & Burger, 2011) in a country where the latter would otherwise have risked being stifled by a conservative, peasant and rural society dominated by the cultural deference promoted by the Catholic Church and a dominant public discourse privileging solidarity and cohesion over public engagement (Murphy, 2011); thirdly, by integrating distinctive writer profiles such as that of polemicist and arguer through language tools and argument schemes that lead the discourse of disagreement to fulfil the social function of 'coexistence in dissent' Amossy (2011) sees as a desirable way to handle conflict verbally.

While extant research on Irish journalism mainly focuses on the historical development of both the national and the provincial press, this study has delved into aspects of text and discourse structure. By means of a qualitative analysis that uncovered evidence of the strategies for the expression of disagreement, this small-scale research has implemented a method that could be applied to comparative studies of correspondence to Irish editors from a few decades later. This would shed light on any pattern of (dis)continuity in the text structure or the argumentative discourse of the average Irish letter-writer, as it were.

REFERENCES

- Amossy, R. (2009). Argumentation in discourse: A socio-discursive approach to arguments. In J. Ritola (Ed.), *Argument cultures: Proceedings of OSSA '09* (pp. 1-12). Windsor: OSSA.
- Amossy, R. (2011). La coexistence dans le discours. *Semen*, 31, 25-42.
- Amossy, R. & Burger, M. (2011). Introduction: La polémique médiatisée. *Semen*, 31, 7-24.
- Angenot, M. (1982). *La parole pamphlétaire. Typologie des discours modernes*. Paris: Payot.
- Breathnach, C. (2005). *The Congested Districts Board, 1891-1923. Poverty and development in the west of Ireland*. Dublin: Four Courts Press.

- Brinton, L. J. (2001). Historical discourse analysis. In D. Schiffrin, D. Tannen & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 138-160). Oxford: Blackwell.
- Brownlees, N. (2009). "Alwayes in te Orbe of honest Mirth, and next to Truth": Proto-infotainment in *The Welch Mercury*. In A. J. Jucker (Ed.), *Early modern English news discourse* (pp. 57-72). Amsterdam: Benjamins.
- Brownlees, N. (2016a). A Medici agent's newsletter to Florence during the Leghorn crisis of 1653. In A. Assonitis & B. Sandberg (Eds.), *The Grand Ducal Medici and their archive (1537-1743)* (pp. 203-211). Turnhout: Brepols.
- Brownlees, N. (2016b). Reporting the United States of America in the Liverpool Mercury (1812). In M. Dossena & S. Rosso (Eds.), *Knowledge dissemination in the long nineteenth century* (pp. 137-156). Newcastle-upon-Tyne: Cambridge Scholars.
- Burger, M. (2011). Une caractérisation praxéologique du désaccord polémique: Ce qu'informer dans les médias veut dire. *Semen*, 31, 61-80.
- Charaudeau, P. (2005). *Le discours politique. Les masques du pouvoir*. Paris: Vuibert.
- Dunlop, A. (1911). *Fifty years of Irish journalism*. Dublin: Hanna and Neale.
- Foley, M. (2004). Colonialism and journalism in Ireland. *Journalism Studies*, 5(3), 373-385.
- Fürsich, E. (2009). In defense of textual analysis. Restoring a challenged method for journalism and media studies. *Journalism Studies*, 10(2), 238-252.
- Garand, D. (1998). Propositions méthodologiques pour l'étude du polémique. In A. Haitward & D. Garand (Eds.), *États du polémique* (pp. 211-268). Québec: Nota Bene.
- Glandon, V. (1985). *Arthur Griffith and the advanced-nationalist press, 1900-1922*. New York: Peter Lang.
- Ilie, C. (2010). Managing dissent and interpersonal relations in the Romanian parliamentary discourse. In C. Ilie (Ed.), *European parliaments under scrutiny. Discourse strategies and interaction practices* (pp. 193-222). Amsterdam: Benjamins.
- Jacquin, J. (2011). Le/La polémique: Une catégorie opératoire pour une analyse discursive et interactionnelle des débats publics? *Semen*, 31, 43-60.
- Kee, R. (2000) [1972]. *The green flag. A history of Irish nationalism*. London: Penguin.
- Kerbrat-Orecchioni, C. (1980). La polémique et ses définitions. In N. Gelas & C. Kerbrat-Orecchioni (Eds.), *Le discours polémique* (pp. 3-40). Lyon: Presses Universitaires de Lyon.
- Kissane, B. (2002). *Explaining Irish democracy*. Dublin: University College Dublin Press.
- Kurzon, D. (2001). The politeness of judges: American and English judicial behaviour. *Journal of Pragmatics*, 33(1), 61-85.
- Legg, M. L. (1999). *Newspapers and nationalism: The Irish provincial press, 1850-1892*. Dublin: Four Courts Press.
- Mac an Iomaire, S. (2000) [1938]. *The shores of Connemara*. Newtownlynch: Tír Eolas.

- Mazzi, D. (2019). *Views of place, views of Irishness. Representing the Gaeltacht in the Irish press, 1895-1905*. Oxford: Peter Lang.
- McDowell, R. B. (1952). *Public opinion and government policy in Ireland, 1801-1846*. London: Faber and Faber.
- Morash, C. (2010). *A history of the media in Ireland*. Cambridge: Cambridge University Press.
- Morash, C. (2014). Ghosts and wires: The telegraph and Irish space. In K. Steele & M. de Nie (Eds.), *Ireland and the New Journalism* (pp. 21-33). New York: Palgrave Macmillan.
- Murphy, M. P. (2011). Civil society in the shadow of the Irish state. *Irish Journal of Sociology*, 19, 170-187.
- Nir, L. (2011). Disagreement and opposition in social networks: Does disagreement discourage turnout? *Political Studies*, 59, 674-692.
- O'Brien, M. & Rafter, K. (Eds.) (2012). *Independent newspapers. A history*. Dublin: Four Courts Press.
- Paltridge, B. (1996). Genre, text type, and the language learning classroom. *ELT Journal*, 50(3), 237-243.
- Rafter, K. (Ed.) (2011). *Irish journalism before independence. More a disease than a profession*. Manchester: Manchester University Press.
- Robles, J. S. (2011). Doing disagreement in the House of Lords: 'Talking around the issue' as a context-appropriate argumentative strategy. *Discourse & Communication*, 5(2), 147-168.
- Shiffman, G. (2002). Construing disagreement. Consensus and invective in "Constitutional" debate. *Political Theory*, 30(2), 175-203.
- Steele, K. & de Nie, M. (Eds.) (2014). *Ireland and the New Journalism*. New York: Palgrave Macmillan.
- Tobin, M. (2018). *Literacy and society in Ireland 1900-1980*. Unpublished Ph.D. dissertation, Department of Sociology, National University of Ireland (Maynooth).
- Tognini Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: Benjamins.
- Uzelgun, M. A., Mohammed, D., Lewiński, M. & Castro, P. (2015). Managing disagreement through *yes, but...* constructions: An argumentative analysis. *Discourse Studies*, 17(4), 467-484.
- Van Eemeren, F. H., Houtlosser, P. & Snoeck Henkemans, A. F. (2007). *Argumentative indicators in discourse. A pragma-dialectical approach*. Dordrecht: Springer.
- Van Poppel, L. (2012). Pragmatic argumentation in health brochures. *Journal of Argumentation in Context*, 1, 97-112.
- Walsh, O. (2008). Landscape and the Irish asylum. In Ú. Ní Bhroiméil & G. Hooper (Eds.), *Land and landscape in nineteenth-century Ireland* (pp. 156-170). Dublin: Four Courts Press.
- Walton, D. (1999). The fallacy of many questions: On the notions of complexity, loadedness and unfair entrapment in interrogative theory. *Argumentation*, 13, 379-383.

An Argumentative Approach to the Justification of Abduction

PAULA OLMOS

Universidad Autónoma de Madrid

paula.olmos@uam.es

The philosophical debate over the justification of abduction can be modelled as the critical assessment a *warrant-establishing argument* allowing “H explains D” to be used as a reason for “H can be inferred from D”. Philosophers discuss conditions under which such kind of generic argument could be accepted. Five kinds of conditions are identified and commented on: a) dialectical/procedural restriction; b) claim restriction; c) restriction over acceptable explanatory principles; e) balancing restriction and d) epistemic restriction.

KEYWORDS: abduction, backing, explanation, inference to the best explanation, justification, metaphilosophy, Toulmin model, warrant.

1. INTRODUCTION. PHILOSOPHY AS A CASE-MAKING ACTIVITY

This paper is an essay on *metaphilosophy* inasmuch as it tries to review, describe and categorise different ways philosophers have approached a certain self-assumed philosophical task. In our case, this task is the justification of a certain mode of reasoning and arguing, namely abduction. Such a kind of endeavour responds to a tradition which has been mainly represented by the well-known and long-standing philosophical discussion on the “justification of induction”.

Our inquiry is, in any case, presided by the assumption that what philosophers mainly do is arguing. So the idea is to approach their pursuit as an *argumentative activity* and, more specifically, under a conception of arguing and argument that does not aim at capturing what *follows from* what, of what *is implied* by what, but tries instead to understand how something (some content) is presented/proposed by someone to others as a reason for something else in a communicative setting (Marraud 2013).

This idea of philosophy, not only as an argumentative activity but, more specifically, a kind of *case-making* activity under the model of legal discussion was already proposed by Friedrich Waismann in his definitely metaphilosophical and influential paper “How I See Philosophy” (1968 [1956]).

The essential difference between philosophy and logic is that logic constrains us while philosophy leaves us free: in a philosophic discussion we are led, step by step, to change our angle of vision [...] a thing profoundly different from deducing theorems from a given set of premisses (Waismann, 1968 [1956], p. 21)

[philosophical arguments] were, quite mistakenly as I hope to have shown, supposed to be proofs and refutations in a strict sense. But what the philosopher does is something else. He builds up a case. First, he makes you see all the weaknesses, disadvantages, shortcomings of a position; he brings to light inconsistencies in it or points out how unnatural some of the ideas underlying the whole theory are by pushing them to their farthest consequences [...] On the other hand, he offers you a new way of looking at things not exposed to those objections. In other words, he submits to you, like a barrister, all the facts of his case, and you are in the position of a judge (Waismann, 1968 [1956], p. 30).

But one interesting twist is that philosophy is a case-making or reason-giving activity that is particularly interested on *other* case-making (reason-giving) activities: from the most ordinary and universal to the most sophisticated and heavily institutionalized ones. Thus, Jonathan L. Cohen goes as far as characterizing philosophy as “the reasoned investigation of reasons” or “the reasoned discussion of what can be a reason for what” (Cohen, 1986, pp. 49–50, 57). This last remark fits exactly a discussion as that of the “justification of abduction” in just the way we are going to reconstruct it.

On the other hand, it should not be disregarded that, precisely in our case, the philosophical interest in the justification of abduction has been particularly encouraged by its centrality for discussions regarding science and its own methods *within* philosophy of science (Olmos, 2018a). Although John Woods’s paper on the *logic of abduction* (2016) aims at a more general, epistemological, picture of what is for him basically a naturalistically (evolutionarily) developed mode of reasoning, he must nevertheless deal with the nature of scientific enquiry and its own justificatory standards. It is in this regard that he offers us a usefully *argumentative* characterization of science itself, another *forensic*, case-making (communicatively reason-giving) activity:

Not unlike the law, science is in significant measure a case-making profession –a *forensic* profession– made so by the premium it places on demonstrating that knowledge has been achieved, rather than just achieving it. This has something to do with its status as a profession, subject to its own exacting requirements for apprenticeship, standard practice, and advancement. These are factors that impose on people in the showing professions expectations that regulate public announcement. [...] Publication is a vehicle for case-making, and case-making is harder than knowing. Journal editors don't give a toss for what you know. But they might sit up and notice if you can show what you know (Woods, 2016, 143-144).

So the philosophical justification of abduction deserves a *doubly* argumentative approach: as an argumentative activity dealing with another argumentative activity. If I now reveal that I favour what's usually dubbed as a *first-order construal* of metaphilosophy for which “the application of philosophy to philosophy itself, is simply one more instance of philosophy” (Joll, 2017; Cf. Wittgenstein, *PI*-§121; Cf. Williamson, 2007, p. ix), it is clear that the reader of this paper is dealing with no less than *three* argumentative layers which should, on the one hand, be clearly distinguishable and, on the other, be understood and dealt with the *same* practical and conceptual tools.¹

Keeping this in mind, section 2 will just try to clarify, for the purposes of the present discussion, my own argumentative account of abductive argument, as developed in previous papers (Olmos 2019a; 2019b; forthcoming). Section 3 will show different argumentative ways philosophers have essayed for the global justification of abduction, some of which (3.1, 3.2, 3.3, 3.4) simply demand restrictions compatible with the current critical assessment of particular abductive arguments, while others (3.5) try to reach a deeper level in which the very grounds of “what constitutes a reason for what” are directly confronted.

Finally, some conclusions about the nature of these discussions are offered in section 4.

¹ This is not necessarily so, as philosophy could claim to use different methods or argumentatively distinct strategies and grounds. However, although this cannot be dealt with here, my contention is that there is nothing fundamentally diverse in philosophical argument.

2. ARGUMENTATIVE APPROACHES TO ABDUCTION²

Since D. Walton's extensive monographic work (Walton, 2004) there have been some distinct argumentative approaches to abduction. Probably not as many as could be expected but certainly covering different trends within argumentation theory. Walton's work is based on the methodology of "dialogue types", offering both a "dialogical model of explanation" plus a "dialogical model of justification of best explanation" (which is, for him, equivalent to abduction). Wagemans's approach (2016) is pragma-dialectical and Yu & Zenker (2018) employ a methodology of argumentation schemes plus critical questions.

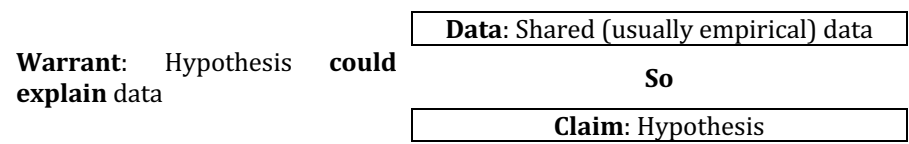
My own proposal offers a model of abduction based on a distinctly Toulminian argumentation scheme (in which the role of the warrant is duly emphasized), supplemented with some analytic tools accounting for *inter*, *counter*, and *meta-argumentative* structures.³ According to this framework, presenting in a communicative interchange an abductive argument is to support an in principle theoretical or factual claim (typically mentioning either unobservable or merely unobserved entities, properties and processes) on the basis of shared data (typically observable, well-known, taken for granted or assumed) *precisely because* it provides some conceivable explanation to them. The basic elements of an abductive argument, thus understood are:

1. **Conclusion/Claim:** an "explanatory hypothesis" H, usually presented as a factual statement, although, depending on the requirements of the context it may be easily reinterpreted as a practical or even evaluative conclusion of the kind: "we should explore hypothesis H", "Hypothesis H is worth exploring".
2. **Reasons/Data:** usually empirical, observable but in any case presented as *shared* or *agreed upon* data, and nevertheless *surprising data*, i.e. data *requiring explanation* (what makes of them a potential *explanandum*). However this preparatory condition might be contested in an ensuing discussion.
3. **Warrant:** what makes of the data a justificatory reason for the conclusion (the hypothesis) is that such hypothesis *could explain* them.

² This section is based on previous work (Olmos, 2019a, 2019b, forthcoming). Concrete quotations won't be indicated in the body of text.

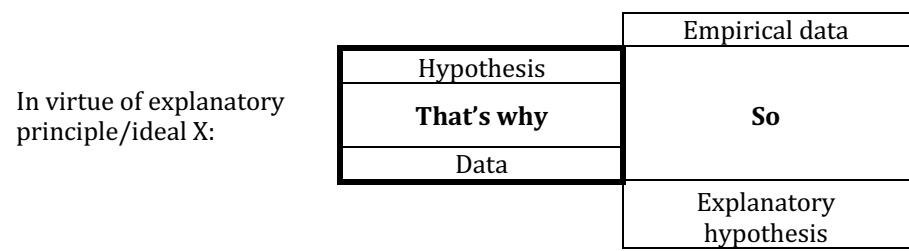
³ These tools are best described in a series of papers by H. Marraud's: Cf. <https://uam.academia.edu/hubertmarraud>.

These elements may be represented thus:



In the case of abductive arguments, it could seem as though the kind of warrant I am proposing is not exactly “of a more general scope” than the data and claim of the argument, referring exactly to the same items as its basic constituents. Nevertheless, this is just apparently so as such a warrant introduces between these items a *substantive kind of relation*, namely “explanation”, which responds as well to a variety of *principles* or *explanatory warrants* on which it may be based. The concretion of the type of *explanans* that the Hypothesis is vis-à-vis the Data (taken as *explanandum*) provides the degree of *principled generality* that an argument requires to be so recognized and understood (and eventually assessed) by an interlocutor.

According to this idea and taking in account the structural homogeneity of argument and explanation as the products of pragmatically different *acts of giving reasons* (Álvarez, 2016) it is possible (and useful) to construe an “expanded diagram” of abduction, including the details of the *related* explanation –an explanation that is not exactly *given* in the abductive argumentative act but just, *mentioned* or *alluded to* as what makes of the empirical data a justificatory reason for the (theoretical) hypothesis– thus:



However, argumentative models of abduction are not really my topic in this paper. All these models, in fact, assume both the widespread presence of abductive arguments in a variety of argumentative practices and their *evaluable character*, proceeding then to determine ways in which abductive arguments are in fact or should be assessed.

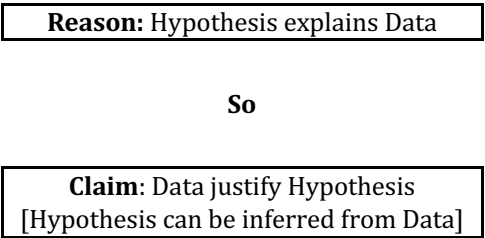
But these conditions are precisely the nub of what is *at stake* in philosophical discussions about the justification of abduction. At least in principle, although, depending on the particular solution provided they are questioned to a greater or lesser degree.

3. PHILOSOPHICAL DISCUSSIONS ON THE JUSTIFICATION OF ABDUCTION

So, the idea of philosophical discussion on the justification of abduction is to discuss *the grounds themselves* that allows for the understanding, conception and presentation of abductive arguments. Using Toulminian terminology, they would discuss *up to what point* the kind of (substantive) link between data and claim invoked in an abductive argument (i.e. expressed in its warrant) makes of the data a justifying reason for that claim. Thus, such philosophical debates might be considered as *warrant-discussing* argumentative activities aiming at *establishing warrants* (or warrant-types, Cf. Toulmin, 2003 [1958], 111-113, 125-126).

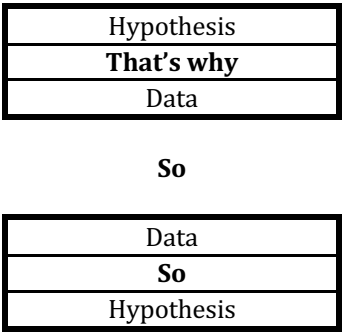
According to these ideas, philosophical discussions on the justification of abduction might be modelled as constituting an, either restricted or more radical but, in any case, reasoned (and argumentative) *critical assessment* of the following scheme in which “H explains D” (or “H, that’s why D”) is taken as a reason to support that “D justify H” (or “D, so H”):

Warrant: That a hypothesis explains some data is a reason to consider those data a justifying reason for that hypothesis:



For the purposes of my discussion, I will, in fact, be using *more expanded* versions of this same scheme as the following:

That a hypothesis explains some data is a reason to consider those data a justifying reason for that hypothesis:



In virtue of “explanatory principle X”:	Hypothesis
	That’s why
	Data

That a hypothesis explains some data is a reason to consider those data a justifying reason for that hypothesis:

So

Hypothesis can be inferred from Data

Now, if you ask, *what kind of reason* could that one be for such a claim? the answer should be expressed by the warrant I have provided which is, I must agree, *frustratingly redundant*, and so kind of useless. It might be considered (taking in account how it is articulated) a “more general” statement than the argument it covers, but it surely does not add any new substance to it, beyond formal or informal *subsumption* (depending on quantification).

This is a problem according to my own reading of Toulmin’s warrants.⁴ However, I will advance two excuses for it. The first is that we might be reaching a really deep (cognitive, logical) level of what counts as a reason, not really based on more reasons-for-*reason-being* than sheer “intuition”. But, this is really what is *at stake* and what some philosophers (as those mentioned in 3.5) will really try to handle by suggesting “backings” for such a rule.

The other excuse is simply operative. Such a scheme (as will be shown in what follows) makes room for both the philosophically deeper and the more *accommodating* (and qualified) restrictions on the use and acceptability of abductive arguments that different authors have supplied and is, therefore, a good instrument to compare them.

Now, philosophers such as J. Woods, I. Hacking, P. Lipton, B. van Fraassen or I. Douven (many of them interested in attacking or defending “scientific realism”) have discussed abduction starting with the idea that it is not a mode of reasoning and arguing that could be unqualifiedly or unrestrictedly admissible. So their strategy is demanding additional restrictions or well-defined conditions under which such kind of generic scheme (and therefore, the particular abductive arguments generated by it) could be accepted.

Five such modes or levels of restriction might be identified in philosophical literature and will be described in the following subsections.

⁴ See, my Commentary on J.A. Blair’s paper in these same *Proceedings*.

3.1 Global dialectical/procedural restriction (or field limitation)

The generic scheme and so the use of abductive arguments might be admissible in some argumentative fields while not in others: e.g. it is all right to use abductive arguments in everyday life but not in scientific inquiry. Such global restrictions are usually made dependent on considerations regarding:

- i) The higher or lower degree of certainty demanded from proofs in that field.
- ii) The greater or lesser need (or institutional obligation) to reach a conclusion.

Both kinds of considerations obviously act in opposite directions as to the admission of abductive arguments. B. van Fraassen's contention about the non-scientific character (because they go beyond empiricism) of conclusions based on abduction is a good example of this:

A person may believe that a certain theory is true and explain that he does so, for instance, because it is the best explanation he has of the facts or because it gives him the most satisfying world picture. This does not make him irrational, but I take it to be part of empiricism to disdain such reasons (van Fraassen, 1985, 252).

3.2 Claim restriction (related to the argument's sufficiency)

The generic scheme and so the use of abductive arguments might be admissible adjusting the *mode of validity* of its claim (or the kind of claim presented by its conclusion). In his paper on the logic of abduction, J. Woods emphasizes what he sees as one of the Peirce's insights: "Rather than believing them, the proper thing to do with abduced hypotheses is to send them off to experimental trial. (CP, 5. 599, 6. 469-6. 473, 7. 202-219)" (Woods, 2016, 138). According to this idea, abduction may not really provide reasons (enough? any?) to assert the hypothesis but may function as a directive "practical argument" in the following way:

In virtue of “explanatory principle X”:	Hypothesis
	That’s why
	Data

That a hypothesis explains some data is a reason to consider those data a justifying reason to **experimentally test** such hypothesis:

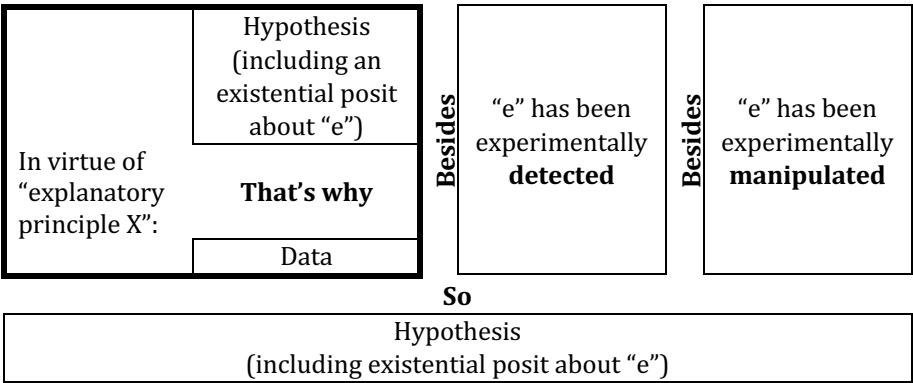
So

Data
So
Hypothesis should be empirically tested

Now, using similar intuitions, I. Hacking, in *Representing and Intervening* (1983: 271-272) mentions the *insufficiency* of abduction, i.e. of the mere explanatory power of a hypothesis, to support a “realist claim” regarding a theoretical entity (e.g. the electron), included among the posits of that hypothesis. His point is that we are finally justified in supporting such a “realist claim” as a theoretical statement about the entity’s “existence” when our experimental practices have allowed us:

- a) to (directly/indirectly) *detect* the entity, and
- b) to effectively *manipulate* it in further experiments *unrelated* to its establishment (Cf. Douven 2002, 360-362).

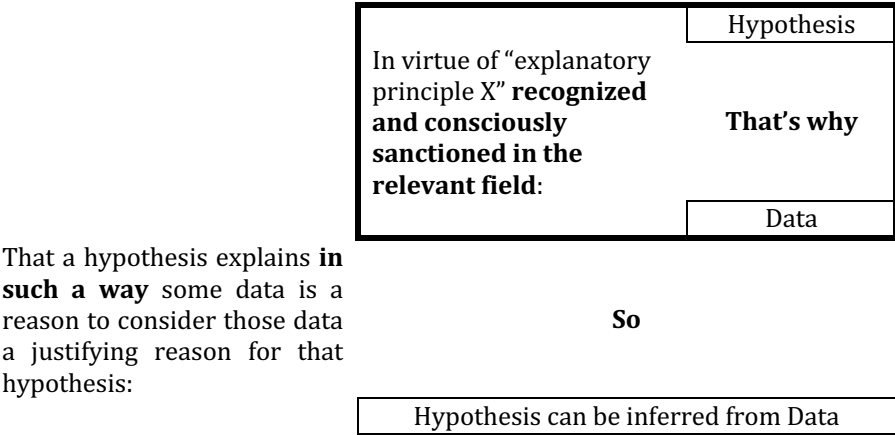
Hacking’s suggestion might be argumentatively modelled as requiring for the sought for conclusion (the assertion of the hypothesis) a *conjunction of arguments* (Marraud, 2013, p. 59-62; Cf. Olmos, 2018b, p. 23) which functions as a *coordinative argumentation* (Snoeck-Henkemans, 2003).



3.3 Restriction over acceptable explanatory principles

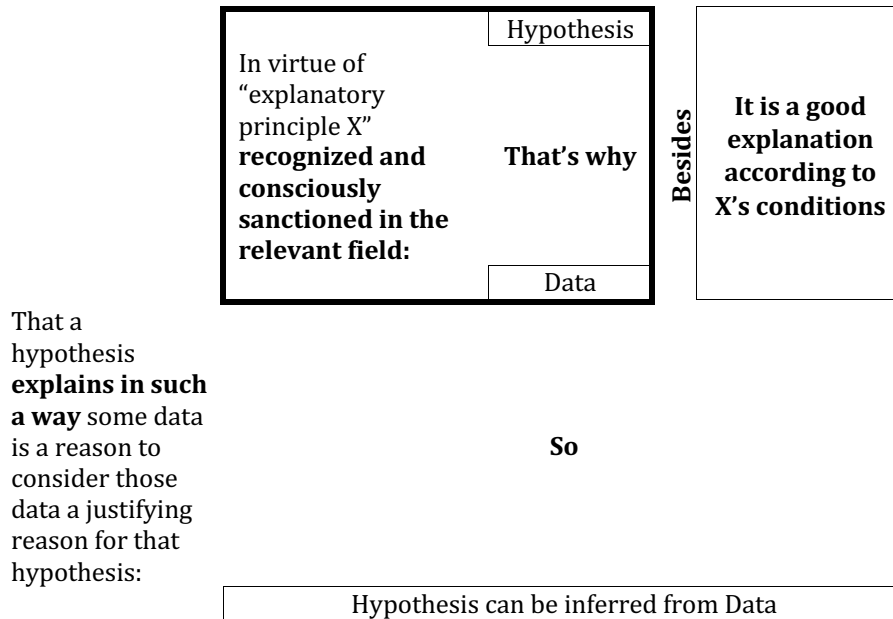
A different kind of discussion arises when one concentrates on the idea that explanations, as arguments, may also be classified according to the diverse kinds of “explanatory warrants” they are based on. A possible answer to the admissibility of abduction makes it dependent on the kind of associated explanation that serves as its warrant. In this sense, the generic abductive scheme and so the use of abductive arguments might be admissible just in case some normative requirements are placed on it regarding three different aspects.

- a) A restriction over the *kinds* of explanation (or explanatory principles) acceptable in a given field.⁵ This restriction might be added to the diagram of an acceptable abduction, thus:

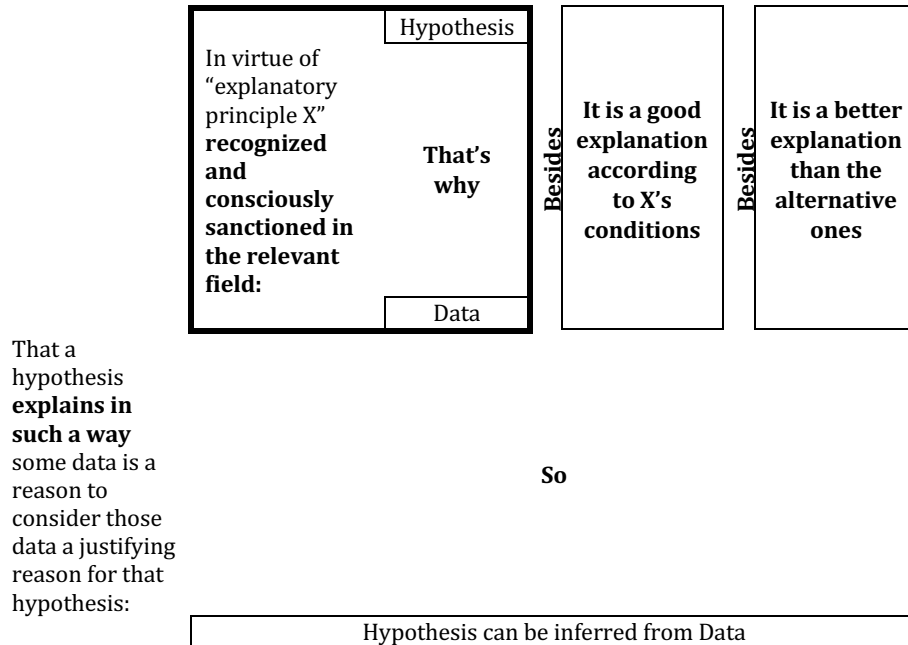


⁵ This is usually what’s at stake in contemporary philosophical discussions on the notion of “mechanism” as what may be the basis of a scientific explanation (Glennan & Illari, 2016).

- b) A restriction over the *quality* of such an explanation, based on standardized criteria associated to the “explanatory principle” in case. Here what would really be the conclusion of a process of critical assessment of the explanation is represented as an additional reinforcing reason.



- c) Finally, a further restriction could be called for regarding the *comparative quality* of the explanation *vis à vis* other alternative explanations. Here, again, what would really be the conclusion of a more complex argumentative process is represented as just an additional reason for the conclusion.



But the interest of such a justification process in itself has led some philosophers to concentrate on it as what's really the true key to the assessment of abduction which could only be acceptable and even assessable under the model of an "inference to the best explanation". The next kind of restriction in our classification accounts for this idea.

3.4 Balancing restriction (stipulation of comparative weighing)

Philosophers such as I. Douven (2017) consider abduction is only justifiable under an IBE model. That means that emphasis is placed over the justification of the comparative premise: "It is a better explanation than the alternative ones".

Now, according to my own analysis (Olmos 2019b) this implies the systematic weighing of the arguments supporting each alternative explanatory hypothesis which of course can be very different and based on different warrants. Thus, my claim is that there cannot be a rigid universal model of IBE-type argumentations. In any case, among the simplest (and probably most interesting) possibilities is the comparison between two abductive arguments (supporting each an alternative

explanatory hypothesis) based on the superiority of one of the explanatory principles invoked over the other. This could be thus represented:

Explanatory principle X is superior to explanatory principle Y (in the relevant field)				
So				
In virtue of explanatory principle X:	Explanatory Hypothesis A		Explanatory Hypothesis B	In virtue of explanatory principle Y:
	That's why	>	That's why	
	Empirical data		Empirical data	
So				
Hypothesis A offers a better explanation than Hypothesis B				

3.5 Epistemic restriction (or backing-request)

So far, the ideas advanced by different authors for a qualified acceptance of abduction as a *prima facie* plausible and assessable mode of arguing, even if conducted in a universal and allegedly conceptual way, do not really go much beyond the *usual* criteria we are accustomed to in assessing *particular* abductions as establishing stronger or weaker arguments. Thus, the critical questions usually associated with abduction as an argumentative scheme (Cf. Marraud, 2017, p. 5) may easily take care of requirements such as those expressed in 3.1 (usability of the scheme in a certain context), 3.3 (grounds and quality of the associated explanation) and 3.4 (comparison of claimed hypothesis with alternative hypotheses). So, finally, what these kinds of (philosophical) restrictions really amount to is the acceptability of *good-enough* abductions.

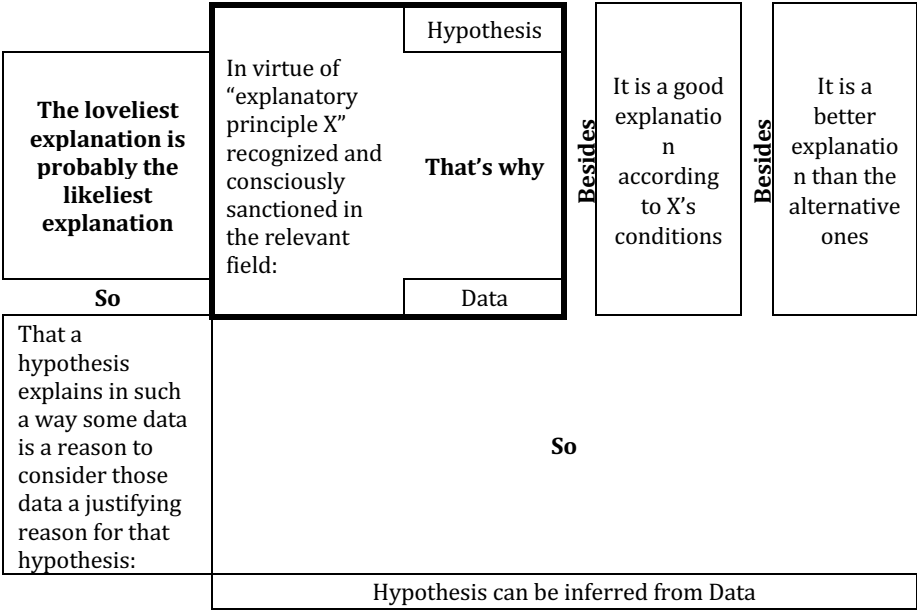
Concerns expressed in 3.2 are probably harder to accommodate in such a methodology, as they address the very definition and comprehension of abduction itself as a theoretically or pragmatically bounded (and accordingly evaluated) way reasoning and arguing. But still, the abductive *principle* is taken for granted.

This is not the case with discussions attaining precisely the *tenability* of the principle itself as an epistemic⁶ rule (i.e.: that a hypothesis explains some data is a reason to consider those data a

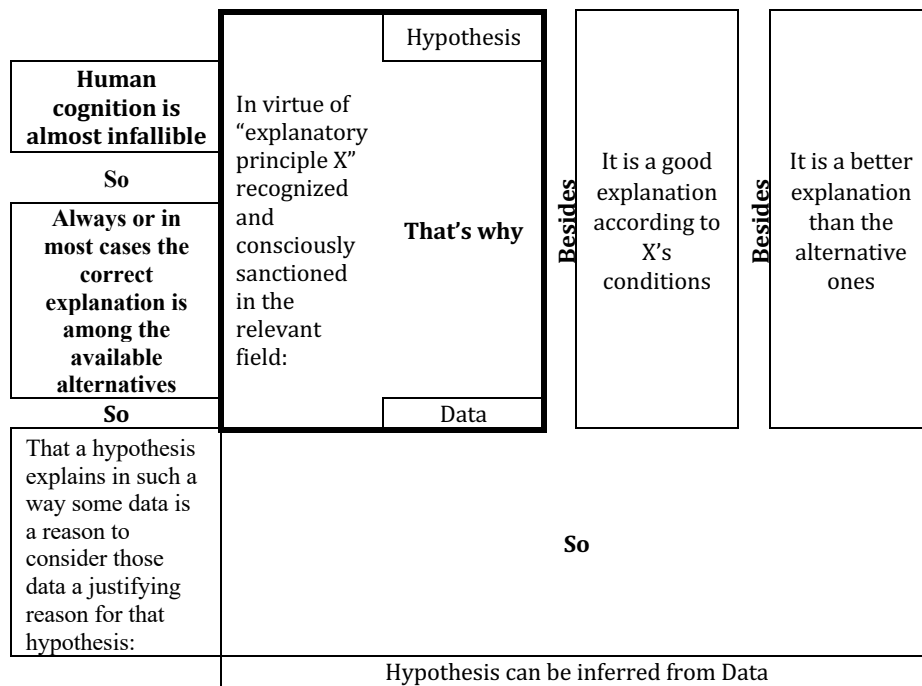
⁶ The terms and context of such a discussion need not be focused on purely “epistemic” concerns, but in fact this is what we mainly find in philosophical literature.

justifying reason for that hypothesis). If we question a principle acting as a warrant (not its limits or scope, but the principle itself) the answer would be an attempt to “back” it, find reasons in its favor. That’s what P. Lipton (1991, 1993), or B. van Fraassen (1989) try to do, albeit adducing really opposite intuitions (cf. Douven 2002, 356-360) and philosophical agendas.

Lipton (who wants to defend the justification of abduction and the feasibility of scientific realism) enunciates an *optimistic* “epistemic principle” as the appropriate backing for our warrant: “the loveliest explanation is probably the likeliest explanation”. It is somewhat redundant but makes explicit the kind of hopeful belief behind our abductive behavior.



Van Fraassen, instead, (who wants to attack abduction, precisely as the *lifeboat* of “scientific realism”) states that the principle could only be established by a rather more polemical epistemic hope: “always or in most cases the correct explanation is among the available alternatives” which, in turn, could only be justified in case “human cognition be almost infallible”. Which is obviously something quite difficult to swallow.



4. CONCLUSIONS

In this paper, I have tried to analyze and categorize, from an argumentative point of view, diverse philosophical approaches and responses to the perceived problem of the justification of abduction.

Part of my interest in this inquiry is to expose not just the argumentative character of philosophical practice (which is a rather widespread assumption) but more precisely its *ordinary* argumentative character, at least in *structural* terms. Philosophical argumentation might be understood and modelled with the same *structural tools* we use to address any other kind of argumentative practice and it is not necessarily more sophisticated than everyday domestic argumentation in those terms.

It addresses, though, substantive questions that are not ordinarily addressed or probed and this implies concentrating on somehow deeper (*grounding*) levels of argument. That's where Toulmin's *distinctions* come to help in characterizing the particularities of philosophical argument. First, there's the distinction between warrant-using and warrant-establishing arguments, which has proved helpful in modelling and understanding discussions regarding "what can be a reason for what" (Cohen, 1986).

Even more significantly, it is Toulmin's whole strategy in dismantling the "undifferentiated premises *plus* conclusion" model –discriminating the different roles the argument's constituents play and the different ways to question them– that has helped us identify specifically *philosophical* concerns regarding the grounds and principles of principles themselves.

Once the iterative and recursive structure of arguing understood, philosophers' concentration of *backings* (for both justificatory and explanatory warrants) is a salient feature of philosophical argumentative practice.

Thus, our metaphilosophical itinerary ends up with the centrality (or at least usefulness) of argumentative analysis for the three levels I mentioned in the Introduction: i) for that of metaphilosophical practice itself, ii) for philosophical practice iii) and for the practices that are of interest for philosophers.

ACKNOWLEDGEMENTS: This work has been made possible by funds provided by the Spanish Ministry of Science, Innovation and Universities through the Research Project PGC 2018-095941B-100, "Argumentative practices and pragmatics of reasons"

REFERENCES

- Álvarez, M. (2016). Reasons for Action: Justification, Motivation, Explanation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition).
 URL = <https://plato.stanford.edu/archives/spr2016/entries/reasons-just-vs-expl/>
- Cohen, J.L. (1986). *The Dialogue of Reason: An Analysis of Analytical Philosophy*. Oxford: Clarendon Press.
- Douven, I. (2002). Testing Inference to the Best Explanation. *Synthese*, 130, 355-377.
- Fraassen, B. van (1989). *Laws and Symmetry*. Oxford: Clarendon Press.
- Fraassen, B. van (1985). Empiricism in the Philosophy of Science. In P. Churchland and C. Hooker (Eds.) *Images of Science: Essays on Realism and Empiricism* (pp. 245–308). Chicago: University of Chicago Press.
- Glennan, S. & P. Illari, eds. (2016). *The Routledge Handbook of Mechanisms and Mechanical Philosophy*. Oxford: Routledge.
- Hacking, I. (1996 [1983]). *Representar e intervenir*, trad. de Sergio Martínez. México: Paidós.
- Joll, N. (2017). Metaphilosophy. *Internet Encyclopedia of Philosophy*. URL: <https://www.iep.utm.edu/con-meta/>
- Lipton, P. (1991). *Inference to the Best Explanation*. Londres: Routledge.

- Lipton, P. (1993). Is the best good enough? *Proceedings of the Aristotelian Society*, 93, 89-104
- Marraud, H. (2013). *¿Es lógic@? Análisis y evaluación de argumentos*. Madrid: Cátedra.
- Marraud, H. (2017). *Guía de Campo de Esquemas Argumentativos*. URL: <https://www.academia.edu/>
- Olmos, P. (2018a). La justificación de la abducción en el contexto del debate sobre el realismo científico: una aproximación argumentativa. *ArtefaCToS. Revista de estudios de la ciencia y la tecnología*, 7(2), 35-57.
- Olmos, P. (2018b). Commentary on Bermejo-Luque's The Appraisal of Conductions. In S. Oswald & D. Maillat (Eds.) *Argumentation and Inference: Proceedings of the 2nd European Conference on Argumentation, Fribourg 2017* (Vol. I, pp. 19-27). London: College Publications.
- Olmos, P. (2019a). Un enfoque argumentativo sobre la abducción y sobre la ponderación de hipótesis explicativas. *Theoria* 34(1), 5-30.
- Olmos, P. (2019b). A meta-argumentative approach to abduction and IBE
- Olmos, P. (forthcoming). Abduction and comparative weighing of explanatory hypotheses. An argumentative approach. *Logic Journal of IGPL*.
- Sonoeck-Henkemans, F. (2003) Complex argumentation in a critical discussion. *Argumentation*, 17, 405-419
- Toulmin, S. E. (2003 [1958]). *The uses of argument*. Cambridge: Cambridge University Press.
- Wagemans, J.H.M. (2016). Argumentative Patterns for Justifying Scientific Explanations. *Argumentation* 30(1), 97-108.
- Waismann, F. (1968 [1956]). How I see philosophy. In R. Harré (Ed.), *How I See Philosophy by F. Waismann* (pp. 1-38). London: Palgrave Macmillan.
- Walton, D.N. (2004). *Abductive reasoning*. Tuscaloosa, AL: The University of Alabama Press.
- Williamson, T. (2007). *The Philosophy of Philosophy*. Malden MA and Oxford: Blackwell.
- Wittgenstein, L. (2001). *Philosophical Investigations. The German Text, with a Revised English Translation*. Malden MA and Oxford: Blackwell.
- Woods, J. (2016). Reorienting the Logic of Abduction. In L. Magnani & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science* (pp. 137-150). Dordrecht: Springer.
- Yu, S. & F. Zenker (2018). Peirce Knew Why Abduction Isn't IBE. A Scheme and Critical Questions for Abductive Argument, *Argumentation*, 32(4), 569-587.

When You Can't Trust What You Don't Know: organ and tissue donation, public trust, the voice of dissent and informed consent

HARMONY PEACH

*Department of Argumentation Studies, University of Windsor, CA
peachh@uwindsor.ca*

This paper examines the apparently good reasons for becoming an organ and tissue donor (OTD) in "opt-in" public recruitment and registration discourse (PRRD). It explores how public trust is employed to encourage assent to the notion that donating is the right thing to do, and it considers whether expert ethical concerns for OTDs omitted from PRRD may offer good reason for public dissent. It also examines how persuasion in PRRD contravenes the tenets of informed consent.

KEYWORDS: organ and tissue donation, public trust, public recruitment and registration discourse, expert ethical debates, dissent, informed consent, presumed consent

1. INTRODUCTION

Good reasons and rational justification for becoming an Organ and Tissue Donor (OTD) are replete in the dominant Public Recruitment and Registration Discourse (PRRD) in so called "opt-in" jurisdictions where the public must actively consent to donation. Government officials, medical and legal professionals, *experts* in the field, contend with, and rely on, public trust to adhere that organ and tissue donation is "the right thing to do." On the one hand, the public's acceptance of the trustworthiness of experts is used to assuage, and even to discount, public dissent to advance a claim that the public *should* assent to organ and tissue donation. This also demonstrates that on the other, the public may not wholly trust that it should and therefore requires expert reassurance. Hence, policy-sanctioned public service announcements, which prompt people to trust the official "ask" for their consent to donate, employ a narrative discourse consisting of "good" reasons for donation which focuses on saving or improving lives. Donation helps people to walk and see again; it restores burn victims. Donors are

framed as “heroes” who bestow “gifts of life.” However, there is also an Expert to Expert Debate (EED) or discourse of dissent within the medical literature about potential risks to organ and tissue donors which seems to either be omitted from or presented as resolved in the PRRD. Debate continues about definitions of death, waiting periods for organ procurement post-declaration of death, and even whether there may be the potential for latent feeling and/or resuscitation. Given that the public is predisposed to view expert discourse as trustworthy, I suggest this EED may then afford the public good reasons for its own dissent, yet it is being discouraged from doing so through the misuse of public trust. Moreover, I argue the EED in the medical dissent discourse suggests the aforementioned “good” reasons for why people *should* or *ought* to donate amount to rhetorical device and hence persuasion in the OTD recruitment and registration discourse which contravenes the avowed tenets of informed consent. The medical voice of dissent suggests serious ethical blind spots for OTDs. Yet, the public is asked, based on the expert discourse disclosed to it, to accept the narrative that leads people to conclude they “do not need their organs and tissues after death” seeming to suggest that OTD “death” is beyond reproach or is definitive in both its medical context and meaning. In this paper I unravel the web woven out of public trust as it relates to the various levels of discourse operational in the context of organ and tissue donation. The relationship between public trust and expert discourse in this field demonstrates that a public grooming process that organ and tissue donation is the “right thing to do” is in contrast to a lack of medical information regarding the procedures involved in the OTD process, which violates trust in informed consent.

2. INTRODUCING TRUST

Apparently good reasons and rational justification are replete in Public Recruitment and Registration Discourse (PRRD) in opt-in (actively expressed consent) and opt-out (presumed consent) jurisdictions. Here, I focus generally on the North American context and explain why the Canadian context is particularly relevant. I suggest the public may have reason for dissent from organ and tissue donation when the PRRD is tested against informed consent tenets in opt-in jurisdictions, particularly when we consider that much of the dissent in the Expert to Expert Debate (EED) is either omitted or presented as resolved in the PRRD. The Journal of Transplantation notes that organ and tissue donation

“constitutes a complex ethical and value laden field of interdisciplinary interventions. It is a surgical and medical

field that requires the highest scientific standards, but likewise one, where ethics, values, and personal beliefs play an immense role. Not surprisingly, then, extensive research has been done on attitudes to [OTD]...often with the explicit aim of investigating whether they “have the knowledge needed to maximize organ donation rates” or “to inform strategies to improve organ donation rates” (Hvidt et. al., 2016, para. 3).

Indeed, maximizing donation rates is often the impetus for much research within the field of organ and tissue donation (OTD). At the crux of the matter for maximization is public trust in the OTD process and the experts within it. In order to unravel the web woven out of public trust in the OTD process, the place to begin is to first discuss the general characteristics of trust, before offering an account of public trust. I can then address (only) some of the PRRD and EED against informed consent tenets in organ and tissue donation in order to pinpoint how some of the arguments for donation are misusing public trust and contravening informed consent.

Generally, researchers are still debating the characteristics of trust, but they have come to some agreement about its key features which bioethicist David Resnik (2010) has summarized nicely. First, trust is a relationship between or among people. It can be explicit (like contractual promises) or implicit (like following traffic rules). It can be concrete (trusting a doctor) or abstract (trusting the medical profession). Trust’s main purpose is to “facilitate cooperative social interactions” (Resnik, 2010, para. 6). It involves risk-taking or a leap of faith. Essentially, you trust in what you do not know for certain. It has a relationship to trustworthiness. The trustor requires evidence that the trustee has qualities which include competence, experience, sound judgment, reliability, good will or benevolence which merit giving them their trust. This is what separates trust from faith which involves belief without evidence. “Trustworthiness can be earned, enhanced, or lost (Baier, 1986, Blomqvist, 1997)” (Resnik, 2010, para. 7). Finally, “trust can generate ethical and legal duties (Baier 1986). The entrusted person has an obligation to do what is expected of [them] in the relationship” (Resnik, 2010, para. 8). Informed consent, for instance, is a legal and ethical obligation and when trust has these obligations it is a form of promise-keeping (Resnik, 2010). The public relies on a hierarchy of medical, legal and policy experts, ordained by governments, for its understanding of, beliefs about, and interaction with the various levels of OTD, particularly for affirming informed consent. Canadian Philosopher Trudy Govier notes in her book *Social Trust and Human Communities*, that we cannot help but to “depend on each other for knowledge and evidence” especially when we consider specialized knowledge and our ability to verify it or not for ourselves through direct

experience (Govier, 1997, p.53). On most topics, most people are not, in fact, experts and therefore “depend on experts for many of their beliefs about these topics” (Govier, 1997, p. 54). And while this does not preclude us from being able to fact check some claims made by others, it is “only by relying on the claims and reports of still other people” that we do so (Govier, p. 54). So, trusting experts, even when their opinions differ, is a necessary part of our decision making. Thus, hearing from those experts who offer varying opinions is also necessary, particularly when informed consent is a standard which must be met when we make our decisions.

3. PUBLIC TRUST

We often use the term *public trust*, but there really is no concise definition of public trust. Again, Bioethicist David Resnik has attempted to tease apart what, exactly, *public* means in public trust, particularly as it relates to scientific discourse. He is motivated by helping various experts make better arguments for public trust in science. He says “[t]he idea that it is important to promote public trust in scientific research has been used by so many different authors in so many different contexts that it is in danger of becoming a platitude. Even worse, overuse of this concept may lead to ambiguity” (Resnik, 2010). In a general sense, what we mean when we say “the public” is actually *society as a whole* and he argues that it makes sense that a society *generally* trusts science to be careful with public resources, and to provide knowledge and expertise that can inform public policy. However, “drawing specific ethical and policy implications from this idea can be problematic because” essentially, society is highly diverse and made up of many publics that may have divergent expectations of science (Resnik, 2010, para. 23) which means there needs to be nuance in how science, and in this case organ and tissue donation PRRD, communicates with different publics which requires trustworthiness.

In their consideration of moving trust towards trustworthiness Aiken, Cunningham-Burley & Pagliari also recognize that “[g]iven the central role of scientific knowledge within society, publics have little choice but to trust in science. But [they clarify] this trust remains conditional and does not mean that [publics] will inevitably have confidence in...scientists or scientific institutions” (Aiken, Cunningham-Burley & Pagliari, 2016, para. 14). Moreover, they say a binary of either the public trusts or distrusts is not nuanced enough to capture the complexity of the trust relationship the public has with science. Instead, the way to advance public trust is not to focus on creating an automatic trust response, but to have scientific institutions and their experts consciously share the trust responsibility. They argue “[t]here is a need

for more symmetrical and reflexive considerations of what it means for publics to trust science, and equally of what it means for science to be trustworthy” (Aiken, Cunningham-Burley & Pagliari, 2016, para. 17) and I agree. Organ and tissue donation sits somewhere between the general sense of public trust as ‘society as a whole’ because the goals of those seeking public trust are to maximize donor rates which involves trust from everyone, and ‘individual publics’ because there are cases that differ across context and culture like religious practices and recruitment differences by region. For instance, in the United States the PRRD often speaks to tissue donations helping military veterans to walk again, but this is less prevalent in the Canadian recruitment context. So, OTD PRRD often seeks to establish, maintain or build public trust in a general sense, and in a more targeted or specifically contextual sense. Prospective donors or their proxy decision makers, if they choose donation (and are not themselves experts in the field), then trust what they do not know directly for themselves because they are relying on the PRRD presented by experts which shapes their beliefs about consenting to donation.

There are many reasons why people do not agree to donation, and some of those reasons are motivated by mistrust. I will handle two such reasons here: 1) there are concerns about donors not truly being dead, and 2) whether they will receive adequate healthcare if they register consent. These can be characterized as necessary public dissent as the EED will later show. Meanwhile, the PRRD not only presents “facts” about OTD, but reasons for consenting to donation in order to increase public trust for the express purpose of increasing donation and quieting dissent. This nexus is where I argue the discrepancy exists between the PRRD in opt-in jurisdictions and informed consent tenets, and where what may be necessary public dissent is being dismissed as a need for more education while either omitting or downplaying similar dissent in the EED.

3. PUBLIC TRUST

The general tenets of informed consent are similar across jurisdictions, particularly in North America in both healthcare and medical research and are a necessary standard which must be met in order for people to opt-in to donation. Distilled, in order for informed consent to be affirmed, it must have been given freely and voluntarily, free of fraud, misrepresentation, coercion or manipulation and have involved a reasonable disclosure of all the facts needed for a reasonable person to make a choice that demonstrates risk and benefit assessment and that offers a dialogue for any needed clarification (Canadian Medical Association, 2014; American Medical Association, 2019).

In the Canadian context, The Canadian Medical Association states “[t]he purpose for providing the opportunity to choose to donate organs or tissues may be to procure organs or tissues for transplant. [But] [s]uccess in achieving this outcome should not be construed as a criterion for measuring the quality of the process of free and informed decision making. *The quality of this process depends on whether the choice is adequately informed and voluntary and not on whether the outcome is a decision to donate*” (Canadian Medical Association, 2014, p. 2, emphasis added). Therefore,

[i]n order for the choice to donate organs or tissues to be duly informed, prospective donors or proxies should be provided with meaningful, understandable information pertinent to the choice. [Most relevant for the purposes here, this] includes information about... the benefits and *risks* of donation... procedures concerning the determination of death... measures that may be required to preserve organ function until death is determined and surgical procurement can occur... [and] what will happen to the body once death has been declared (Canadian Medical Association, 2014, p. 2-3, *emphasis added*).

Essentially, people need to make the choice freely, voluntarily, and be given information about the donor process that includes the *benefits and risks of donation*; all “requiring the exchange and understanding of information and absence of coercion” (Canadian Medical Association, 2014, p. 2), fraud and misrepresentation (Health Care Consent Act, 1996).

In ‘*opt-in*’ jurisdictions, people must actively provide their explicitly *expressed consent* to donate which **requires informed consent**, and trust is used in recruitment and registration. In ‘*opt-out*’ jurisdictions *presumed consent* means people must explicitly withdraw their consent for donation and trust is used in retention. In both cases, the family can still hold the final say over donation, even by overriding registered consent so, trust is operational here too. However, a more recent practice in opt-out countries is that organs will be procured from eligible donors even if the family cannot be reached for consultation. The World Health Organization (2015) suggests moving towards a presumed consent model globally because there is evidence it yields substantially more donations. However, the British Medical Association has reported that a taskforce review shows not all opt-out jurisdictions have high rates of donation. It does not overtly state this is because of a lack of public trust, but it did advise that “improvements in... education, including public awareness campaigns” would significantly increase donation. These solutions hinge on high public trust (Prabhu, 2018, para. 6). The nuances between opt-in and opt-out are especially topical

in the Canadian context where OTD is currently opt-in, and handled provincially. However, the province of Nova Scotia has most recently decided to implement the opt-out consent model come 2020 while the rest of Canada, at this point, will remain opt-in. As this consent softening in one province may lead to justification for others to follow suit, teasing apart these complex relationships is necessary, especially when considered against informed consent in the opt-in model.

4. ARGUMENTS IN THE PRRD

Because the need for organs and tissues outweighs the available supply, the arguments presented in PRRD spend much effort offering apparently good reasons for becoming an OTD with the explicitly stated goal, not of informing consent, but of maximizing (or at least improving) donation rates. Most of these good reasons, however, are based on altruism which uses scientific expert discourse as evidence for its justification. This altruistic angle has a recognizable rationale that transcends geography across opt-in jurisdictions and relies generally on enthymeme for its effect which is ultimately meant to simultaneously encourage active consent registration, often through emotion, while discouraging public dissent. Essentially, both the stated and unstated are crafted to persuade the public to arrive at a decision that OTD is “the right thing to do” or what people should or ought to do. For instance, the PRRD features versions of providing people with the “opportunity” to “save or improve lives”; the “gifts” of solid organs allow the dead to “live on” through their “donations”; and “gifts” of tissue “help people to walk or see again or help to restore burn victims” or “improve lives.” Such narratives as “you can’t take your organs and tissues with you when you die so why not give them to someone who needs them and save a life?” are ubiquitous. Donors are regularly framed as “heroes” and PRRD often suggests that people “ought” to or “should” or “be responsible” or “kind” and “do the right thing” and register their consent to donation. Moreover, it has become a ubiquitous argument to assert that public dissent amounts to “barriers” of ignorance: a need for addressing questions purely for clarification and education, and a public belief in myths which need debunking so people will “make the right choice” to donate. Thus, the PRRD generally offers a list of “facts” which are meant to assuage “fears”, “answer frequently asked questions” and ultimately, to argue that the information presented by the PRRD about the practice of OTD is trustworthy, so people *should* choose to consent. Essentially the PRRD provides people with reasons specifically towards one choice-consent. If they do not choose consent, enthymematically, they are “making the wrong choice” or need better education so they can make the right choice or are selfish, or not heroes or kind, which I suggest is a

coercive framing of the discourse which poses serious concerns about just how much freely decided and informed “choice” is involved in the consent process. Deliberately peppered enthymeme in the framing of PRRD leaves the prospective donor expected to draw the conclusion that they will not be responsible or kind, or heroic if they choose the wrong option.

Some specific examples that are not at all unique include one from California’s online registration and recruitment website www.donatelifecalifornia.org where its “vision is that one day all Californians will embrace organ, eye and tissue donation as their personal responsibility” (Learn More About Donate California, 2019, para. 2). Unstated, is that those not donating are shirking a personal responsibility. In New York, the www.donatelife.ny.gov website says “[y]our kindness could save eight lives through organ donation, restore sight with cornea donations and improve 75 more lives with tissue” (Register to Become an Organ and Tissue Donor Today, n.d., para. 2). Unstated is that people who do not donate are unkind and do not want to save lives or help people see or improve the lives of others. In Ontario, Canada the registration and recruitment website www.beadonor.ca offers “[r]egistering to be a donor makes you a hero” (www.beadonor.ca More Ways to Register, 2017, para. 1) and that “[b]y registering, you are essentially making a decision to help save lives after death through organ and tissue donation” (www.beadonor.ca About Donation: Donation Process, 2017, para. 2). Here what is missing is that people will be villains or the opposite of heroes; by not registering they are essentially making a decision to end lives. Even the various names of donor registration and recruitment institutions are often directives like “be a donor” and “donate life” which also has subtle negative implications for free choice in informed consent.

5. ARGUMENTS IN THE EED

Further confounding the ethical implications for informed consent are the arguments occurring in the EED which also feature dissent, but are either omitted or presented as resolved in the PRRD. The guiding ethical principle in declaring death in OTD is the Dead Donor Rule (DDR) which is a “deontic constraint that categorically prohibits causing death by organ removal” (Nair-Collins & Green, 2014, para. 1). So, a declaration of death *must* be made before donation. There are two ways a person can die for OTD though not all are accepted definitions everywhere. The first is death by neurological criteria or “brain death”. Following no response to a series of neurological tests by two separate doctors, a person is determined to have no brain function, and will remain on life support for the purposes of donation. The second definition is Donation

after Circulatory Death or DCD which occurs when a person is removed from life support (and/or has wishes not to be resuscitated if they code while on life support). It is characterized by the irreversible cessation of cardiac and respiratory functions... [However] the brain is still capable [of] sustaining consciousness, integrative and vital functions" which is why there is a waiting period following DCD before organ retrieval. (Canadian Medical Association, 2014, p. 4-5).

Despite its generally well-regarded rigour, some deeply respected experts in the medical community (Verheijde, Rady, & McGregor, 2009) still question the efficacy of the brain death test which amounts to EED dissent. Generally, the brain death EED points to patients who meet all the criteria for brain death but do not in fact have "irreversible cessation of all functions of the entire brain," because some of the brain stem's homeostatic functions remain, such as temperature control and water and electrolyte balance, and some patients enter puberty or continue to carry pregnancies to term. To counter, others have argued that not all the functions of the brain need to be lost for a patient to be dead, only those that are critical to maintaining integration of the body functions, and that loss of these will inevitably lead over hours or days to cardiac arrest, even with continuing intensive life-support. Yet occasionally the bodies of some patients who meet all the criteria for brain death can survive for many years with all their bodily functions intact except for consciousness and brain stem reflexes" (Sade, 2011, p. 146). There is also a small contingent of the EED that questions whether donors in this state have some potential for pain or some latent, untestable awareness (Verheijde, Rady, & McGregor, 2009). EED dealing with circulatory death hovers on the practice itself and waiting periods which are in place to ensure enough time has passed for the body's death process to terminate, but not so long as to lose viable organs. As little as 75 seconds has been waited in Denver, Colorado and as long as 20 minutes is the standardized wait in France (Aita, 2000; Dead Enough, 2014). In Canada the standard is five minutes, but some hospitals wait 10 and there are variations on whether institutions accept DCD at all. (Dead Enough, 2014, para. 13-14).

EED also questions whether what is currently transpiring in OTD violates the Dead Donor Rule and therefore whether it should be abolished altogether (Rodriguez-Arias, Smith & Lazar, 2011). Given that live organs are procured from dead donors (according to the widely accepted medical and legal policies and practices), one side of the debate argues there will always be problems associated with the practice. Some, like Dubois (2011), insist even without the DDR, many people would still want a ruling of death before organ removal as not having one would create fears among donors and physicians. Others say the DDR death ruling allows for the public belief that donors cannot be

harmed during procurement as dead people cannot be aware of or feel pain. Here, dissent in the EED questions not at which point death occurs, as mandated by the DDR, but instead they want to debate when it might be acceptable to procure organs from dying, but not dead, patients which they argue is more akin to what is actually transpiring currently despite the DDR standard (Rodriguez-Arias, Smith & Lazar, 2011). I have not been able to find evidence of disclosure of this EED in the PRRD. I have found extensive discourse that assures people they will be “dead” when organs and tissues are procured which omits the dissent aspect in the EED from the public’s risk assessment necessary for informed consent. I have also found extensive evidence that employs the prospective donor’s supposedly non-debatable “death” as a kind of guilt mechanism which is meant to quash its dissent. It is because of the above dissent in the EED that the public may indeed have justification for its own dissent from OTD. If we must on some level trust the experts, then their dissent matters to informed consent as much as their arguments for OTD.

CONCLUSION

I have only been able to scratch the surface here in order to offer some of the concerns related to public trust in OTD with regard to discrepancies in PRRD and EED as compared to the tenets of informed consent. My purpose here was not to evaluate arguments, but to begin to tease apart how trust, and specifically “public trust”, are being deployed in the practice of recruitment, and to show how and where it may be being used to increase donation rates while not wholly enforcing the tenets of informed consent. As I mentioned at the outset, OTD is complex, and public trust is just one of the many moving parts which animate the various concerns within the practice. The purpose of this work has been to begin to sketch an account of how public trust is at work in OTD, and whether the ways in which it is being utilized is discouraging what may be necessary public dissent within the OTD practice.

ACKNOWLEDGEMENTS: Thank you to Dr. Brian Brown. His guidance supervising my master’s research which performed a Critical Discourse Analysis of the Ontario-based BeADonor.ca website was invaluable. Thank you to Dr. Suzanne McMurphy for her contributions and discussions with regard to trust and trustworthiness. The members in the audience of the Centre for Research and Reasoning, Argumentation and Rhetoric (CRRAR) at the University of Windsor are all owed a debt of gratitude for their comments and questions. Finally, thank you to the

members in the audience at the ECA presentation of this material for their thoughtful questions and insights.

REFERENCES

- Aitken, M., Cunningham-Burley, S., & Pagliari, C. (2016). Moving from trust to trustworthiness, Experiences of public engagement in the Scottish Health Informatics Programme. *Science and Public Policy*, 43(5), 713-723. doi:10.1093/scipol/scv075
- American Medical Association. (2019). Informed Consent. Retrieved from <https://www.ama-assn.org/delivering-care/ethics/informed-consent>
- Canadian Medical Association. (2014). *Organ and Tissue Donation and Transplantation* (pp. 1-7, Publication). CMA Policy. Retrieved April 12, 2017 from <https://www.cma.ca/Assets/assets-library/document/en/advocacy/PD14-08-e.pdf>
- Govier, T. (1997). *Social trust and human communities*. Montreal: McGill-Queens Univ. Press.
- Hvidt, N. C., Mayr, B., Paal, P., Frick, E., Forsberg, A., & Büssing, A. (2016). For and against Organ Donation and Transplantation: Intricate Facilitators and Barriers in Organ Donation Perceived by German Nurses and Doctors. *Journal of Transplantation*, 2016, 1-13. doi:10.1155/2016/3454601
- Learn More About Donate Life California. (2019). Retrieved from <https://donatelifecalifornia.org/about-us/>
- Nair-Collins, M., & Green, S. R. (2014, September 26). *Abandoning the dead donor rule? A national survey of public views on death and organ donation*. Retrieved December 10, 2016, from <http://jme.bmj.com/content/early/2014/09/26/medethics-2014-102229.full>
- Prabhu, P. K. (2018). Is presumed consent an ethically acceptable way of obtaining organs for transplant? *Journal of the Intensive Care Society*, 20(2), 92-97. doi:10.1177/1751143718777171
- Register to Become an Organ and Tissue Donor Today. (n.d.). Retrieved from <https://donatelife.ny.gov/about-do>
- Resnik, D. B. (2010). Scientific Research and the Public Trust. *Science and Engineering Ethics*, 17(3), 399-409. doi:10.1007/s11948-010-9210-x
- Rodriguez-Arias, D., Smith, M. J., & Lazar, N. M. (2011). Donation After Circulatory Death: Burying the Dead Donor Rule. *The American Journal of Bioethics*, 11(8), 36-43. doi:10.1080/15265161.2011.583319
- The Lancet. (2016). Organ donation depends on trust. *The Lancet*, 387(10038), 2575. doi:10.1016/s0140-6736(16)30886-8
- Verheijde, J. L., Rady, M. Y., & McGregor, J. L. (2009). Brain death, states of impaired consciousness, and physician-assisted death for end-of-life organ donation and transplantation. *Medicine, Health Care and Philosophy*, 12(4), 409-421. doi:10.1007/s11019-009-9204-0

- World Health Organization. (2013, December 11). Donation and transplantation. Retrieved from <https://www.who.int/transplantation/donation/en/>
- World Health Organization. Increasing organ donation by presumed consent and allocation priority: Chile. (2015, February 26). Retrieved from <https://www.who.int/bulletin/volumes/93/3/14-139535/en/>
- Jackson, S., & Jacobs, S. (1980). Structure of conversational argument: Pragmatic bases for the enthymeme. *Quarterly Journal of Speech*, 66(3), 251-265.
- Tindale, C. W. (2015). *The philosophy of argument and audience reception*. Cambridge: Cambridge University Press.
- Wenzel, J. W. (1990). Three perspectives on argument: Rhetoric, dialectic, logic. In J. Schuetz & R. Trapp (Eds.), *Perspectives on argumentation: Essays in honor of Wayne Brockriede* (pp. 9-26). Prospect Heights: Waveland.

“The worst ever conceived by a man of genius”
Hume’s probability argument in *A Treatise*

JEANNE PEIJNENBURG
University of Groningen, Netherlands
jeanne.peijnenburg@rug.nl

DAVID ATKINSON
University of Groningen, Netherlands
d.atkinson@rug.nl

The probability argument in Hume’s *A Treatise of Human Nature* (Section 1.4.1) has been widely criticized, with David Stove calling it “the worst [argument] ever conceived by a man of genius”. We explain that the argument is open to two interpretations: one that is in accordance with probability theory and one that is not. We surmise that Hume failed to distinguish between the two, and that this contributed to the confusion surrounding the argument.

KEYWORDS: diminution, Hume, infinity, iteration, probability, regression, *Treatise*

1. INTRODUCTION

In *A Treatise of Human Nature* David Hume presents an argument which purports to show that, if we rely purely on reason and ignore the sensitive part of our natures, then all our beliefs will be destroyed. In the literature, the argument has been given several names: Hume’s probability argument, the ‘probability reduces to nothing argument’, the ‘iterative probability argument’, ‘the reductio’, or the ‘regress argument’. The section in which the argument occurs (‘Of scepticism with regard to reason’, *Treatise* 1.4.1), has been deemed indispensable for grasping Hume’s theory of reason. In the words of William Morris:

If we ever are to understand Hume’s view of the role of reason, ... we should first figure out how to integrate ‘Of scepticism with regard to reason’ into the picture (Morris, 1989, p. 58).

Yet the argument itself has been widely criticized. David Stove (1965, p. 174) went so far as calling it “the worst ever conceived by a man of genius”, while Robert Fogelin (1985, p. 16) and Mikael Karlsson (1990, p. 126) dubbed it simply “a morass”.

Recently the argument has attracted new attention through the work of Don Garrett, David Owen, Donald Ainslie and others, which has greatly improved and deepened our understanding of this “least understood” (Morris, 1989, p. 58) passage in Hume’s writings. Still there remain confusions. In this paper we surmise that they spring from Hume’s failure to distinguish between two readings of his argument, one that is in accordance with probability theory and one that is not.

In Section 2 we describe Hume’s argument in some detail. In Section 3 we discuss two different ways in which it has been analysed, a formal and an informal one, and we conclude that neither is satisfactory. In Section 4 we explain the precise sense in which the argument fails and in which it is correct.

2. THE PROBABILITY ARGUMENT IN *TREATISE* 1.4.1

Hume’s argument basically consists of three steps. The first encompasses the idea that all “knowledge degenerates into probability” (T 1.4.1.1).¹ By this Hume means that we can never know for sure whether a particular proposition is true. This applies not only to empirical propositions, but also to mathematical ones. In fact, Hume begins the section ‘Of scepticism with regard to reason’ with a reflection on the demonstrative sciences:

In all demonstrative sciences the rules are certain and infallible; but when we apply them, our fallible and uncertain faculties are very apt to depart from them, and fall into error. We must, therefore, in every reasoning form a new judgment, as a check or controul on our first judgment or belief; ... this means all knowledge degenerates into probability; and this probability is greater or less, according to our experience of the veracity or deceitfulness of our understanding, and according to the simplicity or intricacy of the question (T 1.4.1.1).

¹ *A Treatise of Human Nature*. References to this work are given by “T”, followed by four numbers, which indicate Book, Part, Section, and paragraph as in the volume edited by D.F. Norton and M.J. Norton, Oxford University Press, 2006 (first published 2000).

An example may help to understand this first step in the argument. Imagine that we have just performed by hand the addition of the first thousand natural numbers, and that we concluded:

A: The sum S is equal to 500500,

where S equals $1+2+3+ \dots +1000$. Hume's point is that we can never know that A is correct – there is always the possibility that we made a mistake. Of course, we can ask colleagues to do the addition, and if they arrive at the same result this will raise our confidence that A is true, but the salient point is that we can never be sure. The best we can do is to say that A is *probably* true. So our belief in A is gradual, and it leads to a belief in a new proposition, B :

B: A is probably true.

The second step is to apply this reasoning to B itself. For we cannot be certain that B is true either; the best we can say is that it is probably true, which leads to proposition C :

C: B is probably true.

And so on. A more quantitative version of the second step yields:

B: $P(A) = x$
 C: $P(B) = y$
 D: $P(C) = z$

and so on, where x, y, z are values between 1 and 0. Thus the second step gives the argument the form of a regress of higher and higher order subjective probability judgements:

$P(P(P(A)=x)=y)=z \dots$

According to Hume this regress is vicious, because it will inevitably lead to the conclusion that the subjective probability or credence for the first statement, $P(A)$, is zero:

at last there remain nothing of the original probability, however great we may suppose it to have been, and however small the diminution by every new uncertainty. No finite object can subsist under a decrease repeated in infinitum; and even the vastest quantity, which can enter human imagination, must in this manner be reduc'd to nothing (T 1.4.1.6).

Together the two steps entail “a total extinction of belief and evidence” (T 1.4.1.6). This may be a welcome conclusion for “those sceptics, who hold that all is uncertain, and that our judgment is not in *any* thing possest of *any* measures of truth and falsehood” (T 1.4.1.7), but Hume hastens to say that he is no part of “that fantastic sect” (T 1.4.1.8). Although he declares that one “can find no error” in the above steps (ibid.), he recalls that we *do* have beliefs, both in philosophy and in daily life.

Then he takes the third step, which is to say that the former two steps reveal what would happen if reason were left to its own devices: reason would simply annihilate itself and all our beliefs would “terminate in total suspence of judgment” (ibid.). He concludes that reason is “deriv’d from nothing but custom” and that belief is “more properly an act of the sensitive, than of the cogitative part of our natures” (ibid.). Hume’s argument is therefore a regress argument, but it is also a reductio. It shows that he regress leads to an absurdity (namely that we do not have any beliefs or any knowledge), and the way out is to realise that we should not, and in fact do not, rely on reason alone.

3. FORMALIST AND ANTI-FORMALIST APPROACHES

Among the many disagreements that Hume’s argument has provoked, there is the controversy about whether or not formal tools, especially taken from probability theory, can be used to understand and evaluate the argument. Some think they can, and we will call them the ‘formalists’. Others, the ‘anti-formalists’, are strongly opposed to using probability theory. In this section we explain their positions further.

Formalists tend to point out that Hume talks about subjective degrees of probability and strongly suggests that these degrees can be measured. Moreover, formal probability theory was very much in vogue during Hume’s lifetime: Jacob Bernoulli’s *Ars Conjectandi* had just been published, and Hume was a contemporary of Thomas Bayes, whose famous essay on probability was posthumously published by Richard Price, Bayes’s literary executor and a friend of Hume. Doubtlessly Hume realized that formal probability theory existed, and that some of his contemporaries were making significant contributions to it.²

² According to Bernard Peach, Richard Price convinced Hume that some part of his reasoning was inconclusive (Peach, 1980). Since the regress argument in 1.4.1 is among the arguments from the *Treatise* that are not repeated in Hume’s later writings, David Raynor suggested that Price, perhaps in early conversations, convinced Hume that this piece of reasoning is incorrect (Raynor, 1981). Price explicitly criticizes Hume’s regress argument in *A Review of the Principal Questions in Morals* of 1787 (albeit not very successfully: Price

Among the formalists there are some well-known names: C.S. Peirce (1905), G.H. Von Wright (1941), W.V. Quine (1946/2008), and R. Popkin (1951). They all explicitly or implicitly assume that Hume's reasoning in the first two steps of his argument involves a simple multiplication of probabilities. This can be explained as follows.

According to Hume, we cannot be certain that A is correct because we cannot fully trust our calculational capabilities: there is always the possibility that we made a mistake. Now suppose we trust our calculational abilities only to at least 75%. So we believe

$$B: P(A) \geq \frac{3}{4}.$$

However, we are not sure of B either. Suppose we trust it also to a degree of at least 75%, so we have C :

$$C: P(B) \geq \frac{3}{4}.$$

The same goes for C , and D , and so on. The formalists then appear to assume that Hume reconstructed the unconditional probability of A as:

$$P(A) = P(A|B)P(B). \quad (1)$$

Formula (1) is of course incorrect, and in Section 4 we will identify this error as the heart of what goes wrong in Hume's argument. Here we restrict ourselves to the observation that none of the formalists criticized Hume for having used the wrong formula (1). If they have criticized Hume at all, then it is because Hume apparently assumed that a product of factors smaller than one always yields zero (we will shortly return to this).

Similarly, formalists assume that Hume sees the probability of B as $P(B|C)P(C)$, and if we insert the latter formula into the right hand side of (1), we obtain a new formula for $P(A)$:

$$P(A) = P(A|B)P(B|C)P(C). \quad (2)$$

By the same procedure, insertion of $P(C|D)P(D)$ for $P(C)$ in (2) gives us an even longer formula for $P(A)$, namely

$$P(A) = P(A|B)P(B|C)P(C|D)P(D). \quad (3)$$

argues that doubting one's doubt of A will make one believe A *more* because the higher order doubt cancels the doubt of a lower order).

If we repeat this procedure infinitely many times, then we evidently will end up with an infinite chain:

$$P(A) = P(A|B)P(B|C)P(C|D)P(D|E) \dots \quad (4)$$

in which the right hand side contains only conditional probabilities. In each of the four formulas above, $P(A)$ is a product of factors all less than one. Hence the longer the formula is, the smaller $P(A)$ will be, and in the limit that the chain goes to infinity, $P(A)$ will converge to zero. Thus Richard Popkin concludes:

Since [the] probabilities are smaller than 1, the product is smaller than either of them. ... This process of introducing new probabilities ... can go on ad infinitum, and thus, the probability that we could ever recognize ... that a particular piece of reasoning was correct, approached to zero (Popkin, 1951, p. 390).

Popkin does not seem to find anything wrong with this reasoning, and many formalists appear to have come to the same conclusion. Some formalists, however, have criticized Hume on the grounds that a multiplication of numbers smaller than one need not yield zero. A necessary condition for this to occur is that the higher order probabilities approach ever closer to one. Thus Quine (1946/2008) pointed out that in very special cases the product of numbers smaller than one might be positive, and he reproached Hume for having failed to see this.

Quine's criticism is however beside the point. For it is clear that Hume is not talking about these special cases. Hume is addressing the situation where a continual diminution takes place: he talks about something that in the end becomes nothing at all. It is simply irrelevant to explain, as Quine does, that Hume's reasoning in very exceptional circumstances has a non-zero outcome.

Anti-formalists such as David Owen (1999, 2004, 2015) and Don Garrett (2000, 2004, 2006, 2015) vehemently deny that the probability calculus can help us to understand Hume's argument. In their view, a formal rendering is not only useless, but actively blocks an understanding of what Hume was after. It can be noted that nowadays practically all the scholars who have studied Hume's probability argument adopt a more or less anti-formalist approach.

Anti-formalists do seem to agree with the formalists that, according to the formal calculus, $P(A)$ is computed as a multiplication that converges to zero in the limit. David Owen even calls this a "mathematical truism" (Owen, 2015, p. 114). However, he maintains that Hume cannot have had this alleged mathematical truism in mind.

The main reason is that such a truism has very unhumane consequences. After all, if in the limit $P(A)$ is zero, then in the limit $P(\neg A)$ is one. This would mean that we have certainty after all, and this goes against everything Hume is trying to do in T 1.4.1. In the well-chosen word of David Owen:

The point of Hume's argument is 'the total extinction of belief and evidence' ... It is a sceptical argument, not the argument of a negative dogmatist (Owen, 2015, p. 114).

Anti-formalists conclude that the word 'probability' in Hume's argument is not 'probability' as explicated in the calculus. It rather means 'force', 'vivacity' or 'retention' – all notions that cannot be explained in terms of formal probability theory.

In the next section we explain why we think this conclusion is too quick. We will argue that the standard formalist reading of Hume's argument, as explained above, is not in accordance with the probability calculus (although it may be in accordance with what Hume had in mind). If we reconstruct Hume's argument in a way that *is* in agreement with the calculus, then it becomes clear that there is indeed something that goes to zero, although it is different from what Hume may have meant (*cf.* Atkinson & Peijnenburg *forthcoming*).

4. NOT A PRODUCT, BUT A SUM

In this section, we argue for two claims. The first is that both formalists and anti-formalists are mistaken when they assume that, according to the calculus, $P(A)$ goes to zero in the limit. Not only is it not a mathematical truism, as Owen maintained, it is simply false. The reason why both factions made the mistake is that, as we will explain, both saw $P(A)$ as a product, whereas it is a sum. The second claim is that in a correct formal rendering of Hume's argument something goes to zero, but it may be something other than what Hume had envisaged.

Let us start with the first claim. We have seen how the formalists reconstruct Hume's argument. If we have proposition A (in our example: 'The sum S is equal to 500500'), and we believe to at least 75% that A is true, then we have a new belief B : $P(A) \geq \frac{3}{4}$. Since we also trust B to a degree of at least 75%, we believe C : $P(B) \geq \frac{3}{4}$. And so on. As we have seen, formalists implicitly or explicitly assume:

$$P(A) = P(A|B)P(B). \quad (1)$$

But (1) is wrong. In determining the probability of A on the basis of B , we should also take into account what the probability of A is given that B is false. So rather than (1) we have

$$P(A) = P(A|B)P(B) + P(A|\neg B)P(\neg B). \quad (1')$$

which is not a product, but a sum. Of course, the same goes for $P(B)$, which is not given by $P(B|C)P(C)$ but by:

$$P(B|C)P(C) + P(B|\neg C)P(\neg C),$$

and similarly for $P(C)$, $P(D)$, et cetera. It is somewhat puzzling that none of the formalists have noticed this. Perhaps it is because in their lifetime the application of formal methods to philosophical problems was not as common as it is today (note that most of the formalists we mentioned wrote their works quite some time ago). Or perhaps the formalists were, like Quine, focussed on the fact that a multiplication of factors smaller than one may in exceptional circumstances yield a non-zero number, and consequently overlooked the fact that Hume made a much more fundamental mistake. Be that as it may, if we use the correct formulas for $P(A)$ and $P(B)$, then what we obtain as the new value for $P(A)$ is not, as the formalists thought,

$$P(A) = P(A|B)P(B|C)P(C), \quad (2)$$

which is a multiplication, but rather

$$P(A) = P(A|B)[P(B|C)P(C) + P(B|\neg C)P(\neg C)] + P(A|\neg B)[P(\neg B|C)P(C) + P(\neg B|\neg C)P(\neg C)] \quad (2')$$

which is a sum.

What happens if we repeat these transformations infinitely many times? The answer is: one still gets a sum rather than a product. Moreover, it can be proven that $P(A)$ converges to a unique and positive number, not zero. We will not stop to give the proof here, but readers who are interested can find it in (Atkinson & Peijnenburg, 2017).

This takes us to our second claim. If we iterate the correct formulas (1'), (2') et cetera infinitely many times, then it turns out that there is *something* that converges to zero. This is however not the probability of A , $P(A)$, but rather the influence exerted on $P(A)$ by the propositions in the chain. The further away a proposition is from A , the smaller is its contribution to $P(A)$, and in the limit this contribution vanishes completely. Again the proof can be found in (Atkinson & Peijnenburg, 2017). Here we restrict ourselves to giving an illustration.

Imagine again that we trust our calculational capabilities to at least 75% (nothing depends on the latter; our argument goes through with any exact or inexact number). Suppose further that our credence in A , given B , is 0.9:

$$P(A|B) = 0.9.$$

And let us suppose that the probability of A , given the falsity of B , is 0.5:

$$P(A|\neg B) = 0.5.$$

In the first instance we assume B to be true, so $P(B)=1$, and from (1') we find $P(A)=0.9$. Let us further assume that the numbers in the rest of the chain are the same (again, this assumption of uniformity is not essential):

$$\begin{aligned} P(B|C) &= 0.9 \text{ and } P(B|\neg C) = 0.5 \\ P(C|D) &= 0.9 \text{ and } P(C|\neg D) = 0.5, \end{aligned}$$

and so on. Now the first humean doubt assails us: we begin to doubt B after all, but we (provisionally) suppose at least C to be true, $P(C)=1$. This allows us to recalculate $P(B)$, which drops from 1 to 0.9; and as a consequence $P(A)$ drops from 0.9 to 0.86. Next we doubt C but believe D fully, and so on. This regress of doubting yields an infinite sequence of revisions of $P(A)$. A few steps are given in Table 1:

no. of propositions	1	2	3	5	10	∞
value of $P(A)$	0.9	0.86	0.84	0.83	0.833	5/6

Table 1. Decreasing higher-order probabilities of A

Two things attract our attention. First, in the limit the final value of $P(A)$ is not zero, but 5/6. Second, the further away a proposition is, the smaller is its contribution to that final value. $P(A)$ with only proposition B is 0.9, but when we also take C into account, then the probability is reduced to 0.86, which means that C contributes a (negative) correction is 0.04. With D , the value goes down still further to 0.844, so D contributes a correction of $0.86 - 0.844 = 0.016$. The combined effect of the sixth to the tenth orders, as can be seen from the table, produces a correction of less than two parts in a thousand.

In Table 1 the probabilities of A decrease, but they could actually increase. Whether they decrease or increase depends on the values of the conditional and unconditional probabilities. Suppose we set $P(A|B)$, $P(B|C)$, and so on, equal to 0.8, and $P(A|\neg B)$, $P(B|\neg C)$, and so on, equal to

0.3, while the values of the unconditional probabilities are 0.5, rather than 1 (the latter reflects the idea that we initially think we might be just as well right as wrong about *B*, *C*, *D*, et cetera). That leads to Table 2:

no. of propositions	1	2	3	5	10	∞
value of $P(A)$	0.5	0.57	0.58	0.59	0.599	$3/5$

Table 2: Increasing higher-order probabilities of *A*

As in Table 1, the value of $P(A)$ is a well-defined number, namely $3/5$. However, in Table 2 the probability goes up rather than down as the number of doubttings increases. Yet the contribution of the higher orders to the final value of the probability of *A* once again decreases. Further it can be proved that this final value, after an infinite number of doubttings, does not depend at all on whether we set the unconditional probabilities equal to a half or to one: it is a function solely of the conditional probabilities.

Both tables illustrate that the probability of the original belief not only fails to go to zero, but generally approaches a positive number that is unique and well-defined; and this is what usually happens (the only situation in which this does not happen is when the – nonuniform – conditional probabilities in the chain rapidly approach 1, that is, when they are close to material implications). Moreover, the tables show that there is something that invariably diminishes as the chain of doubttings increases, namely the effect of higher-order doubttings on the unconditional probability of *A*. The further away a proposition is from *A*, that is the more intermediate doubttings there are, the smaller is its influence on the final value of $P(A)$.

The tables reveal that the approach to the limit can be rather rapid. This should remove any feeling of uneasiness that one might have about drawing conclusions from reasoning that goes on forever. In line with Hume's claim that the diminution already occurs in a finite sequence of doubttings, the tables tell us that we do not need to go all the way to infinity in order to see the effect that we have been talking about: a few steps suffice to indicate that the significance of the higher orders diminishes as their number increases. This is because a few steps are enough to suggest that a regress of higher and higher-order probabilities converges to a non-zero value.³

³ Of course, we need a mathematical proof to demonstrate that what we are actually observing is a firm fact rather than a fluctuation. But this has been provided.

5. CONCLUSION

Anti-formalists have protested that the standard formalist reading of Hume's probability argument in *Treatise* 1.4.1 turns Hume into a negative dogmatist – and they are right. What they appear to have missed, however, is that the standard formalist reading implicitly accuses Hume of having made an elementary formal mistake. That reading is based on a faulty formula for $P(A)$ and wrongly presupposes that the higher and higher order doubts form a multiplication rather than a sum. It thus takes Hume as claiming that the credence or subjective probability in A decreases to zero as the chain of doubts lengthens, and such a claim violates the probability calculus.

In this paper we have investigated what happens if we rectify the formal mistake. If we reconstruct the chain of Humean doubts in a way that agrees with the probability calculus, then we discover that indeed something goes to zero. What decreases is however not the credence in A , nor is it the force or vigour of that credence. Rather it is the contribution to that credence of the successive doubts in the chain. The further away a doubt is from A , the smaller is its contribution, and in the limit the latter peters out completely.

There are thus two formal interpretations of Hume's probability argument, a valid and an invalid one. Which of these interpretations did Hume have in mind? In 'Of scepticism with regard to reason' Hume appears to go back and forth between them: most expressions point to the invalid interpretation, a few indicate that he was thinking of the valid one. We are therefore drawn to the conclusion that Hume failed to distinguish between the two. This conclusion appears to be supported by David Owen's analysis of Hume's argument. Owen, an outspoken anti-formalist, has paraphrased Hume's argument as follows:

As the number of intermediate ideas increases and the chain of reasoning becomes longer, the relationship between the ideas at each end of the chain of ideas becomes more indirect and the certainty of the conclusion is lessened (Owen, 2015, 120).

If Owen is right, then Hume failed to distinguish between a valid and an invalid version of his argument. For probability theory teaches us that the first part of Owen's sentence hits the mark, but the second part is false. It is indeed the case that, as the chain becomes longer, the relationship between the ideas at each end of the chain becomes less direct. It is however not so that the certainty of the conclusion is lessened. No matter how long the chain is, the conclusion can still be almost certain, and moreover be believed with great force and vivacity.

ACKNOWLEDGEMENTS: We thank the audience at the third European Conference on Argumentation in Groningen. This paper was also presented at the Erasmus University Rotterdam, the University of Leiden, and the Heinrich Heine University in Düsseldorf. We thank Maria van der Schaar, and Gerhard Schurz and his group. Special thanks go to Tim de Mey, Fred Muller, Han van Ruler and their students of the Erasmus School of Philosophy for having organized after the talk an unforgettable evening in the local philosophical café.

REFERENCES

- Atkinson, D., & Peijnenburg, J. (2017). *Fading foundations. Probability and the regress problem*. Cham, Switzerland: Springer, Open Access.
<https://link.springer.com/book/10.1007%2F978-3-319-58295-5>
- Atkinson, D., & Peijnenburg, J. (forthcoming). 'Till at last there remain nothing': Hume's *Treatise* 1.4.1 in contemporary perspective. *Synthese*.
- Fogelin, R. (1985). *Hume's skepticism in the Treatise of Human Nature*. London: Routledge and Kegan Paul.
- Garrett, D. (2000). Owen on Humean Reason. *Hume Studies*, 26(2), 291-304.
- Garrett, D. (2004). 'A small tincture of Pyrrhonism': Skepticism and naturalism in Hume's science of man. In W. Sinnott-Armstrong (Ed.), *Pyrrhonian skepticism* (pp. 68-98). Oxford: Oxford University Press.
- Garrett, D. (2006). Hume's conclusions in 'Conclusion of this Book'. In S. Traiger (Ed.), *The Blackwell guide to Hume's Treatise* (pp. 151-75). Oxford: Blackwell.
- Garrett, D. (2015). *Hume*. Abingdon (Oxon) / New York: Routledge.
- Hume, D. (1739/2006). In D. Norton & M. Norton (Eds.), *A treatise of human nature*. Oxford: Oxford University Press.
- Karlsson, M. (1990). Epistemic leaks and epistemic meltdowns: A response to William Morris on scepticism with regard to reason. *Hume Studies*, 16(2), 121-130.
- Morris, W. (1989). Hume's scepticism about reason. *Hume Studies*, 15(1), 39-60.
- Owen, D. (1999). *Hume's Reason*. Oxford: Oxford University Press.
- Owen, D. (2004). Hume and the irrelevance of warrant, manuscript available at <http://davidowencourses.arizona.edu/David's%20Papers/Irrel5a.pdf>
- Owen, D. (2015). Scepticism with regard to reason. In D. Ainslie & A. Butler (Eds.), *The Cambridge companion to Hume's Treatise* (pp. 9-26). Cambridge: Cambridge University Press.
- Peach, B. (1980). Hume's mistake. *Journal of the History of Ideas*, 41(2), 331-334.
- Peirce, Ch. S. (1905). Notes on portions of Hume's *Treatise of human nature*, manuscript.
- Popkin, R. (1951). David Hume: His Pyrrhonism and his critique of Pyrrhonism, *The Philosophical Quarterly* 1(5), 385-407.
- Quine, W.V.O. (1946/2008), Lectures on David Hume's philosophy. In D. Føllesdal and D. Quine (Eds.), *Confessions of a confirmed extensionalist* (pp. 37-137). Cambridge: Harvard University Press.

- Raynor, D. (1981). Hume's mistake: Another guess. *Hume Studies*, 7(2), 164-166.
- Stove, D. (1965). Hume, probability, and induction. *The Philosophical Review*, 74(2), 160-177.
- Von Wright, G.H. (1941), *The logical problem of induction*. New York: Barnes & Noble.

Character, Dog Whistles, and the Limits of Charity

KATHRYN PHILLIPS

University of Rochester

kathryn.phillips@rochester.edu

Both the principle of charity and responsibility condition are thought to be central elements of argument reconstruction and productive discourse. These conditions are problematic in arguments that contain various forms of deception. In this paper, I will focus on multivocal appeals (popularly known as dog whistles,) which are meant to be heard by only certain audience members. I will argue that arguments containing dog whistles require more nuanced tools to reconstruct the argument.

KEYWORDS: [Argument Reconstruction, Dog Whistles, Principle of Charity, Virtue Argumentation]

1. INTRODUCTION

Many philosophers tend to prefer logical models of argumentation¹ because we hope to distinguish good argument from manipulation and bullshit (Frankfurt, 2005). The reliance on truth conditions in logical models, rather than more relativistic criteria such as acceptability or consensus from rhetorical models, appear to give us the tools to make such distinctions. More broadly, argumentation theorists interested in distinguishing argument from related bad practices tend to build in idealizing conditions such as assumptions that argumentative agents are rational, cooperative, and aim to uncover what is reasonable to believe or true. Idealized conditions are present across argumentative traditions—not just the logical tradition.

Sometimes idealizations come in the form of principles of rational engagement. One such principle, some form of which is popular across argumentative traditions, is the principle of charity. From the logical tradition, Richard Feldman says that the “fundamental principle

¹ And also epistemic models that focus on reasonable or justified belief. For an overview of the distinction see: (Lumer, 2005).

of argument analysis is the ‘principle of charity,’” because reconstructing arguments such that they are valid while giving the most charitable interpretation of the premises, implicit premises, and conclusions “leads us to consider the best available arguments and thus to gain the most insight into the issue we are studying” (Feldman, p. 115). In the dialectical tradition, Frans van Eemeren and Rob Grootendorst give us the responsibility condition stipulating that we ought to take the speaker as committed to her stated, externalized claims even if she is lying (Jørgensen, 2007). In the rhetorical tradition, Christopher Tindale claims that argumentation is invitational and that it fundamentally aims at understanding between arguer and audience. This, too, suggests a kind of charitable engagement. Idealizations and abstract principles raise questions about argument and argument evaluation in non-ideal conditions—in particular in common cases of deception. To define deception and ill-intent out of argumentative theories is to exclude a great deal of what appears to be common argumentative practice.

In this paper I will focus on a particularly complicated kind of deception—the linguistic phenomenon that Bethany Albertson calls the “multivocal appeal,” also sometimes referred to as “code words” or “dog whistles,” so called because they target “those predisposed to respond favorably to the message and [go] over the heads of those who might be turned off by it” (Albertson, 2015, p. 4). If dog whistles are meant to convey problematic content, while the same words in the same order are sometimes mere innocent assertions, how and when should charitable readings of arguments that use them be deployed?

I will argue that dog whistles add weight to theoretical perspectives like Tindale’s rhetorical argumentation, which insist on the central role of audience in understanding and evaluating argumentation. In addition, beyond merely giving us further reason to appreciate the necessity of including the situated nature of speaker and audience in argument evaluation, I will argue that dog whistles show virtue argumentation is an excellent complement to rhetorical argumentation and provides the tools for audiences and arguers to effectively discern limitations of the principle of charity in reconstructing and evaluating arguments. Dog whistles show us that charity is not limitless, and taking aspects of the rhetorical perspective such as the focus on audience, as well as the centrality of character in virtue argumentation as necessary starting points for argument analysis we can construct a strong argumentative framework to meaningfully navigate these limitations.

2. DOG WHISTLES

While code words, multivocal appeals, or dog whistles have been studied in politics (Albertson, 2015; Haney-López, 2014) and political psychology (White, 2007) for some time, analysis of this phenomenon in philosophy of language is relatively new. Jason Stanley gives an initial analysis of code words in his 2015 *How Propaganda Works* primarily as a means of explaining a particularly problematic mechanism for eroding democratic norms. Stanley's linguistic analysis of code words depends on distinguishing between at-issue and not-at-issue content. At-issue content refers to content that the speaker is proposing be added to shared common ground, while not-at-issue content is presupposed and not explicitly up for question or debate. Another way to say this is that it would take more work to question not-at-issue content. In the case of code words, Stanley argues that frequent connection between particular words and ideas over time creates not-at-issue-content connected to particular concepts. For example, he says:

When the news media connects images of urban Blacks repeatedly with mentions of the term “welfare,” the term “welfare” comes to have the not-at-issue content that Blacks are lazy (Stanley, 2015).

According to Stanley, code words are a particularly effective tool to erode rational and democratic norms because they give speakers the means to deploy veiled, targeted attacks on particular groups within a society as a means of domination and control, while maintaining some form of plausible deniability that the attack is intentional. Some evidence for Stanley's argument, particularly concerning the erosion of norms, comes from empirical work, such as White (2007), that demonstrates racial attacks are much more likely to be effective when they are disguised by terms like “inner-city” rather than referring directly to negative stereotypes about African Americans.

Several authors have objected to Stanley's linguistic analysis of code words arguing that the non-cancelability of not-at-issue content would undermine the purpose and effectiveness of dog whistles (Henderson & McCready, 2018; Khoo, 2017). For instance, if someone says: “John stopped riding his bike,” it would not make sense for that person to immediately follow that utterance by saying, “John never rode his bike.” Non-cancelability is what explains this confusion—the second utterance would cancel the not-at-issue, that John used to ride his bike, present in the first claim but doing so doesn't make sense. With standard examples of dog whistles, however, the deniability is exactly the point—that is, dog-whistled content *must* be able to be cancelled, and in fact it looks like it can. Taking Stanley's example about welfare

again, if one were to object by saying, “your words are intended to further oppress poor black people,” the speaker could respond without creating any confusion by saying: “I’ve said nothing at all about poor black people – how dare you madam!” And so it seems that dog whistles don’t get their power through encoding actual semantic content, but through some other mechanism.

An alternative analysis by Justin Khoo claims that code words function through the activation of inferences based on background knowledge of stereotypes. According to Khoo, statements like: “The food stamp program will primarily benefit inner-city Americans,” get their racial content from hearers, who may already believe things like “The inner city is mostly populated by poor African Americans,” thus licensing inferences such as: “The food stamp program will primarily benefit poor African Americans” (Khoo, 2017, p. 47). According to Khoo, in these cases speakers may or may not be intending for hearers to make these inferences, but the racist content of such speech relies on activating pre-existing belief that lead to the relevant inference.

Henderson and McCreedy provide an argument similar to Khoo’s, but add that dog whistles also signal speaker personae, suggesting that speaker intent is an important element of dog whistles (Henderson & McCreedy, 2018). So, it continues to be a matter of dispute the extent to which dog whistles involve speaker intention. For Khoo dog whistles are not tied to speaker intent but rather how hearers receive and interpret certain claims based on background information, and according to Khoo this is practically relevant because the most effective way to combat the problematic effects of dog whistles is not call out *speakers* but rather to call out the problematic *inferences* which are likely to be drawn because doing so would avoid the question of speaker deniability by focusing instead on possible interpretations. While there is some appeal to this strategy, speaker character is still relevant to the level of suspicion that might be warranted in encountering particular speakers, even if the best strategy for stripping dog whistles of their power is to call out potential inferences.

On Khoo’s account of dog whistles it seems that a variety of inferences can be licensed from a particular utterance dependent on the listener’s background knowledge and which stereotypical beliefs are triggered by particular phrases. More broadly, we might say that on Khoo’s account it is the interplay of the utterance and the audience that gives it meaning and it would be mistaken to suggest that one discrete argument is generated out of such an exchange. This suggests that a necessary framework for reconstructing and evaluating arguments should begin from theoretical work like Tindale’s rhetorical argumentation, which insists on the central role of audience in understanding and evaluating argument. I will address audience in the

next section. Furthermore, dog whistles provide a strong rationale for resisting the application of universal principles in argument evaluation, such as the principle of charity or universal applications of its negative counterpart—the ad hominem fallacy. Negotiating appropriate limitations on the application of principles goes beyond merely assessing the role of the audience in how speakers construct their arguments and I will argue in the final section that virtue argumentation provides the tools to understand these limitations.

3. ARGUMENT AND AUDIENCE

To understand the importance of audience in argument analysis, a bit of history may be useful. A deep division has developed between logical (and epistemological) models of argumentation that aim at reason or truth and rhetorical models that aim at persuasion and effectiveness. Logical models are common in contemporary philosophy, as we can see from examples like Feldman's *Reason and Argument* where he defines rational argument as a direct contrast to rhetoric. More generally, a variety of post-enlightenment models of argumentation value reason and truth in a manner that is taken to be odds with rhetorical models that insist on important roles for identity, context and the situated nature of argumentation and arguments.

Christopher Tindale has provided a strong argument, especially for those of us who come from the logical tradition, to take seriously the possibility that the rhetorical dimension of argumentation is foundational to logical or dialectical dimensions. For Tindale argumentation is collaborative and invitational and it is the process prior to the product. Audiences are a *central* element of argumentation and their presence ensures that they are co-authors in arguments as the primary arguer must take audience needs into account and shape her reasoning such that the audience can engage. As a result, on this view, both arguer and audience are altered by their interaction. While rhetorical argumentation focuses on audience and social context, many contemporary philosophers would challenge the notion that *who we are* does or should affect the *argument*. Dog whistles give us reason to believe that that's not true.

In order to better understand the importance of audience in evaluating arguments and the challenges dog whistles present, let's consider the following excerpt from Donald Trump's first speech to the UN general assembly:

We appreciate the efforts of United Nations agencies that are providing vital humanitarian assistance in areas liberated from ISIS, and we especially thank Jordan, Turkey and

Lebanon for their role in hosting refugees from the Syrian conflict.

The United States is a compassionate nation and has spent billions and billions of dollars in helping to support this effort. We seek an approach to refugee resettlement that is designed to help these horribly treated people, and which enables their eventual return to their home countries, to be part of the rebuilding process.

For the cost of resettling one refugee in the United States, we can assist more than 10 in their home region. Out of the goodness of our hearts, we offer financial assistance to hosting countries in the region, and we support recent agreements of the G20 nations that will seek to host refugees as close to their home countries as possible. This is the safe, responsible, and humanitarian approach ("Remarks by President Trump to the 72nd Session of the United Nations General Assembly," 2017).

How ought we to understand this excerpt and reconstruct the argument? Using the principle of charity would suggest that we ought to reconstruct a valid argument and take Trump at his word that he is compassionate and wants the United States to do the best that it can, for humanitarian reasons, to help refugees of the Syrian war. We might get an argument that looks something like this:

1. We ought to deal with Syrian refugees in the safest, most responsible, and most humanitarian way possible.
2. The safest way to deal with Syrian refugees is to resettle them as close to their home as possible and to help them return to their homes.
3. The most responsible way to deal with Syrian refugees is to do so in the most cost effective as possible.
4. The most cost-effective way to deal with Syrian refugees is to resettle them as close to home as possible and to help them return to their homes.
5. The most humanitarian way to deal with Syrian refugees is to resettle them as close to home as possible and to help them return to their homes.
6. The safest, most responsible, and most humanitarian way to deal with Syrian refugees is to resettle them as close to home as possible and to help them return to their homes.

So,

7. We ought to resettle Syrian refugees as close to their home as possible and help them return to their homes.

Furthermore, charitably we might suggest that "safe" refers to the refugees themselves given the apparent humanitarian focus of this argument. It appears, especially if taken charitably, to be an argument

about how to best deal with refugees for the refugees themselves, as opposed to an argument about how other countries can avoid what is sometimes perceived as the troublesome influx of refugees. And yet, for anyone who has any knowledge of Trump's previous public remarks or knowledge about his policy proposals (such as a call for a complete ban on Muslims in the United States), would likely object that a charitable analysis, especially one that reads his claims about safety as referring to the refugees *themselves*, looks unjustified. Here we see at least one dog whistle—the surrounding argument as stated suggests a humanitarian claim while simultaneously signaling fear and distrust of Middle Eastern refugees, Muslims in particular. Furthermore, we can see it is packaged up as a message that is palatable to his intended audience—the UN General Assembly. Tindale's work on rhetorical argumentation predicts precisely this—that the construction of this argument and apparent conciliation about Syrian refugees is developed out of interest in communicating with the General Assembly. A charitable reconstruction of Trump's argument would be a mistake both because it misses both the role of audience in shaping these specific remarks, but also because some knowledge of his character as an argumentative (and moral) agent is what allows us to understand the principle of charity does not apply here. A person trying to understand how to treat refugees humanely, or really how to treat refugees at all, who starts with the remarks on a racist fear monger is inevitably making a mistake. To see why, we should return to Stanley's broader point about the ways that propaganda distorts rational and democratic norms.

While Stanley's account of the *mechanisms* that give code words their power may have been mistaken, his account of their *function* remains compelling. According to Stanley, the function of such speech within debate is to silence particular groups while maintaining the guise of rational debate (Stanley, 2015). This is accomplished, in part, by how the debate is framed. He says: "The function of these expressions is to mask the demagogic nature of the contribution, by creating flawed ideological beliefs to the effect that the perspectives of a designated group are not worthy of reasonable consideration" (Stanley, 2015, p. 129). This seems to be at work in Trump's speech if instead of taking the charitable interpretation that when invoking "safety" it refers to the safety of the refugees, or what safety would *mean* from a refugee's perspective, we instead take it to mean safety *from* refugees, which obviously leaves refugees out. This suggests the acceptance of a frame in which the refugees are effectively silenced. In the next section I will argue that virtue argumentation provides the most effective tools for discriminating between genuine and malicious content such that we appreciate the variety of ways that arguments can be reconstructed as well as relevant limitations of the principle of charity.

4. VIRTUE

Virtue argumentation, like its predecessors virtue ethics and virtue epistemology, focuses on the cultivation of character traits rather than the articulation and application of universal principles. In virtue ethics, seemingly intractable disputes between theories that articulate universal principles of right action based on central concepts such as duties or consequences, led to a resurgence in ancient accounts that asked first who we should *be* rather than what we should *do*. For instance, a principle that says one should not lie seems like perfectly good advice in many cases, but also becomes bad advice in cases where the lie is the obviously correct choice. This ranges from relatively benign choices like telling your friend he looks good when he needs it to more unlikely examples such as a murderer asking if you are harboring their intended victim in your home when, in fact, you are.

It is no surprise then, that when looking away from purely theoretical, abstracted argumentation siloed from the practical realm, and instead looking at the interplay of our ideals and our practice, we would also look to virtue in the realm of argumentation.

Part of taking practice seriously includes significant attention to argumentative agents. In a recent introduction to an issue of *Topoi* dedicated to virtue argumentation, Andrew Aberdein and Daniel Cohen say: “A distinctive feature of the aretaic turn in the study of argumentation is its focus on agents: arguers, rather than (just) arguments” (Aberdein & Cohen, 2016, p. 340). Virtue argumentation is still quite new, and while there are many details to be figured out, the centrality of argumentative agents is fundamental to virtue argumentation. This gives us the room to evaluate argumentative character traits in thinking through how to reconstruct arguments rather than relying on broad principles, such as the principle of charity, alone. Virtue argumentation importantly connects the argument(s) with the arguer, and Trump often demonstrates argumentative vices such as unwillingness to change position, unwillingness to listen to others,² and other more obvious vices such as mendaciousness, which should make a careful interlocutor approach engagement in argumentation with Trump more cautiously. This means that the virtuous argumentative agent has more nuanced tools for argument reconstruction particularly in cases where the virtuous arguer has evidence of their interlocutor’s character. This doesn’t suggest that all arguers should be approached with suspicion—for instance the blanket application of a hermeneutic of

² These vices come from the taxonomy of argumentative vices developed in (Andrew Aberdein, 2016)

suspicion would be equally misguided because one might not yet have reason to be suspicious.

Given the agent-centric nature of virtue argumentation, it's no surprise that many objections to virtue argumentation have focused on the *ad hominem* fallacy, which is generally thought to be committed when one objects to an *arguer* rather than an *argument*. While a number of authors have pushed back against the idea that *every* instance of addressing the arguer's character is problematic, (Aberdein, 2014; Battaly, 2010; Boudry, Paglieri, & Pigliucci, 2015) the idea persists that the arguer and the argument are fully separable. For instance, Tracy Howell and Justine Kingsbury have argued that we cannot give a complete account of argumentation using virtue theory because virtue theory alone cannot give an account of good argument. Howell and Kingsbury say that "[w]hen we put forward an argument, we seek to rationally persuade others to accept our conclusion" (Howell & Kingsbury, 2013, 23). While they accept some of the restrictions on *ad hominem* that have previously been argued for, for instance the legitimate criticism of a speaker who presents herself as an authority in a domain that she has no expertise, Howell and Kingsbury admit that such appeals can cast doubt on *conclusions* but argue that they cannot undermine arguments themselves because an argument is valid or inductively forceful independent of the arguer. Thus, virtue argumentation cannot account for the good argument in a satisfying way.

In a response to this objection, Aberdein (2014) distinguishes several different possible forms of *ad hominem*. According to Aberdein, Howell and Kingsbury have argued that *ad hominem* can legitimately be used as a rebuttal, but cannot undercut an argument as is required by the virtue theorist. The distinction, according to Aberdein, is that a rebutting *ad hominem* refutes the conclusion, while an undercutting *ad hominem* requires that, on the basis of character, the speaker has not shown that the conclusion follows from the premises. Crucially, Aberdein points out that while his critics object that a valid argument cannot become invalid because of their character, natural language arguments are rarely presented in standard form. It is true that the *form* of the argument cannot be undercut by character, but how we reconstruct the argument *can*. It is the ascription of form of the argument and the meaning of its terms we are considering when reconstructing arguments, and that is where charity is relevant, as Aberdein makes clear.

Here again, the Trump example becomes relevant. Rather than working to make Trump's natural language argument valid, we could instead start with a rough account of his apparent argument: "The best way to manage Syrian refugees is in the safest, most responsible, and

most humanitarian way possible, so we ought to help them resettle close to Syria and return.” Obviously, there is much left that could be filled in on this account, but before we do so we ought to start with considerations taken from rhetoric and virtue argumentation to see how we should go about reconstructing the argument. First, we might ask who is part of the conversation: most directly it is the members of the UN General Assembly, but such events are televised, printed, etc. and also connect to a variety of political or educational audiences. At the same time, we can think about the speaker and what we know about his character such as the extent to which he’s genuinely willing to engage in debate, consider issues, change his mind, seek out relevant evidence (such as the perspective of Syrian refugees themselves) to understand if we ought to go further in developing a charitable account of the argument. Given some obvious concerns – no refugee voices being directly represented in this argument and the speaker having an empirically verifiable racist history that is relevant to his argumentative situation, coupled with his documented propensity to lie and mislead, it is clearly a waste of time and a possible danger to charitably reconstruct this argument.

Given the vices we can reasonably attribute to Trump discussed above, it’s much more plausible to accept the dog-whistled version of the argument than the charitable one. In addition, we’ll benefit by avoiding the problematic frames that Stanley outlined. It’s tempting to think that Trump’s moral failings undercut his arguments, but on a virtue argumentation it’s Trump’s *argumentational* vice that does the undercutting. His argumentative vices such as unwillingness to listen to others or change his position, certainly seems to be influenced by moral failings – his lack of regard for truth or open-mindedness seems driven by racism and cruelty, which suggests certain moral failings may be tightly connected to argumentative vices such as unwillingness to change one’s mind or consider relevant perspectives.

5. CONCLUSION

I have focused in this paper on how to analyze natural language arguments that use dog whistles—a particularly complicated form of deception. I argued that in order to address such deception when reconstructing arguments, we ought to eschew the application of broad principles and instead use rhetorical knowledge and virtue argumentation to reconstruct arguments more effectively. For instance, if someone who sees clear dog whistles in a speech like Trump’s encounters someone else who believes that Trump is a true humanitarian, those two interlocutors have the opportunity to engage in arguments of their own about what evidence they have for the speaker’s

character and how to proceed with reconstruction. By deliberating about the relevant characters of arguers and opening up a space to identify possible deceptions without going all the way to blanket suspicion, arguers have more pluralistic means to reconstruct arguments and also a framework that still provides normative guidance and constraints in how to effectively reconstruct arguments.

ACKNOWLEDGEMENTS: Thank you to the ECA 2019 conference organizers and participants for great discussion, and to John Komdat for his thoughtful feedback. Special thanks to generous readers from the University of Rochester's Writing, Speaking, and Argument Program including Katherine Schaefer and Rachel Lee.

REFERENCES

- Aberdein, A. (2014). In defence of virtue: The legitimacy of agent-based argument appraisal. *Informal Logic*, 34(1), 77. <https://doi.org/10.22329/il.v34i1.3938>
- Aberdein, A. (2016). The vices of argument. *Topoi*, 35(2), 413–422. <https://doi.org/10.1007/s11245-015-9346-z>
- Aberdein, A., Cohen, D. (2016). Introduction: Virtues and arguments. *Topoi*, 35(2), 339–343. <https://doi.org/10.1007/s11245-016-9366-3>
- Albertson, B. L. (2015). Dog-whistle politics: Multivocal communication and religious appeals. *Political Behavior*, 37(1), 3–26. <https://doi.org/10.1007/s11109-013-9265-x>
- Battaly, H. (2010). Attacking character: Ad hominem argument and virtue epistemology. *Informal Logic*, 30(4). <https://doi.org/10.22329/il.v30i4.2964>
- Boudry, M., Paglieri, F., & Pigliucci, M. (2015). The fake, the flimsy, and the fallacious: Demarcating arguments in real life. *Argumentation*, 29(4), 431–456. <http://dx.doi.org.ezp.lib.rochester.edu/10.1007/s10503-015-9359-1>
- Bowell, T., & Kingsbury, J. (2013). Virtue and argument: Taking character into account. *Informal Logic*, 33(1), 22–32.
- Feldman, R. (1999). *Reason and Argument* (2nd ed.). Upper Saddle River, N.J: Prentice Hall.
- Frankfurt, H. G. (2005). *On Bullshit*. Princeton, N.J: Princeton University Press.
- Haney-López, I. (2014). *Dog Whistle Politics: How Coded Racial Appeals Have Reinvented Racism and Wrecked the Middle Class* (Reprint edition). Oxford: Oxford University Press.
- Henderson, R., & McCready, E. (2018). How dogwhistles work. In S. Arai, K. Kojima, K. Mineshima, D. Bekki, K. Satoh, & Y. Ohta (Eds.), *New Frontiers in Artificial Intelligence* (pp. 231–240). Springer International Publishing.

- Jørgensen, C. (2007). The relevance of intention in argument evaluation. *Argumentation*, 21(2), 165–174. <https://doi.org/10.1007/s10503-007-9044-0>
- Khoo, J. (2017). Code words in political discourse. *Philosophical Topics*, 45(2), 33–64. Retrieved from JSTOR.
- Lumer, C. (2005). Introduction: The epistemological approach to argumentation: A map. *Informal Logic*, 25(3). <https://doi.org/10.22329/il.v25i3.1134>
- Remarks by President Trump to the 72nd Session of the United Nations General Assembly. (September 19, 2017). Retrieved June 5, 2019, from The White House website: <https://www.whitehouse.gov/briefings-statements/remarks-president-trump-72nd-session-united-nations-general-assembly/>
- Stanley, J. (2015). *How Propaganda Works*. Princeton, New Jersey: Princeton University Press.
- White, I. K. (2007). When race matters and when it doesn't: Racial group differences in response to racial cues. *American Political Science Review*, 101(2), 339–354. <https://doi.org/10.1017/S0003055407070177>

The Rhetorical Potential of Metaphor

An experimental study on the effect of metaphors in argumentation

ROOSMARYN PILGRAM

Leiden University & University of Amsterdam
r.pilgram@hum.leidenuniv.nl & r.pilgram@uva.nl

Metaphors are deemed to have rhetorical potential. In what way and to what extent they affect argumentation is, however, not entirely clear. How exactly does metaphor presence influence argument evaluation? To answer this question, an experiment was conducted in which respondents had to evaluate dialogue fragments in which novel, direct metaphors were used to present a premise of the argumentation. The results show that metaphor presence negatively affects the reasonableness evaluation of argumentation

KEYWORDS: argumentation, experiment, metaphors, reasonableness evaluation.

1. INTRODUCTION

Metaphors are said to possess rhetorical potential; through the comparison of a source and target domain, a metaphor could make a message more convincing. Various reasons for this supposed rhetorical potential have been given: metaphors provide relief or pleasure, enhance a speaker's *ethos*, reduce counter-arguing, increase cognitive elaboration, induce associations in semantic memory and improve comprehension (see Charteris-Black, 2011; Oswald & Rihs, 2014; Van Stee et al., 2018; Thibodeau, Hendricks & Boroditsky, 2017).

Sopory & Dillard (2006, pp. 408-409) show in their meta-analysis that the presence of a metaphor in persuasive messages positively affects a receiver's attitude. The researchers nevertheless advocate caution when drawing conclusions about the rhetorical potential of metaphors: the messages in their meta-analysis were highly fine-tuned and, even for these finely-tuned messages, almost 40% decreased – as opposed to increased – persuasiveness. Sopory & Dillard (2006, p. 413) suggest that familiarity with the metaphor's target domain, the metaphor's novelty and its extendedness positively impact

its effectiveness, but also call for future researchers to “investigate other variables that have the potential to moderate the effectiveness of metaphorical messages”.

From the field of argumentation theory, at least one variable that impacts the persuasiveness of argumentation is known: the soundness of the argumentation (see O’Keefe, 2005; Meuffels, 2006). As O’Keefe (2005, p. 220) puts it: “normatively-good argumentative practices commonly engender persuasive success”. In the literature on metaphor’s rhetorical potential, the soundness of the argumentation put forward in the tested messages has so far not been given due consideration. In this contribution, the results of an experimental study on the use of metaphor in normatively good (“sound”) argumentation and argumentation that can be regarded as normatively bad (“fallacious”) will therefore be discussed. To what extent does the presence of a metaphor affect the evaluation of sound and fallacious argumentation?

Before delving into this study, first the research question will be refined by specifying the type of metaphor on which the study focuses (section 2). Subsequently, the organisation of the experiment will be outlined (section 3) and the results obtained in it will be presented (section 4). Lastly, the implications of these results will be discussed (section 5).

2. METAPHORS AND ARGUMENTATION

Metaphors can vary in novelty. Novel metaphors are those metaphors in which that which is meant about the target domain has not been (commonly) said before in terms of that which is said about the source domain. To understand novel metaphors therefore requires mapping of the source domain onto the target domain (Lakoff & Johnsen, 2003, p. 53). This is for example required when, in his speech at the annual *Conservative Political Action Conference*, US President Trump (2018) described immigrants in terms of a snake that, despite being cared for by a tender-hearted woman, bites this woman. Through his rather uncommon description, Trump encouraged the audience to connect the reference to the snake and its bad behaviour towards the woman (the source domain of animal care) to immigrants and the consequences of their admission to the US (the target domain of immigration).

In contrast to novel metaphors, conventional metaphors are used to such an extent that it is in principle not necessary anymore to map the source domain onto the target domain – in other words, these metaphors need not be processed as metaphors. Examples of conventional metaphors include expressions such as “*running out of time*” or “the *core* of the problem”, which are perfectly understandable

even without thinking of a certain type of physical activity or the centre of a particular kind of fruit. Indeed, Bowdle & Gentner (2005) stipulate in their Career of Metaphor theory that conventional metaphors are processed by means of categorisation; the metaphor's meaning is already in place in the recipient's mind and simply has to be retrieved. To encourage subjects to process metaphors as metaphors, or in Steen's (see 2008; 2010; 2017) terminology as "deliberate metaphors", the present study strived to focus on metaphors that were as novel as possible.

Additionally, the present study focused on direct metaphors. To explain why, it is helpful first to take a look at indirect metaphors. For indirect metaphors, there is a contrast between the contextual and basic meaning of the lexical unit (i.e., word or word combinations) used to express the metaphor (Steen, 2010, p. 54). The contextual meaning is the lexical unit's meaning when taking into account the co-text and situation in which it is used, while the basic meaning refers to the more concrete, precise, bodily action-related and/or older sense of the lexical unit (Pragglejaz Group 2007, p. 3). For example, when talking about "the war on cancer", the contextual meaning of "war" is "a determined and organized effort to control or stop something" (in this case cancer), while its basic meaning concerns "fighting between two or more countries that involves the use of armed forces and usually continues for a long time" (Macmillan Dictionary, 2018).

For direct metaphors, such a contrast between contextual and basic meaning of the lexical unit is not present. In fact, the contextual meaning of a direct metaphor is the same as its basic meaning. Such a metaphor is nevertheless still distinct from non-metaphorical language because it presents a contrast in meaning of the metaphor-related words (i.e., the words used to express the metaphor) at the level of the text or topic (Steen, 2010, p. 55). For example, in the utterance "The patient wasn't hooked to the drip feed!" that is used to explain why a phone is not charged, the words "patient", "hooked" and "drip feed" do not require any cross-domain mapping at the level of the lexical unit, as they activate the medical concepts directly referred to in the text. However, a new referent is introduced that is incongruous with the context: that of phone charging. Thus, understanding these metaphor-related words in the context of phone charging requires cross-domain mapping at the level of the topic.

In the present study, only direct metaphors are used. The reason for this is that direct metaphors enable the expression of an entire premise of the argumentation metaphorically, which has the advantage that we know with certainty that processing of the metaphor as a metaphor is essential for understanding and subsequently evaluating the argument.

3. ORGANISATION OF THE STUDY

3.1 Set-up

Based on the considerations presented in the previous section, the research question central to the present study can be refined as: to what extent does the presence of a (maximally) *novel, direct* metaphor affect the evaluation of sound and fallacious argumentation *in which this metaphor is used to convey a premise in the argumentation?*

To answer this question, a questionnaire was constructed in which respondents had to rate the perceived reasonableness of a discussion party's contribution to short dialogue fragments of informal discussions. This contribution consisted either of a pragma-dialectically sound argument, or an argument that could be considered as fallacious from a pragma-dialectical perspective (cf. Van Eemeren, Garssen & Meuffels, 2009). Furthermore, the premise in these sound or fallacious arguments was either expressed non-metaphorically or presented by means of a (maximally) novel, direct metaphor. This way, the effect of the presence of the metaphor on the perceived reasonableness of the argumentation could be contrasted with the controls consisting of literally expressed arguments.

In line with Sopory & Dillard's (2006, p. 413) suggestion to investigate variables that could moderate metaphors' rhetorical effects, respondents were asked to evaluate the comprehensibility and naturalness of the discussion contributions in addition to evaluating their reasonableness. The reason for this is that understanding metaphors (as reflected in their perceived comprehensibility) has been positively linked to metaphor appreciation, and the aptness of metaphors (as reflected in their perceived naturalness) is positively linked to metaphor comprehensibility (see Chiappe, Kennedy & Chiappe, 2003; Trick & Katz, 1986). It is therefore expected that the perceived comprehensibility and naturalness of metaphors might also affect the perceived reasonableness of their use in argumentation.

3.2 Materials

The experimental study combined a multiple-message design with a repeated-measures design: the questionnaire consisted of 36 short dialogue fragments, by means of which multiple instantiations of sound and fallacious arguments with and without metaphors were tested. In half of the items, the presented argument was pragma-dialectically sound (the other half being fallacious), and in one out of three items, the premise in the argument was conveyed by means of a (maximally)

novel, direct metaphor. Since the study aims to establish the effect of metaphor on the reasonableness evaluation of argumentation in general, different types of argumentation were systematically varied (i.e., pragmatic, causal and symptomatic argumentation). An overview of the item distribution can be found in Table 1.

It should be noted that the experiment combined a between and within subjects design. A between subjects design was used with respect to metaphor presence; respondents did not see the same items with and without metaphor. A within subjects design was used with respect to soundness of the argumentation; respondents saw several items with both sound and fallacious arguments.

	TYPE OF ARGUMENTATION	WITH METAPHOR	WITHOUT METAPHOR	TOTAL
SOUND	PRAGMATIC	2	2	12
	CAUSAL	2	2	
	SYMPTOMATIC	2	2	
FALLACIOUS	SLIPPERY SLOPE	2	2	12
	INCORRECT CAUSAL RELATION	2	2	
	HASTY GENERALISATION	2	2	
EXTRA CONTROLS	SOUND		6	12
	<i>AD HOMINEM</i> (DIRECT)		3	
	<i>AD HOMINEM</i> (TU QUOQUE)		3	
TOTAL		12	24	36

Table 1 – Distribution of the items in the questionnaire. The items “with metaphor” included a novel direct metaphor to present a premise in the argumentation. The items “without metaphor” did not include any metaphorical language.

The items themselves followed a fixed pattern (see Figure 1). At the start of each dialogue fragment, a brief description was provided of the discussion topic and the relationship between speakers A and B. To make sure that respondents evaluated the quality of B’s argumentation, a variety of discussion topics that are as uncontroversial as possible was included (e.g., coffee, work habits and exercising). Additionally, to

ensure that respondents regarded the dialogues as informal discussions, it was made clear that speakers A and B know each other (e.g., they are colleagues, family members or friends) and that the discussions were not conducted in specific institutionalised contexts.

Subsequent to the item's background information, a two-turn dialogue followed: speaker A puts forward a standpoint in turn 1 and speaker B contradicts A's standpoint and provides a sound or fallacious argument for his or her own standpoint in turn 2. If a metaphor is included in B's argumentation, the metaphor is (maximally) novel and direct. In half of all the items with metaphors, the metaphor was signalled as a metaphor by means of an indicator of analogy (e.g., "similarly", "it's just like", "equally").

-
- (i) *Background information on the topic of the discussion and the relationship between speakers A and B*
 A Puts forward a standpoint.
 B Contradicts A's standpoint and provides a sound or fallacious argument for his or her own standpoint, in which the premise is presented by means of a (maximally) novel, direct metaphor or without a metaphor.

 - (ii) *Two friends are discussing work.*
 A People should always try to stay in their job, no matter whether they like it or not.
 B I disagree; tigers in small cages will get ill as well.

 - (iii) *Two friends are discussing work.*
 A People should always try to stay in their job, no matter whether they like it or not.
 B I disagree; employees who stay in a job in which they aren't appreciated will get unhappy.
-

Figure 1 – A (i) schematic representation of the dialogue fragments in the questionnaire, a (ii) concrete example of such a fragment with sound pragmatic argumentation and metaphor, and the (iii) non-metaphorical counterpart of example (ii).

Each item was directly followed by the question "How comprehensible do you find B's contribution to the conversation?", which was in turn followed by similarly formulated questions about reasonableness and naturalness. Each of the questions had to be answered on a 7-point Likert scale, ranging from "very incomprehensible / unreasonable / unnatural" (= 1) to "very comprehensible / reasonable / natural" (= 7).

Thereby, the reasonableness scale corresponded with the scale used in Van Eemeren, Garssen & Meuffels (2009), and the comprehensibility and naturalness scales allowed for comparison between the scales.

To make the sound and fallacious argumentation in the questionnaire as comparable as possible, they each use the same argument schemes (i.e., pragmatic, causal and symptomatic argumentation, see Table 1). In the fallacious arguments, there is however an incorrect application of this argument scheme: the slippery slope fallacy can be regarded as an incorrect application of the negative variant of pragmatic argumentation, the fallacy of incorrect causal relation is, as the name suggests, an incorrect application of causal argumentation, and the fallacy of hasty generalisation amounts to an incorrect application of symptomatic argumentation (Van Eemeren & Grootendorst, 1992; 2004). Extra controls were included in the questionnaire by way of adding additional sound arguments as well as direct *ad hominem* fallacies and *tu quoque* fallacies (see Table 1).

The dialogue fragments underwent several rounds of revision before they were deemed suitable for inclusion in the study. Two colleagues at the University of Amsterdam (a metaphor researcher and an argumentation theorist) helped evaluating whether speaker B *contributes to the discussion in the dialogue fragments in the intended way* (by using a deliberate metaphor in the envisioned types of argumentation). Two native speakers of English checked the *correctness* and *naturalness* of the formulation used in the dialogues, as well as the *overall uniformity* of the dialogue fragments (does the conversation in the dialogues proceed in an identical manner?) and their *stylistic variety* (do the speakers use various kinds of formulations?). Attention was also paid to the *length of the dialogues*: the dialogues were kept as concise as possible to ensure that filling out the questionnaire would not take too much time (can the dialogue not be kept shorter?).

Before filling out the questionnaire, respondents received some information about the study. To refrain from giving away the purpose of the study, the terms “metaphor” and “argumentation” were carefully avoided. Instead, respondents were told “Depending on the specific topic and context of a conversation, people might have different ideas about the comprehensibility, reasonableness and naturalness of particular contributions to the conversation; what one might find perfectly fine, others might not. The purpose of this research is to determine which conditions influence people’s ideas about conversational contributions”.

It was furthermore emphasised that respondents may assume that the speakers in the questionnaire’s dialogue fragments always speak the truth, that these speakers are completely different people in each of the different dialogue fragments, that no special expertise is

required for filling out the questionnaire and that there are no right or wrong answers. Subsequently, it was stressed that respondents could stop participation at any moment or withdraw their data up to 8 days after completion, and that their data was processed anonymously. They were then provided with the contact details of the author and of the *Ethics Committee Faculty of Humanities* of the *Amsterdam Institute for Humanities Research*, and asked for their consent.¹

3.2 Respondents

A total of 408 native speakers of English were randomly selected via online research platform *Prolific Academic*. Only native speakers were asked to participate to minimise the influence of language proficiency problems. Other than that, no specific requirements were laid down for participation. *Prolific Academic* was used for practical reasons: since the study was designed in the Netherlands and needed be conducted in English, distributing the questionnaire via this online research platform allowed for obtaining sufficient responses from native English speakers. In compliance with *Prolific Academic's* policy, respondents were rewarded £1.50 for their participation.

Respondents were between 18 and 73 years old, with an average age of 33.9 (SD = 11.4). Somewhat more respondents were female (63.2%); 36.3% were male and 0.5% non-binary. A slight majority of the respondents was higher educated (54.3% was higher educated versus 45.6% lower educated).² It took the respondents between 15 and 20 minutes to complete the questionnaire. It should be noted that respondents were given maximally 45 minutes to complete the questionnaire in *Prolific Academic* to avoid responses that were, for some reason, interrupted. Any responses that were timed out or incomplete for other reasons were excluded from the analyses.

4. RESULTS

Now that organisation of the study has been outlined, we can take a look at the results. First, the effect of the presence of novel, direct metaphors on argumentation on the perceived reasonableness scores of sound and fallacious arguments will be determined. Second, possible relations

¹ Approval for the experiment was provided by the *Ethics Committee Faculty of Humanities* on 6 July 2018 (dossier 2018 -52).

² Respondents were considered to be 'higher educated' if their highest completed level of education consisted of a university Bachelor's degree, a university Master's degree, or a PhD or equivalent. They were considered to be 'lower educated' if their highest completed educational level was elementary school, high school or vocational school or equivalent.

between the comprehensibility and naturalness scores on the one hand, and the reasonableness scores on the other hand, will be checked.

4.1 The effect of metaphor presence

Before determining the effect of metaphor presence and argumentative soundness on reasonableness evaluation, we should first take a look at the reasonableness evaluations of the control items, as they provide the benchmarks against which the results can be interpreted.

Table 2 shows that the respondents gave the sound control items (in which no metaphors were used) on average a score of 5.44 (SD = 1.45), meaning that they regarded these arguments to be between “fairly reasonable” and “reasonable”. As expected, this is not significantly different from the reasonableness scores of the sound items of interest without metaphor (which obtained an average score of 5.59, SD = 1.53) as presented in Table 3a. It should also be noted that these scores are similar to the average reasonableness scores that Van Eemeren, Garssen & Meuffels (2009) report in their research³, which indicates the robustness of these findings.

	REASONABLENESS SCORE WITHOUT METAPHOR
SOUND	5.44 (1.45)
FALLACIOUS (<i>AD HOMINEM</i> , DIRECT)	4.00 (1.70)
FALLACIOUS (<i>AD HOMINEM</i> , <i>TU QUOQUE</i>)	5.37 (1.54)

Table 2 – Mean reasonableness scores of the sound and fallacious control items without metaphorical language (scale values are means with SD between parentheses)

Since the scores for the controls provide us with the means to interpret the other reasonableness scores, we can now take a look at the main results of interests. As Table 3a shows, the mean reasonableness scores of the items in which a premise in the argumentation is presented by means of a metaphor is structurally lower for both the sound and fallacious items than these scores are for items in which the premise is presented directly, without the use of a metaphor. Table 3b indicates

³ Van Eemeren, Garssen & Meuffels (2009), for example, report mean scores of 5.27 (SD = 0.60) and 5.32 (SD = 0.60) for the sound items in their research on the *argumentum ad consequentiam*, and a score of 5.31 (SD = 0.66) in their research on the slippery slope fallacy.

that this structurally lower score is independent of the type of argumentation used.

	REASONABLENESS SCORE WITH METAPHOR	REASONABLENESS SCORE WITHOUT METAPHOR
SOUND	3.58 (1.89)	5.59 (1.53)
FALLACIOUS	2.96 (1.67)	4.46 (1.71)

Table 3a – Mean reasonableness scores of the sound and fallacious items of interest with and without novel, direct metaphors as premises (scale values are means with SD between parentheses)

	TYPE OF ARGUMENTATION	REASONABLENESS SCORE WITH METAPHOR	REASONABLENESS SCORE WITH METAPHOR
SOUND	PRAGMATIC	4.69 (1.94)	5.59 (1.61)
	CAUSAL	3.52 (1.87)	5.88 (1.24)
	SYMPTOMATIC	3.52 (1.83)	5.30 (1.66)
FALLACIOUS	SLIPPERY SLOPE	2.60 (1.55)	4.15 (1.76)
	INCORRECT CAUSAL RELATION	3.53 (1.71)	4.77 (1.61)
	HASTY GENERALISATION	2.76 (1.57)	4.44 (1.68)

Table 3b – Mean reasonableness scores of the sound and fallacious items of interest with and without novel, direct metaphors as premises specified for each argument type (scale values are means with SD between parentheses)

To determine whether the difference in the reasonableness scores of arguments with and without metaphor in Table 3a is significant, a mixed model analysis was conducted. Overall, there seems to be a significant effect of metaphor presence ($F(1, 9374) = 2440.48$; $p < .001$): respondents evaluated sound and fallacious argumentation combined to be significantly more reasonable if it did not include any metaphorical language (on average, they evaluated this argumentation with a score of 5.02 (SD = 1.72)) than if a novel, direct metaphor was used to present a premise in this argumentation (which they gave an average score of 3.27 (SD = 1.80)).

When taking the soundness of the argumentation into account, there is, as expected, a significant difference between the reasonableness evaluation of the sound and the fallacious

argumentation, irrespective of metaphor presence ($F(1, 9374) = 1490.28$; $p < .001$): when the combined group of sound arguments with and without metaphors are contrasted with the combined group of fallacious arguments with and without metaphors, the sound arguments were evaluated as significantly more reasonable (on average with a score of 4.58 ($SD = 1.99$)) than the fallacious ones (which received an average score of 3.71 ($SD = 1.84$)).

Interestingly, the analysis shows that there is also a significant interaction between soundness and metaphor presence ($F(1, 9374) = 111.71$; $p < .001$), meaning that metaphor presence obscures the effect that the soundness of the argumentation has on perceived reasonableness. In other words, even though respondents find fallacious argumentation less reasonable than sound argumentation in argumentation with and without metaphors, the extent to which they find fallacious argumentation less reasonable is smaller when a novel, direct metaphor is used in the argumentation than when no metaphors are used at all.

It was additionally checked whether these results hold equally for the different types of argumentation (i.e., for the argument schemes of pragmatic, causal and symptomatic argumentation, see *table 3b*). Rather unexpectedly, they did not ($F(2, 6098) = 110.75$; $p < .001$). There seems to be a significant interaction between argument type, metaphor presence and soundness ($F(2, 6337) = 3.51$; $p = .03$). A pairwise comparison of means shows that only for causal and pragmatic sound argumentation there are no significant differences in reasonableness evaluation, meaning that the interaction is predominantly due to differences in evaluation of the fallacious counterparts of these schemes. Apart from argument type, no other extraneous variables seem to have impacted the study.

4.2 The effect of comprehensibility and naturalness

In addition to reasonableness, respondents were also asked to evaluate the comprehensibility and naturalness of the dialogue fragments. The idea behind this is that, especially when respondents had to evaluate the reasonableness of arguments in which a premise is presented by means of a novel, direct metaphor, the more comprehensible and natural they find this argument, the more reasonable they would find it.⁴

Table 4a shows that there is indeed a significant positive correlation between the overall comprehensibility, reasonableness and

⁴ For this reason, only the items that respondents saw with metaphors were included in the correlation analysis; the extra sound controls and the *ad hominem*-fallacies were left out of the equation.

naturalness scores of the argumentation in the items of interest: the more comprehensible the argumentation was to the respondents, the more reasonable and natural they found it, and the more natural they judged the argumentation to be, the more reasonable and comprehensible they found it as well. A partial correlation in which metaphor presence is regarded as a covariate (see Table 4b) shows that these correlations are, in fact, independent of metaphor presence. Thus, comprehensibility and naturalness of the argumentation itself already seems to be a factor in the reasonableness evaluation of argumentation, irrespective of the use of metaphors in this argumentation.

	1	2	3
1. COMPREHENSIBILITY	-		
2. REASONABLENESS	.68**	-	
3. NATURALNESS	.72**	.78**	-

Table 4a – Pearson correlations among comprehensibility, reasonableness and naturalness scores. ** Correlation is significant at the 0.01 level (2-tailed).

	1	2	3
1. COMPREHENSIBILITY	-		
2. REASONABLENESS	.61**	-	
3. NATURALNESS	.66**	.73**	-

Table 4b – Partial correlations among comprehensibility, reasonableness and naturalness scores with metaphor presence as a covariate. ** Correlation is significant at the 0.01 level (2-tailed).

5. CONCLUSION

So, to what extent does the presence of a novel, direct metaphor affect the evaluation of sound and fallacious argumentation in which this metaphor is used to convey a premise in the argumentation? This study indicates that the presence of such a metaphor affects the reasonableness evaluation of sound and fallacious argumentation, but not in the positive way that was suggested in the extant literature on metaphors in general: rather, the presence of a novel, direct metaphor

that is used as a premise in an argument negatively impacts the reasonableness evaluation of both sound and fallacious argumentation. What is more, our analysis shows that the extent to which this negative impact arises depends on an interaction between soundness and metaphor presence: a fallacious argument is less affected by metaphor presence than a sound argument

This finding seems to go against the idea that metaphors might reduce counter-arguing or improve comprehension, as mentioned in the argumentation literature. It also challenges the idea that metaphor presence, in general, positively affects reasonableness evaluation, which can be formed based on studies in the field of communication, and the acclaimed positive correlation between reasonableness and effectiveness in empirical research in the field of argumentation. Although the results might therefore be regarded as surprising, it might not be so surprising after all. As mentioned in the introduction of this paper, meta-analyses on the persuasiveness of metaphor already showed that the effects of metaphor presence on argumentative discourse are varied: sometimes metaphor presence yields a more positive evaluation, sometimes a more negative one. This could be due to the lack of controlling the soundness of the argumentation, the novelty of the metaphors, and the metaphor's argumentative use, as well as failing to take into account the comprehensibility and naturalness of metaphors. Indeed, this was the reason why this study only focussed on a particular type of metaphor with a particular kind of argumentative use in the present study: novel, direct metaphors that were used to express a premise in the argumentation. It was also the reason why respondents were asked how comprehensible and natural they found the metaphorically expressed argumentation. This study now seems to have found one of the reasons why the meta-analyses gave such mixed results: neither the particular use and type of metaphor was typically taken into account, nor its overall perception in the discourse itself.

While this does explain why we obtained different results from those in previous research, it does not explain the interaction that was found in the present study between soundness and metaphor presence. We found that the negative impact that metaphor presence has on the reasonableness evaluation of sound argumentation was greater than on this evaluation of fallacious argumentation. This could perhaps be explained by the idea that metaphors increase cognitive elaboration and therefore leave less room to critically evaluate the message that they convey, which has been put forward in the literature. Yet, whether this is indeed the case should be explored in more detail in future research.

ACKNOWLEDGEMENTS: I hereby thank Gerard Steen and Jean Wagemans for their helpful comments on the experiment and earlier versions of this contribution. My gratitude extends additionally to Francisca Snoeck Henkemans for her help with the formulation of the items and to Christian Burgers for his help with the statistical analysis of the data.

REFERENCES

- Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review* 112(1), 193–216. doi:10.1037/0033-295X.112.1.193.
- Charteris-Black, J. (2011). *Politicians and rhetoric: The persuasive power of metaphor*. Palgrave Macmillan: Basingstoke, UK.
- Chiappe, D.L., Kennedy, J.M., & Chiappe, P. (2003) Aptness is more important than comprehensibility in preference for metaphors and similes. *Poetics* 31, 51-68.
- Eemeren, F.H. van & Grootendorst, R. (1992). *Argumentation, communication, and fallacies: A pragma-dialectical perspective*. Hillsdale: Lawrence Erlbaum Associates.
- Eemeren, F.H. van & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge: Cambridge University Press.
- Eemeren, F.H. van, Garssen, B. & Meuffels, B. (2009). *Fallacies as judgments of reasonableness: Empirical research concerning the pragma-dialectical discussion rules*. Dordrecht: Springer.
- Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. London: The University of Chicago Press.
- Macmillan Dictionary (2018). Accessed from <https://www.macmillandictionary.com/> on 21 October 2019.
- Meuffels, B. (2006). De vrijheidsregel: Convergente operaties in empirisch ad hominem-onderzoek. [The freedom rule: Convergent operations in empirical ad hominem-studies] In: Garssen, B. & Snoeck Henkemans, F. (Eds.), *De redelijkheid zelve: Tien pragma-dialectische opstellen voor Frans van Eemeren* [In all reasonableness: Ten pragma-dialectical essays for Frans van Eemeren]. Amsterdam: Rozenberg, pp. 11-22.
- O’Keefe, D. (2005). News for argumentation from persuasion effects research: Two cheers for reasoned discourse. In: Willard, C.A. (Ed.), *Critical problems in argumentation: Selected papers from the 13th biennial conference on argumentation sponsored by the American Forensics Association and National Communication Association August, 2003*. Washington, DC: National Communication Association.
- Oswald S. & A. Rihs (2014). Metaphor as argument: Rhetorical and epistemic advantages of extended metaphors. *Argumentation* 28(2), 133-159.
- Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse’, *Metaphor and Symbol* 22(1), 1–39.

- Sopory, P. & J.P. Dillard (2006). The persuasive effects of metaphor: A meta-analysis. *Human Communication Research* 28(3), 382-419. Doi:10.1111/j.1468-2958.2002.tb00813.x
- Stee, S. van, S.M. Noar, N. Grant Harrington & L.F. Grant (2018). The effect of metaphor use and message format on cognitive processing and persuasive outcomes of condom promotion messages. *Communication Studies* 69 (1), 23-41. Doi.: 10.1080/10510974.2017.1396543.
- Steen, G.J. (2008). The paradox of metaphor: Why we need a three-dimensional model of metaphor. *Metaphor & Symbol* 23, 213-241.
- Steen, G. J. (2010). When is metaphor deliberate? In N.L. Johannesson, D. C. Minugh & C. Alm-Arvius (Eds.), *Selected papers from the 2009 Stockholm Metaphor Festival* (pp. 43-63). Stockholm: Department of English, Stockholm University.
- Steen, G.J. (2017). Deliberate metaphor theory: Basic assumptions, core tenets, remaining issues. *Intercultural Pragmatics* 14(1), 1-24.
- Thibodeau, P.H., R.K. Hendricks & L. Boroditsky (2017). How linguistic metaphor scaffolds reasoning. *Trends in Cognitive Sciences* 21(11), 852-863. Doi: 10.1016/j.tics.2017.07.001.
- Trick, L., & Katz, A.N. (1986). The domain interaction approach to metaphor processing: Relating individual differences and metaphor characteristics. *Metaphor and Symbolic Activity*, 1(3), 185-213.
- Trump (2018). Conservative Political Action Conference (CPAC) in National Harbor, Maryland, 23 February 2018. Accessed from <https://www.whitehouse.gov/briefings-statements/remarks-president-trump-conservative-political-action-conference-2/> on 21 October 2019.

Separate opinions as argumentative activity type

H. JOSÉ PLUG
University of Amsterdam
h.j.plug@uva.nl

National and international systems of law differ in as far as they allow separate opinions to be published. In the Netherlands, for example, collegial courts speak with one voice. In the European Court of Human Rights, however, court members who disagree with the majority of the court may express their divergent views in a separate opinion. In this paper I will investigate institutional constraints that may affect the argumentation brought forward in separate opinions and I will set about defining separate opinions as a distinct argumentative activity type.

KEYWORDS: argumentative activity type, concurring opinion, dissenting opinion, ECHR, legal argumentation, separate opinion, strategic manoeuvring

1. INTRODUCTION

In many countries judicial decisions taken by a court's judge panel may be accompanied by separate opinions. These separate opinions, written by one or more of the judges forming a minority, serve to comment on the final decision reached by the majority of the court's judge panel. These separate opinions of the minority offer the possibility to provide insight in the considerations on which judges disagree with the decision taken by the majority or with the arguments on which the decision is based. Traditionally separate opinions are allowed in most common-law countries whereas in most civil-law countries the publication of these opinions is not permitted.

However, studies by Raffaelli on behalf of the European Parliaments Committee on Legal Affairs (2012) and by the European Commission for Democracy through Law (2018) signal that among the Member States of the EU there is a distinctive trend to allow separate opinions in civil-law countries as well as in common-law countries. The studies note that only a small minority of the Member States of the EU,

including the Netherlands, have no provision on separate opinions or even explicitly forbid separate opinions. Similar disparities in provisions and practices exist between the European Court of Human Rights which does, and the European Court of Justice which does not allow separate opinions international courts.

These different practices in national and international courts together with the increasing trend towards allowing separate opinions have triggered a discussion on the desirability of separate opinions. One of the arguments presented by the proponents is that separate opinions promote transparency and improve the dialogue with future and lower courts. This raises the question of how the argumentation in separate opinions may contribute to realizing these goals. Insights and concepts from argumentation theory may contribute to answering this question by characterizing separate opinions as an argumentative activity type and determine how institutional constraints may affect the argumentation that takes place in this activity type.

In this paper I will focus on separate opinions in the European Court of Human Rights because the rulings of this court are to be respected by all EU countries, which makes separate opinions also relevant for EU countries that do not allow separate opinions on a national level. In section 2 the stage model of argumentative discussions is used to clarify how the argumentative activity that takes place in separate opinions relates to that in legal proceedings. Section 3 of this paper is concerned with the rules and regulations that govern the publication of different types of separate opinions in the European Court of Human Rights and with the reconstruction of the standpoints adopted in these dissenting and concurring opinions. In section 4 an attempt is made to specify characteristics of a separate opinion as an argumentative activity type and to set out some techniques a judge may use to strategically present his or her criticism on the majority decision in a separate opinion.

2. SEPARATE OPINIONS AS PART OF A DISCUSSION

A model that may serve as a starting point to study argumentation brought forward in judicial decisions and separate opinions, is a model that has been developed in the pragma-dialectical argumentation theory. This general model of critical discussion was introduced by Van Eemeren and Grootendorst (1984) and has been applied to the context of legal discussion by Feteris (1999), Jansen (2003), Kloosterhuis (2002) and Plug (2000). In order to render the general model suitable for the reconstruction of legal argumentation, specific characteristics of legal procedures should be taken into account. The general model presents an overview of the four discussion stages that must be

completed in order to further the resolution of a dispute. These stages are referred to as the confrontation stage, the opening stage, the argumentation stage and the concluding stage, respectively. When modeling legal procedures (Feteris, 2017), the confrontation stage concerns the scope and the content of the dispute: it is to be established what the difference of opinion exactly amounts to in legal terms, within the context of the law. In the second stage, the opening stage, the common legal starting points are established in view of the legal procedural rules (the procedure of the discussion) as well as regarding rules of material law (the rights and obligations of citizens and institutions). In the argumentation stage, the parties to the process defend their standpoints in accordance with the rules of procedure and provide evidence when required to do so. The court evaluates the quality of the argumentation and the evidence. In the final stage of a legal procedure, the concluding stage, the court determines whether the claim of the party who initiated the proceedings can be maintained or should be rejected. By means of a justification of the final decision, the court accounts for the way in which its discretionary power was used to apply and interpret the law.

The concluding stage of a legal procedure may give rise to a new discussion. If this new discussion is initiated by the parties to the process, it takes the form of an appeals procedure. In addition, in legal systems in which separate opinions are allowed, a new discussion may be started by one or more of the judges who took part in the deliberation of the court. These newly initiated discussions may also be reconstructed applying the four-stage model of critical discussion. Within the framework of pragma-dialectics, it is assumed that participants in both discussions maneuver strategically in every stage of the procedure in order to resolve their differences of opinion. 'Strategic maneuvering' refers to the efforts discussants, in this case parties to the process and court members, make to reconcile rhetorical effectiveness with the maintenance of dialectical standards of reasonableness. In the following section I will analyze the differences of opinion that may be established in the confrontation stage of separate opinions of the European Court of Human Rights.

3. DISPUTES IN SEPARATE OPINIONS OF THE EUROPEAN COURT OF HUMAN RIGHTS

The European Court of Human Rights rules on applications filed by states or by individuals raising matters concerning violations of the civil and political rights as set out in the European Convention on Human Rights. Applications to the court that are considered admissible, as a rule, are heard by a Chamber that is composed of seven judges. When a

case raises a serious question of interpretation or involves an issue of general importance, it may be referred to the seventeen-member Grand Chamber. The Parliamentary Assembly of the Council of Europe elects all judges from lists of candidates proposed by each state. Although these judges have been proposed for election by individual states, they hear cases as individuals and do not represent any state.¹ The decisions of the court are taken by a majority of the judges present (Rule 23 of the Rules of Court). The court's ruling consists of both the judges' considerations as well as their individual votes.

The European Court of Human Rights allows judges to publish separate opinions. Article 45 (2) of the European Convention of Human Rights reads: "If a judgment does not represent, in whole or in part, the unanimous opinion of the judges, any judge shall be entitled to deliver a separate opinion." Rule 74 (2) of the Rules of Court then specifies this provision: 'Any judge who has taken part in the consideration of the case shall be entitled to annex to the judgment either a separate opinion, concurring with or dissenting from that judgment, or a bare statement of dissent.' A dissenting opinion contains the reasons for which one or more of the judges taking part in the deliberations voted against the final decision reached by the majority. A concurring opinion is written by one or more of the judges forming part of the majority and serves to provide for different, or additional legal arguments to support the conclusion. The Rules of Court provide that a separate opinion does not need to be justified. However, since this present paper focuses on the argumentation in separate opinions, I will not discuss bare statements.

3.1 Standpoints in different types of separate opinions

In legal practices such as the European Court of Human Rights we may distinguish different types of separate opinions. The propositions from which the judge takes a stance can be divided in two categories. We may specify these categories along the lines of pragma-dialectics (1992).

The first category, dissenting opinions, concerns a standpoint adopted with respect to a proposition that refers to the decision (d) of the majority of the court. This standpoint, if it amounts to 'the decision (d) should be evaluated negatively (-/d)', results in a mixed difference of opinion because both positions relate to the same proposition (d). This mixed difference of opinion between the minority and the majority of the court may be reconstructed as follows: Minority/judge (-/d); Majority (+/d). However, if the standpoint of the minority amounts to 'the decision (d) by the court should be different (+/d)', the difference of

¹ See Bruinsma (2008) on the question whether or not judges in the ECHR (Grand Chamber) may be suspected of partiality or chauvinism.

opinion should be considered to be non-mixed because the minority adopts a standpoint with respect to a different, newly introduced, proposition (d'). This difference of opinion may be reconstructed as: Minority/judge (+/d'); Majority (?/(+/d')).

Dissenting opinion (joint): majority +/d; minority -/d

An example of a standard dissenting opinion is the case *Carson and Others v The United Kingdom*. In this case, the European Court of Human Rights rejected a claim that UK pensioners who had earned pensions by working in Britain but had been living abroad, should have their pensions raised in line with UK inflation. The pensioners claimed discrimination (article 14 of the Convention and article 1 of the First Protocol) because their treatment was in contrast with the position of pensioners who had remained resident within the United Kingdom. The court held by eleven votes to six that there has been no violation of Article 14 of the Convention taken in conjunction with Article 1 of Protocol No. 1. In their (joint) separate opinion, the six dissenters presented their standpoint as follows.

We are unable to find that there has been no violation of Article 14 of the Convention taken in conjunction with Article 1 of Protocol No. 1. (...).

(*Carson and others v. The United Kingdom*, no. 42184/05, 16 March 2010)

In this case positive standpoint of the majority concerns the proposition 'there has been no violation of (...)' (+/d). Since this positive standpoint has been countered with a negative standpoint with respect to the same proposition (-/d), the difference of opinion may be characterized as mixed.

The second category concerns a standpoint adopted with respect to a proposition regarding the argumentation (a) underlying the decision by the majority of the court. This standpoint results in a mixed difference of opinion between the minority and the majority of the court, in case the minority counters the positive standpoint of the majority with the negative standpoint: 'the argumentation (a) should be evaluated negatively' (-/a). The difference of opinion may be reconstructed as follows: Minority/judge (-/a); Majority (+/a). A negative standpoint regarding the argumentation underlying the decision by the court usually functions as an (indirect) attack (-/d) on the decision by the majority of the court. If the decision by the court is disputed via the

argumentation, the separate opinion should be considered as a dissenting opinion.

The minority's standpoint on the argumentation (a) in the decision may also concern a different, newly introduced, proposition (a') and could then be formulated as: The argumentation (a) for the decision by the court could (should, would) be different (+/a'). This standpoint may imply that there is not necessarily a difference of opinion on the decision by the majority. The separate opinion can therefore be considered a concurring opinion.

Concurring opinion (joint): minority +/a'; majority ?/(+/a')

An example of a standard joint concurring decision is the case *Karatas v. Turkey*. In this case the applicant, Mr Karatas, had published poems that the Turkish national court found to be of a terrorist nature because they incited separatist movements. Mr Karatas was convicted and claimed a violation of Article 10 of the Convention (the right to freedom of expression). The majority of the court (twelve votes to five) agreed with Mr Karatas's point of view that his conviction amounted to a violation of Article 10 because the poems did not encourage the use of violence. There were six judges who dissented and six judges who attached concurring opinions. In a joint concurring opinion, five judges presented their collective standpoint as follows.

We share the Court's conclusion that there has been a violation of Article 10 in the present case although we have reached the same result by a route which employs the more contextual approach (...).

(Karatas v. Turkey, no. 23168/94, 8 July 1999)

In this case the judges agree with the decision of the majority of the court (+/d). However, according to the judges, the argumentation underlying the decision should be different: +/a'. Since it may be assumed that this standpoint may be questioned by the majority, ?/(+/a'), the difference of opinion may be regarded as non-mixed. If, in their justification, the court already would have discussed and rejected a', then the difference of opinion would have been mixed.

The differences of opinion may be more complex if the decision or the argumentation that is questioned relates to two or more propositions: d1, d2, dn or a1, a2, an. In case of propositional complexity, the difference of opinion in the confrontation stage of a separate opinion may relate to one or more of these propositions. Separate opinions that discuss a selection of the propositions are referred to as 'partly dissenting opinions' or 'partly concurring

opinions'. An overview of the different implicit or explicit standpoints that are defended in the different types of separate opinions is presented in table 1.

Dissenting opinion:	<ul style="list-style-type: none"> - The decision (d) by the court should be evaluated negatively (-/d) - The decision (d) by the court should be different (+/d') (because) the argumentation (a) for the decision (d) by the court should be evaluated negatively (-/a)
Partly dissenting opinion:	<ul style="list-style-type: none"> - Part of the decision (d1) by the court should be evaluated negatively (-/d1) - Part of the decision (d1) by the court should be different (+/d1') (because) part of the argumentation (a1) for the decision (d) by the court should be evaluated negatively (-/a1)
Concurring opinion:	<ul style="list-style-type: none"> -The argumentation (a) for the decision by the court should/would / could be different (+/a')
Partly concurring opinion:	<ul style="list-style-type: none"> - Part of the argumentation (a1) for the decision by the court should/ could/ would be different (+/a1')

Table 1. Reconstruction of standpoints in different types of separate opinions

4. STRATEGIC MANEUVERING IN SEPARATE OPINIONS

The way in which judges present the justification of their standpoints in separate opinions can be analyzed from the perspective of the strategic maneuvering. Such an analysis clarifies how judges make an expedient choice from the options that constitute the starting points of separate opinions in the context of a legal discussion.

In order to answer the question how strategic manoeuvres in separate opinions compare to strategic manoeuvres in judicial decisions, it is necessary to ascertain the characteristics of the argumentative activity type of separate opinions in contrast to those of the argumentative activity type of judicial decisions. Relevant to the characterization of both activity types is the institutional context in which the judges' decision-making task takes place.

The institutional point of a legal procedure, as prescribed in the law, is that the procedure results in a justified, binding decision by the

judge(s) to whom the legal dispute is submitted. In the judicial decision, the judge is obliged to state the grounds for his decision. Although the judge is not obliged to deal with each argument that was raised by the parties to the proceedings, they should be enabled to ascertain how and to what extent the facts and legal foundations, as presented by them, have been taken into consideration. Moreover, the justification should enable the public at large to monitor the administration of justice as well as gain insight into its proceedings.

As regards the process of drawing up a judicial decision, it is considered to be vitally important in a democracy that individual judges and the judiciary as a whole are impartial and independent of all external pressures and of each other: parties to the process as well as the wider public should have confidence that legal decisions are taken fairly and in accordance with the law. This principle of judicial independence is embedded in international and European codes and differentiates between external and internal independence. The aim of external independence, as laid down in Article 6 of the Convention of Human Rights, is to guarantee every person the fundamental right to have their case decided in a fair trial, on legal grounds only and without any improper influence. The aim of internal independence, as laid down in recommendation 5 and 22 of the Committee of Ministers, is that 'In their decision making judges should be independent and be able to act without any restriction.' Judges should have 'unfettered freedom to decide cases impartially, in accordance with the law and their interpretation of the facts.'

The possibility of issuing a separate opinion can be seen as one that safeguards the judges' internal independence (Anand, 1965, p. 801). The opportunity of allowing individual members of the court to openly challenge the decision and the argumentation of the majority of the court, enables the judiciary to meet the demands that follow from the principle of independence. The institutional point of issuing a separate opinion may, therefore, be considered as a means to publicly ensure the judiciary's independence.² It is precisely because each individual court member should be able to independently account for his or her point of view, that institutional rules on separate opinions are less strict than those on judicial decisions.

The differences between the institutional point of binding decisions and separate decisions result in different options for the majority and the minority to manoeuvre strategically with regard to the following three aspects: audience demand, topical choice and presentational devices. In the following I will give short examples of

² See, for example, Laffranque (2003) and Raffaelli (2013) for an overview of arguments for and against allowing separate opinions.

how the aspect of audience demand and the use of presentational devices may become apparent in the strategic manoeuvring in separate opinions.

4.1 Audience demand

A judicial decision and its justification are addressed primarily to the parties to the process, but are also aimed at a broader audience ranging from legal specialists and the legislator to the society of as a whole. In a separate decision, that not has the status of a binding decision, however, the judge may choose to address the audience that suits him best. That means that the argumentation in a separate decision may be aimed at, for example, members of the Court of Human Rights in order to affect future judgments in cases about situations similar to the case in question (Ginsburg, 2010). Another audience the argumentation may be aimed (exclusively) at is the legislator: the judge may write a separate opinion in the hope that, by means of its argumentation, current legislation may be rectified or reformed or possibly give rise to new legislation.³

Take for instance *Lindon, Otchakovsky-Laurens and July v. France*. In this case, the ECHR found that the criminal conviction of the applicants for defaming Jean-Marie Le Pen and his political party, the Front National, was consistent with the European Convention. To be precise, it held by thirteen votes to four that there had been no violation of the Article 10 freedom of expression, and held unanimously that there had been no violation of the Article 6 right to fair trial. In a concurring opinion, one of the judges brought forward the following.

I agree with the findings of the Court in this case but I would like to express certain views regarding freedom of expression and the right to protection of one's reputation. [...] The Convention expressly protects rights of lesser importance, such as the right to respect for one's correspondence. It is therefore difficult to accept that the basic human value of a person's dignity was deprived of direct protection by the Convention and instead simply recognised, under certain conditions, as a possible restriction on freedom of expression.

(*Lindon, Otchakovsky-Laurens and July v. France* (Grand Chamber), nos. 21279/02 and 36448/02, 22 October 2007)

³ In Plug (2011, 342) I demonstrated how the court may seek to address the legislator by means of an obiter dictum.

In this case the judge agrees with the decision of the majority of the court (+/d). However, according to the judge, the argumentation underlying the decision should be different: +/a'. The judge makes use of the opportunity provided by a separate opinion to address the legislator. His standpoint, aimed at the legislator, is the following: the Convention should expressly protect the basic human value of a person's dignity. Such an amendment to the Convention would provide a more acceptable argumentation for future decisions in similar cases.

4.2 Presentational devices

In judicial decisions by the European Court of Human Rights, the court presents the decision and its justification by using an institutional reference to themselves: 'The court considers that...' and 'The court holds that...'. This third-person perspective underlines the impersonal, institutional role of (the majority of) the court. Whereas the court avoids a first-person perspective, the first-person singular pronoun 'I' or plural pronoun 'We' are used in separate opinions. Kaehler (2013, 551-553) points out that when judges use the first person singular they can best describe their personal attitudes. Moreover, the first person perspective makes the judge(s) accountable for the statement in public.

One of the presentational devices used in particular in (partly) dissenting opinions can be observed in the confrontation stage of these opinions. If we look at the way in which the standpoints in (partly) dissenting opinions are formulated, we may notice that the standpoints are expressed in a courteous manner.⁴ In the example of a dissenting opinion under 3.1, the judges use 'we are unable to...' to introduce the difference of opinion. Other expressions of politeness that can be found are 'To my regret, I cannot agree with...' or 'I respectfully disagree with...'. In *From Consensus to Collegiality* (Anonymous, 2011) it is observed that many Supreme Court of the United States dissents also include the phrase, 'I respectfully dissent...' or some variation thereof. According to the authors, these polite formulations of a standpoint in the beginning of a (partly) dissenting opinion may be used to express collegiality and thereby avoid long apologies for deviating.'

⁴ The European commission for democracy through law (2018, 14) notes that as regards limits in the wording of separate opinions, there are only a few countries which have special provisions in this respect. Linguistic features of separate opinions in other systems of law are discussed by Krapivkina (2016) and Langford (2009).

5. CONCLUSION

In this paper I made a first attempt to characterise separate opinions as an argumentative activity type in the domain of legal communication. I focused on separate opinions in the European Court of Human Rights from an argumentation theoretical perspective in order to investigate how institutional constraints deriving from this specific context may affect argumentation brought forward in these opinions. By applying the stage model of an argumentative discussion, I set out to demonstrate that argumentation in a separate opinion relates to that in a judicial decision, but should be considered as an argumentative contribution to a newly initiated legal discussion. By analyzing standpoints and differences of opinion that may occur in the confrontation stage of a new discussion, the two types of separate opinions, dissenting and concurring opinions, could be specified. Finally, I illustrated how differences between strategic maneuvering in separate opinions and in judicial decisions may be explained by observing characteristic of these two activity types.

REFERENCES

- Anand, R.P. (1965). The Role of Individual and Dissenting Opinions in International Adjudication. *The International and Comparative Law Quarterly*, Vol. 14, No. 3, 788-808.
- Anonymous (2011). From Consensus to Collegiality: The Origins of the "Respectful Dissent," Note, (2011), *Harvard Law Review*, vol. 124, no. 5, 1305-1326.
- Bruinsma, F.J. (2008). The Room at the Top: Separate Opinions in the Grand Chambers of the ECHR (1998-2006), *Ancilla Iuris (anchi.ch)*, 32-43.
- Eemeren, F.H. van (2010). Strategic maneuvering in argumentative discourse. Extending the pragma-dialectical theory of argumentation. Amsterdam/Philadelphia: John Benjamins
- Eemeren, F.H. van & R. Grootendorst (1984). Speech acts in argumentative discussions. A theoretical model for the analysis of discussions directed towards solving conflicts of opinion. Dordrecht/Cinnaminson: Foris Publications, PDA 1.
- Eemeren, F.H. van & R. Grootendorst (1992). *Argumentation, communication and fallacies. A pragma-dialectical perspective*. New Jersey: Lawrence Erlbaum Associates.
- European commission for democracy through law (Venice commission) (2018), *Report on separate opinions of constitutional courts*. Council of Europe [https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD\(2018\)030-e](https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD(2018)030-e)
- European Court of Human Rights (2019). *Rules of Court*. Registry of the Court: Strasbourg https://www.echr.coe.int/Documents/Rules_Court_.pdf

- Feteris, E.T. (1999). *Fundamentals of Legal Argumentation. A Survey of Theories on the Justification of Judicial Decisions*. London, Berlin: Springer.
- Ginsburg, R. B. (2010). The Role of Dissenting Opinions. *Minnesota Law Review*. 95: 1, 1-8.
- Jansen, H. (2003). *Van omgekeerde strekking. Een pragma-dialectische reconstructie van a contrario-argumentatie in het recht*. Amsterdam: Thela Thesis.
- Kaehler, L. (2013). First-person perspectives in legal decisions. *Law and Language*, Vol. 15, 1-32.
- Kloosterhuis, H.T.M. (2002). *Van overeenkomstige toepassing: de pragma-dialectische reconstructie van analogie-argumentatie in rechterlijke uitspraken*. Amsterdam: Thela Thesis.
- Krapivkina, O. A. (2016). Judicial Dissents: Legal and Linguistic Aspects. *Journal of Siberian Federal University. Humanities & Social Sciences*. 10, pp. 2449-2459.
- Laffranque, J. (2003). Dissenting opinions and judicial independence. *Juridica International*, VIII, pp. 162-172.
- Langford, C.I. (2009). Toward a Genre of Judicial Dissent: Lochner and Casey as exemplars. *Communication Law Review*, vol. 9, 2, 1-12.
- Manen, N.F. van (2000). The secret of the court in the Netherlands. *Seattle University Law Review*, vol. 24, 569-576.
- Plug, H.J. (2000). *In onderlinge samenhang bezien. De pragma-dialectische reconstructie van complexe argumentatie in rechterlijke uitspraken*. Amsterdam: Thela Thesis.
- Plug, H.J. (2013). Administrative Judicial Decisions as a Hybrid Argumentative Activity Type. *Informal Logic*, Vol. 36, No. 3 (2016), pp. 333-348
- Raffaelli, R. (2013). *Dissenting opinions in the Supreme Courts of the Member States*. Study. Brussels: Policy Department C: Citizens' Rights and Constitutional Affairs.

Can literary fiction be suppositional reasoning?

GILBERT PLUMER

Test Development, Law School Admission Council, USA (retired)
plumerge@gmail.com

Suppositional reasoning can seem spooky. Suppositional reasoners allegedly (e.g.) “extract knowledge from the sheer workings of their own minds” (Rosa), even where the knowledge is synthetic a posteriori. Can literary fiction pull such a rabbit out of its hat? Where P is a work’s fictional “premise,” some hold that some works reason declaratively (supposing P, Q), imperatively (supposing P, do Q), or interrogatively (supposing P, Q?). True, I will argue, although only with much qualification.

KEYWORDS: argument in fiction, didacticism, literary cognitivism, premise-based reasoning, suppositional reasoning, thought experiments

1. INTRODUCTION

Some come close to claiming that literary fiction can be suppositional reasoning. For instance, Green (2010, p. 360) says that Huxley’s *Brave New World* “can plausibly be construed as taking the form of a *reductio ad absurdum*,” which he states as follows:

- 1 Suppose a society were organized along the lines dictated by hedonistic utilitarianism.
- 2 In such a world, people would lack freedom of thought, freedom of expression, and the ability to cultivate the capacities for critical reflection on their surroundings.
- 3 Therefore, in such a world, life would be intolerable to all but those who have lost the capacity for the activities mentioned in premise (2).
- 4 Therefore such a world would be unacceptable.
- 5 Therefore, hedonistic utilitarianism is an incorrect theory of how to achieve happiness.

Comparatively speaking, suppositional reasoning is a puzzling orphan child in argument studies and epistemology. Authors noting the

inattention range at least from Fisher thirty years ago (1989, p. 401ff.) to Rosa (2019, p. 157; cf., e.g., Green, 2000, p. 377 and Dogramaci, 2016, p. 889). On the other hand, suppositional reasoning is well-established in formal logic, particularly as conditional proof, the core idea of which is: ‘Suppose p . It follows that q [1 and 2 in the *reductio* above]. Therefore, if p then q ’. But even this evokes a vexed question, viz., the extent to which a piece of suppositional reasoning can be analyzed in terms of a conditional and vice versa. For instance, statements of the form ‘if p then q ’ have truth-values, whereas statements of the form ‘supposing p , q ’ may appear to lack truth-values. Barnett (2006, esp. pp. 535-536) argues that such (compound) suppositional statements are true if p and q are each true, false if p is true and q is false, yet in contrast to (e.g.) the material conditional, in the two cases where p is false, the statement cannot be evaluated because evaluating it would amount to determining “*whether* [q] while supposing not just *that* [p] but also *that it is not the case that* [p].” On this suppositional understanding of the conditional, the so-called paradoxes of material implication disappear and a door is open to logics alternative to classical logic.

In fact, suppositional reasoning can seem downright spooky, and it may be no accident that deriving *knowledge* from *fiction* can seem equally spooky. Suppositional reasoners allegedly “extract knowledge from the sheer workings of their own minds” (Rosa, p. 157), even where the knowledge is synthetic a posteriori (Balcerak Jackson & Balcerak Jackson, 2013). The idea is that reasoning itself—in the form of suppositional reasoning—might be a source of knowledge distinct from reasoning understood as the means of transmitting knowledge from premises to conclusion. An example Rosa gives (pp. 156-157, 170; the Balcerak Jacksons cite a similar case, pp. 116, 120) of putatively “a priori” knowledge is: ‘Suppose that Lucy is a feminist philosopher. It follows that some feminists are philosophers. Therefore, if Lucy is a feminist philosopher, then some feminists are philosophers.’ (One does wonder about the extent to which the appearance of the proper name ‘Lucy’ here undermines the a priori claim, but never mind.) In order to reason in this way, one need not have any justification that Lucy is a feminist philosopher (nor any justification that some feminists are philosophers); indeed, one might believe that she is not, and have justification that she is not, but be supposing that she is ‘for the sake of argument’. Thus, here it cannot be that knowledge or justification is transmitted from premises to conclusion, in contrast to beliefs arrived at by nonsuppositional reasoning, wherein the premises are taken or asserted to be true.

Can suppositional reasoning be a source of less trivial-seeming knowledge? The Balcerak Jacksons (p. 120) propose a non a priori case:

A human-cannonball has sustained an injury in his profession and wants to prevent a reoccurrence. He asks himself, “*if the stage lighting rigs had been a bit higher, would I have landed in the net?*” In his imagination he visually models the scenario with the rigs higher, and finds that he safely lands in the net, not on the rigs again. He concludes with an affirmative answer to his question.

But, you say, this is just an ordinary ‘thought experiment’. That is right, yet part of the aim of philosophers like Rosa and the Balcerak Jacksons is to argue that it is quite mundane for reasoning alone to be a source of knowledge or justification. Now many contend that at least some works of fictional literature constitute a class of thought experiments (e.g., Carroll, 2002; Elgin, 2007; Swirski, 2007; Mikkonen, 2013; Green, 2010, 2016, 2017). Be that as it may (next section), if works of fictional literature can holistically exhibit reasoning, our question is whether that reasoning can significantly, not purely, be suppositional reasoning. It does not matter for our purposes should their conclusions be based on premise-beliefs as well as suppositions.

2. THOUGHT EXPERIMENTS AND LITERARY FICTION

Tooming (2018, pp. 685-692) distinguishes between “mere supposition,” presumably as in Green’s statement above of Huxley’s *reductio*, and supposition that requires mental imagery that “matches” the supposed propositions(s), presumably as in the human cannonball example. A critical question is whether cases like the latter actually involve reasoning at all, as opposed to something on the order of imaginative engagement. Could imagination function like perception here in giving immediate prima facie justification for believing that things would be/are the way they appear to be in imagination/perception? If even these cases do not involve reasoning, it might be less clear how works of fictional literature could. For one thing, these works are regarded, in their essence, as ‘invitations to imagine’ (vs. believe). The Balcerak Jacksons plausibly argue (pp. 115-122) that the human cannonball kind of case exhibits two ‘hallmarks of reasoning’. The first is that there is a “content gap” to bridge between the first and final mental states in the thought experiment. Following Williamson (2007, ch. 5), they propose that “we evaluate a counterfactual conditional by ‘developing’ its antecedent in certain ways via mental simulation, and then check to see if the consequent is true in the simulation” (p. 121). In contrast, there is ordinarily no content gap to cross in having a perception with a certain content and “forming a belief with the same content” (p. 119), e.g., that there is a red circle. Second, it is appropriate to subject the transitions that bridge the gap to “epistemic appraisal and criticism” (p. 122). It is typically a propos to

ask a thought experimenter to give reasons or justification for the proposed result, whereas it is normally otiose to ask one to give justification for a perceptual judgement (e.g., that we are approaching a log in the middle of the road) other than it looks that way.

As opposed to fiction on the order of 'bodice rippers', pulp fiction, and the like, it is generally held that *literary* fiction is more nuanced; it has a greater richness and complexity of character development, plot, fine description, etc., and also somehow shows insight into human affairs. Is such insight achieved by holistically exhibiting suppositional reasoning? That is our question, so our focus is on literary fiction. Now certainly, the reasoning in thought experiments, if anything, would be suppositional in nature. But there are problems with regarding a work of fictional literature as a thought experiment. For example, within science the epistemic value of thought experiments is regarded as second best (or worse) to real experiments, but there could be no counterpart to this within literary fiction. Moreover, there is an inverse relationship between parameters of evaluation. Factors that make a thought experiment good (e.g., straightforwardness and precision) tend to make a story bad (lack of nuance and subtlety), and vice versa. Egan mentions this (2016, p. 147), and that in contrast to a literary fiction, "the purpose of a thought experiment is *exhausted* in making or contributing to an argument"; its aesthetic qualities are basically irrelevant, as is indicated by the fact that once "we remember how a thought experiment runs, we have no reason to reread it" (pp. 142-143). For instance, the power and cogency of a good philosophical thought experiment may derive from its being a close analogical argument, not from any embedded fictional narrative being *believable* like a novel, play, or short story; consider, e.g., Thompson's (1971) celebrated, though hardly believable, thought experiment involving a famous violinist plugged into your body for life support. This points toward perhaps the most important difference for our purposes: indirectness is not a distinctive feature of thought experiments, but it is for any global argument in a work of literary fiction.

3. DIDACTICISM AND INDIRECTNESS

If a work of literary fiction has a global argument, why should it be indirect? The answer is that otherwise the work would be didactic, which is a distinctive fault for fictional literature. However, given that the term has a negative connotation, didacticism is a flaw to some degree no matter where it appears, and this can lead to confusion. Consider Repp's view. He says (2012, pp. 271, 283):

Works of literature that are too overtly instructive are commonly faulted for being didactic...if we [as literary cognitivists] value literature for the instruction it affords, why would we ever object to overt instruction?...I propose the following answer: overt instruction can arouse suspicion of intellectual vices in the author, such as intellectual arrogance, dogmatism, and prejudice, which can make the lessons the author seeks to convey less rationally acceptable... Didacticism on my view is just as objectionable in works of computational biology and cinematic history as in literary fictions because it is primarily an epistemic rather than aesthetic fault.

Repp tries to assimilate didacticism in literary fiction with didacticism in nonfiction (whether literary or not) as manifesting the same epistemic flaw. But although works of nonfiction such as computational biology and cinematic history could be didactically flawed because they are marked by arrogance, dogmatism, or prejudice, they could not be “*too overtly instructive*” any more than an instruction manual could. After all, being instructive is their express purpose and *raison d’être*. By definition, works of nonfiction aim at achieving veracity and conveying it; they attempt to stick to the facts or tell what actually happened.

In contrast, fictional works, broadly speaking, at most aim at verisimilitude. Repp is right that literary fictions are didactic if they are “*too overtly instructive*,” yet he does not appear to see why. Trite as it may be to be reminded, fictional literature’s significant cognitive value, if any, is conveyed by *showing* insight into human affairs via the character descriptions, narration of events, etc., not by *telling* it—which would make the work didactic. As this truism suggests, *none* of this showing need be intentional or “*lessons the author seeks to convey*” (pace Repp and others, e.g., Gibson, 2009, sec. II; see my 2017, p. 152ff.) Repp (p. 274) says that a literary fiction’s cognitive value can depend “on the extent to which it provides ‘warrant’ or legitimate grounds for accepting the lesson.” But he has the wrong model. Literary fiction is not science, yet he tries to assimilate the two. Compare Swirski (2007, p. 4), who claims that “historical novels transmit knowledge of history much in the same manner that historians transmit it.” If this were so, then there would not be the following sharp asymmetry: For all we know without history, anything in an historical novel could be invented. History is needed to arbitrate, yet historical novels do not arbitrate history.

From these considerations, we see that literary fiction cannot *be* suppositional reasoning, or any kind of reasoning, in a straightforward way; if it were, it would be didactic ‘overt instruction’, which undermines its status as literary fiction and makes it akin to philosophy or science. Thus, the global argument, if any, in a work of literary fiction would have to be somehow uncovered. No doubt fictional narrative

generally makes a supposition (commonly called a ‘premise’) and determines what would, or could very well, follow. For instance, Golding’s *Lord of the Flies* considers what would happen if a group of English schoolboys were stranded on a deserted jungle island and had to fend for themselves and remake society. But these are primarily ‘real’ and probabilistic (mostly causal) consequences imagined by the author; generally, it is only with critical interpretation¹ that there is a transition to more logical or conceptual—hence, argumentative—consequences. This means that if certain works of fictional literature holistically exhibit suppositional reasoning and thereby can constitute a source of knowledge (if the reasoning is good), they do so indirectly within the context of critical interpretation and all the vagaries that can bring. Egan (2016, p. 147) pushes such a point, contending that “we may be able to extract an argument against Stalinism from *Animal Farm*, but...our argumentative criticism of *Animal Farm* would at best target claims we have come to entertain because we read *Animal Farm*, not *Animal Farm* itself.”

For perspective, notice how the suppositional-reasoning approach differs from another possible argumentative approach to regarding fictional literature as a potential source of knowledge. Elsewhere I argue (most completely and recently in Plumer, 2017) that we have a basic intuitive grasp of human nature and the principles that govern it. A literary fiction may evoke these principles in its storytelling, which makes the narrative believable if it is otherwise coherent. So the believability of a fictional story implicates that there is truth there, which amounts to a transcendental argument, and for the appropriately reflective auditor, this contact with truth becomes knowledge.² Here, critical interpretation—and with it the possibility of error, of course—enters the picture in reflectively trying to determine *which* truths of human nature are implicated by the work’s believability, not *that* there are truths there. On the suppositional approach, critical interpretation is

¹My use of the term ‘critical interpretation’ more or less conforms to Gibson’s (2006, p. 444): “Rather than directed at the recovery of linguistic meaning, critical interpretation marks a process of articulating patterns of salience, value, and significance in the worlds literary works bring to view. That is, critical interpretation marks the moment of our engagement with the world of the work, and it has as its goal the attempt to bring to light what we find of consequence in this world.”

²This satisfies the thesis of Literary Cognitivism (LC)—shortly to be discussed in the next section below—because believability with respect to fiction is quite a different thing than it is with respect to nonfiction. If a work of nonfiction is *believable*, it is *worthy of belief*, but the term cannot mean this with respect to fiction.

necessary to display the suppositional reasoning, which then must be determined to be good reasoning, before any knowledge ensues. Conversely, for a believable fiction without critical interpretation, on the suppositional approach all you may relevantly know is the work's 'premise', whereas on the transcendental view you know that as well as that the psychosocial principles the work evokes are mostly true.

4. THE SUPPOSITIONAL REASONING MODEL

Green may be the most ardent proponent of the view that there is "literary fiction that conforms to our suppositional model," a model not presented merely in the guise of thought experimentation (2016, p. 293; see also his 2010, 2017, and forthcoming, sec. IV). Where P is a work's fictional 'premise', Green holds (2016, p. 289) that some works reason declaratively (supposing P, Q), imperatively (supposing P, do Q), or interrogatively (supposing P, Q?). Of course, premise-beliefs (see section 1 above) may enter the picture, and "it is normally appropriate to appeal to a body of background knowledge to aid our reasoning" (p. 290), so the suppositional reasoning need not be pure. Green sees conformity to the suppositional model as the primary way that the thesis of "literary cognitivism," as he construes it, is satisfied. He casts this thesis as

"Literary Cognitivism [LC]: Literary fiction can be a source of knowledge in a way that depends crucially on its being fictional"

(2010, p. 352; 2016, p. 286; 2017, p. 48; quoted approvingly by Maioli 2014, p. 625). Literary cognitivists and anti-cognitivists are all concerned with fiction literature because, by definition, there is no question that nonfictional literature (e.g., an historical or bibliographical work) may yield knowledge. In LC Green tries to say what is special or distinctive about knowledge arising from fictional literature. However, Green never clearly spells out the idea of dependence in LC, although he distinguishes LC from the stronger thesis (he does not endorse) that "the knowledge literary fiction provides is not available through any other means such as journalism, memoir, or research in social psychology," a thesis that might be called "*literary cognitive uniqueness*" [LCU] (2016, p. 286n4). In LCU, the notion of dependence is the usual idea of *cannot exist without*. I propose that an adequate fleshing out of the idea of dependence in LC is that in the path or route to knowledge from the fictional work, the work's fictionality is integral (not necessarily that there is no other path to the knowledge, as per LCU). It is because or partly because of its fictionality that the work yields knowledge.

Green initially considers cases of suppositional reasoning or “suppositions for the sake of argument” in “everyday life” by way of preparing the ground for considering it in literary fiction. His first example (declarative) is: “Suppose we take the 3:17 train to Union Station. Then we can catch the 4:35 from there to the coast, getting there in time for the ferry unless there is some delay...” He claims that “suppositions such as the proposition that we take the 3:17 to Union Station are a species of fictions” (2016, p. 287). But this seems confused. The proposition that we take the 3:17 to Union Station could be a species of fiction only if we do not take that train. Now, before 3:17, all we know is that it is a future fact in the actual world or it is not. Before 3:17, it is a ‘counterfactual’ in only the weak temporal sense that it obtains in neither the present nor the past in the actual world. Yet the ‘premises’ of literary fictions are paradigmatically metaphysical counterfactual possibilities, that is, they obtain in merely possible worlds—not obtaining ever in the actual world.

Continuing with suppositional reasoning in “everyday life,” Green illustrates the directive structure of ‘supposing P, do Q’: “Imagine animated demonstrations of how to change an automobile’s oil filter...the animation, albeit fictional, shows how to do something” (2016, p. 288). Considered as showing how to engage in a *type* of activity, the animation had better not be fictional or else it would purvey falsehoods and misdirect. Considered as depicting a *token* of that activity type, it could be fictional and the dependence of the learning engendered conform to the idea of dependence in LC (though not in LCU of course, because a video of a real oil filter change could be used).

Turning to cases of literary fiction, in the directive vein Green interprets Flaubert’s *Madame Bovary* as “showing how to justify adultery to oneself” (2016, p. 293). I do not see a problem relevant to our concerns, given that the novel’s ‘premise’ is a metaphysical counterfactual, although his reading may be a little obtuse since Emma’s adultery ends in misery for pretty much all concerned. On the other hand, as we have seen (section 1), Green understands Huxley’s *Brave New World* (declaratively) as working out the negative implications of a supposition in the manner of a loose *reductio ad absurdum*. A problem arises in satisfying the dependence requirement in LC if the supposition could simply be an epistemic possibility (‘suppose X, which for all we know, occurs sometime’) or probabilistic (e.g., ‘suppose X, which could very well happen’), not metaphysical counterfactual supposition, that is, distinctively fictional supposition. It is disputable whether *Brave New World*’s supposition that society is “organized along the lines dictated by hedonistic utilitarianism” is actually true of a society somewhere, or at some time was or probably will be true. The same applies (e.g.) to the supposition of Atwood’s *The Handmaid’s Tale* that women become

extremely subjugated under a U.S. totalitarian theocracy, especially as concerns reproduction. The point is, one cannot say that a work imparts knowledge (partly) because of its fictionality *qua* counterfactuality if in key respects its counterfactuality is not evident.

It might be objected that whether a literary fiction's key supposition is counterfactual may vary with how specifically it is formulated. *Brave New World's* supposition could be cast as including (e.g.) that there are no visible signs of aging in the World State, Soma is the state-distributed hedonistic drug, there are biweekly and state-required orgies, hatcheries produce human embryos—all in contrast to natural processes outside the World State in Savage Reservation in New Mexico. Probably not all of this is needed to make the supposition a metaphysical counterfactual. Determining the right level of generality is no doubt an important and difficult question, perhaps even intractable. Nevertheless, it may be that the more the focus is on *particulars* that make a supposition a nonactual possibility, the less likely it is that knowledge or understanding pertaining to the actual world could be gained. Otherwise, Green's formulating the entire *reductio* he sees in Huxley's work in fully general terms would appear to be accidental.

Green regards Stephen King's *Salem's Lot* as having "an interrogative dimension" in that it compels "readers to ask themselves whether there are any epistemic situations in which rationality would oblige them to give up their naturalistic scruples and believe in the supernatural" (2016, p. 292). This kind of case raises two general concerns. First, even assuming that the question posed is a good one or "helps to build a framework in which an intellectual advance can be made" (Green 2017, p. 51), it seems that what would enhance our knowledge or make the advance is the *answer*. It is not clear that good suppositional reasoning in the interrogative form could support LC. Second, the example here raises the issue of *impossible fictions*, that is, ones that involve a logical or metaphysical (not merely a physical) impossibility. Rather than interrogatively, Green (2017, pp. 57-58) considers Stoker's *Dracula* declaratively, and takes it as similarly supposing that its main protagonists, who are "quite rational people," are "faced with empirical evidence undermining...naturalism." If this story showed, as Green appears to suggest, that "commitment to rationality does not by itself guarantee a commitment to naturalism," then the story would provide *that knowledge* partly because of the story's metaphysical counterfactuality. (Green would say LCU is not at issue since as an alternate route to that possible knowledge, he cites Cleanthes' arguments in Hume's *Dialogues Concerning Natural Religion*.) However, it seems that such a knowledge claim is disputable on logical grounds by making the appropriate conceptual connections between the scientific method, rationality, and naturalism (e.g., a commitment to

naturalism is a condition of the possibility of the scientific method, broadly construed, which in turn defines rationality). Taken as not involving such claims, *Dracula* could provide knowledge by considering nonactual metaphysical possibilities that are important for understanding actuality. At any rate, it is particularly hard to see how knowledge could be gained from impossible fictions.³

5. CONCLUSION

Our topic has been whether literary fiction can be suppositional reasoning. We have seen that the reasoning in thought experiments, if anything, would be suppositional in nature, and although it is often claimed that at least some works of fictional literature constitute a class of thought experiments, this claim is misleading. However, we have found that indirectly, within the context of judicious critical interpretation, works of fictional literature can holistically exhibit suppositional reasoning and thereby constitute a source of knowledge (if the reasoning is good) in a way that supports the thesis of Literary Cognitivism. Evident constraints on this include that the form of the suppositional reasoning needs to be declarative or imperative, and that the fictional 'premise' of the work needs to be a metaphysical counterfactual possibility, not merely a temporal counterfactual and not merely an epistemic possibility or probabilistic supposition.

So, yes, it is true that literary fiction can be suppositional reasoning, although only with significant qualification.

ACKNOWLEDGEMENTS: I am grateful to Olivia Odoffin, Kenneth Olson, Teresa Plumer, Katharina Stevens, and audience members at the 3rd European Conference on Argumentation for helpful comments on earlier drafts.

REFERENCES

- Balcerak Jackson, M., & Balcerak Jackson, B. (2013). Reasoning as a source of justification. *Philosophical Studies*, 164(1), 113–126.
- Barnett, D. (2006). Zif is if. *Mind*, 115(2), 519–565.
- Bourne, C., & Bourne, E. C. (2018). Personification without impossible content. *British Journal of Aesthetics*, 58(2), 165–179.

³Although (e.g.) Nolan (2015) can be regarded as arguing that impossible fictions can have cognitive value, Bourne & Bourne (2018) argue that the fictions he considers are not really impossible.

- Carroll, N. (2002). The wheel of virtue: Art, literature, and moral knowledge. *Journal of Aesthetics and Art Criticism*, 60(1), 3-26.
- Dogramaci, S. (2016). Reasoning without blinders: A reply to Valaris. *Mind*, 125(2), 889-893.
- Egan, D. (2016). Literature and thought experiments. *Journal of Aesthetics and Art Criticism*, 74(2), 139-150.
- Elgin, C. Z. (2007). The laboratory of the mind. In J. Gibson, W. Huemer, & L. Pocchi (Eds), *A Sense of the World: Essays on Fiction, Narrative, and Knowledge* (pp. 43-54). New York: Routledge.
- Fisher, A. (1989). Suppositions in argumentation. *Argumentation*, 3(4), 401-413.
- Gibson, J. (2006). Interpreting words, interpreting worlds. *Journal of Aesthetics and Art Criticism*, 64(4), 439-450.
- Gibson, J. (2009). Literature and knowledge. In R. Eldridge (Ed.), *Oxford Handbook of Philosophy and Literature* (pp. 467-485). Oxford: Oxford University Press.
- Green, M. (2000). The status of supposition. *Noûs*, 34(3), 376-399.
- Green, M. (2010). How and what we can learn from fiction. In G. L. Hagberg & W. Jost (Eds.), *A Companion to the Philosophy of Literature* (pp. 350-366). Oxford: Wiley-Blackwell.
- Green, M. (2016). Learning to be good (or bad) in (or through) literature. In G. L. Hagberg (Ed.), *Fictional Characters, Real Problems: The Search for Ethical Content in Literature* (pp. 282-301). Oxford: Oxford University Press.
- Green, M. (2017). Narrative fiction as a source of knowledge. In P. Olmos (Ed.), *Narration as Argument* (pp. 47-61). Cham: Springer.
- Green, M. (forthcoming). From signaling and expression to conversation and fiction. *Grazer Philosophische Studien*.
- Maioli, R. (2014). David Hume, literary cognitivism, and the truth of the novel. *Studies in English Literature*, 54(3), 625-648.
- Mikkonen, J. (2013). *The Cognitive Value of Philosophical Fiction*. London: Bloomsbury.
- Nolan, D. (2015). Personification and impossible fictions. *British Journal of Aesthetics*, 55(1), 57-69.
- Plumer, G. (2017). The transcendental argument of the novel. *Journal of the American Philosophical Association*, 3(2), 148-167.
- Repp, C. (2012). What's wrong with didacticism? *British Journal of Aesthetics*, 52(1), 271-285.
- Rosa, L. (2019). Knowledge grounded on pure reason. *Pacific Philosophical Quarterly*, 100(1), 156-173.
- Swirski, P. (2007). *Of Literature and Knowledge: Explorations in Narrative Thought Experiments, Evolution and Game Theory*. London: Routledge.
- Thomson, J. J. (1971). A defense of abortion. *Philosophy and Public Affairs*, 1(1), 47-66.
- Tooming, U. (2018). Imaginative resistance as imagistic resistance. *Canadian Journal of Philosophy*, 48(5), 684-706.
- Williamson, T. (2007). *The Philosophy of Philosophy*. Malden, MA: Blackwell.

Conspiracy Theories and Reasonable Disagreement

JUHO RITOLA
University of Turku
juho.ritola@utu.fi

This paper will discuss some of the epistemological problems that conspiracy theories present.

KEYWORDS: conspiracy theories, burden of proof, disagreement, independence principle, uniqueness of evidence

1. INTRODUCTION

On the 40th anniversary of moon landing in 2009, Time magazine published a list of what they thought were the most enduring conspiracy theories. While one might quibble about the expression ‘most enduring,’¹ the list certainly noted theories that most of us have heard or seen circulated, and it is still relevant a decade later. According to the list, the moon landings were faked, the CIA assassinated JFK, the 9/11 was a cover-up, Paul McCartney died 1966 and a replica took his place, and various secret societies or even alien reptiles control the world. Just this little sample indicates the great variation these theories exhibit. Some deal with individual events or agents, some with larger collective bodies and agents, and some can be characterized as total, taking into their scope of explanation the world as we know it. While we have good reason to be sceptical of many pieces of information that get to the public sphere, the surprisingly common phenomenon of conspiracy theorizing still seems puzzling: why are people attracted to them? Is there something special about their epistemic or inferential structure? A sceptically oriented citizen might ask if we really know that even the wacky-seeming ones are false. Who should we trust?

In this essay, I will first make some conceptual remarks about the term ‘conspiracy theory.’ This is necessary to set the scene for three epistemic problems that appear relevant in the context of conspiracy

¹ For example, *The Protocols of the Elders of Zion*, a book detailing an alleged Jewish master plan for world dominion, was first published in 1903 in Russia. It had roots in much earlier works, and its claims keep resurfacing in different places.

theories. The first one of these problems is their general epistemic characteristics: can we, from the point of view of applied epistemology, say anything general about conspiracy theories? I hold we can: we are, with some contextual limitations, justified in putting the burden of proof on conspiracy theories.²

The second problem is a strategy sometimes used to defend conspiracy theories, called the expanding strategy, where the defender of a particular conspiracy theory infers that a given source conflicting with the conspiracy theory must be “in on it.” I claim this strategy can be seen as a violation of the principle called the Independence, debated in the epistemology of disagreement. According to this principle, roughly, when you evaluate the epistemic worth of your opponent, you ought to do it in a way that is independent of the point of contention. I argue that the context of conspiracy theories can contribute to the discussion on the epistemology of disagreement.

The third problem is the nature of evidence. We are often willing to make the assumption that evidence is unique: there is just one possible rational attitude to take to what any given body of evidence shows: either accept, reject, or withhold judgment. This assumption means that when we have our evidence-base fixed, there should be no possibility of rational disagreement. I will discuss this from a very broad perspective: I will try to show that if we do not assume uniqueness to be a quality of evidence, we struggle to understand the concept of justification. Finally, I will discuss the argument that Ballantyne and Coffman (2011) have presented, an argument that seems to show this assumption is mistaken. I argue there is a crucial flaw in the argument.

2. WHAT IS A ‘CONSPIRACY THEORY’?

A conspiracy is a fairly common thing, and most of us can name one without hesitation. They seem as much an integral part of history books (for example, the assassination of Julius Caesar) as they are of modern news coverage (for example, the Volkswagen emission scandal). However, when ‘conspiracy’ is coupled with ‘theory,’ the connotations are completely different. They seem something that are, well, out-there. Perhaps this is the reason that the philosophical interest in them is quite recent, starting with Brian Keeley’s 1999 paper ‘Of Conspiracy Theories’ in *The Journal of Philosophy*.³ This paper discussed unwarranted

² While this claim has been made by various authors, I should note my particular intellectual debt to the work of, and discussions with, Juha Räikkä on this and the conceptual question.

³ Karl Popper famously discussed the conspiracy theory of society in his *Open Society and Its Enemies*, but that had more to do with the general issue of intentional explanation in social sciences than with conspiracy theories as such.

conspiracy theories, claiming that such theories typically run counter to some received or obvious account; assume nefarious intentions⁴; seek to tie seemingly unrelated issues; the truths behind them are well-kept secrets; their chief evidence is errant data, left unexplained by the official accounts but which actually require explanation conflicting with the official account (Keeley 1999, pp. 116-118). However, Keeley also noted that these criteria do not distinguish between conspiracies we are justified in believing (1999, p. 118). Since then, the topic has attracted new authors and there is now a burgeoning literature on the topic, with new titles and articles appearing every year.⁵

Regardless of whether we could come up with necessary and sufficient conditions for a conspiracy theory, I think a crucial difference between a conspiracy, a run-of-the-mill term of historical and social explanation, and a conspiracy theory, is that the former, but not the latter, are explanations that enjoy the support of epistemic authorities. What, then, are the epistemic authorities? In broad terms, they are the mainstream media, investigative journalists, state and local authorities and agencies, and the academic community.⁶ Yet, the fact that an explanation is in conflict with the received view does not imply that it does not receive support from powerful agents.

The idea of taking this difference in the epistemic support as a central feature is to approximate the popular use of the term, and then study that phenomenon, without thereby taking a stand on their objective epistemic quality. An alternative strategy proposed in the literature (see for example Pigden 2007, p. 230; Dentith 2016, p. 587; Basham 2018) is to revise the term so that all explanations that refer to conspiracies are called conspiracy theories. The motivation is that it might remove the stigma involved with the term. Putting explanations perceived as legitimate under the same term might help to expose schemes, which the relevant agents are keen to hide, and thus serve public interest. After all, we know that governments and other powerful agents have engaged in dubious plots, so we should not naïvely believe that powers that be are always to be trusted.

⁴ 'Nefarious' must be relativized to speakers; for example I do not think the Reagan administration did themselves think they were involved in something nefarious.

⁵ A quite comprehensive literature list, put together by an interdisciplinary Cost Action project 15101: "Comparative Analysis of Conspiracy Theories," can be found in <https://conspiracytheories.eu>. For a conspiracy theoretic -view on the motivations and aims of that project, see for example <https://allunreal.com/blog/controlled-academic-opposition/#more1363>.

⁶ This list is neither clear-cut nor immutable. We can lose trust in some agents on it, and agents can gain/regain trust by some actions. Also, various private citizen groups can become part of it by building a respectable track-record.

However, I do not think it is clear that philosophers could affect such a change in the use of an established term.⁷ Ideally, conspiracy theories should get a fair review anyway, but of course, we do not live in an ideal world, and sometimes people voicing legitimate concerns are ridiculed, even smeared, publically. Yet, the proposed shift in the meaning of the term would hardly prevent that or empty languages of terms for character assassination.

3. EPISTEMIC ISSUES WITH CONSPIRACY THEORIES

A conspiracy theory, then, is an explanation of some events that lacks the backing of sources on which we are used to relying. But if we pause to consider them just a little longer, we might notice that they really are a worthwhile subject. Just consider the way one source of the received view in philosophy, *The Stanford Encyclopaedia of Philosophy*, divides the field of social epistemology into three broad topics: testimony and trust, social/collective knowledge, and the reliability of institutions and systems. The issue of conspiracy theory touches on all these topics. Also, to mention one popular topic in modern epistemology: the contents of some of these theories are topics on which I have often disagreed with my epistemic peers. Let me now try to show their intrinsic interest by considering some individual epistemic questions.

3.1 Generality and the burden of proof

The first issue concerns the general nature of the conspiracy theories.⁸ As noted in the beginning, conspiracy theories exhibit a great deal of variation, ranging from the explanation of one single event to total theories that aim to explain the state of the world *in toto*. The amount and type of evidence varies greatly, and while some of them have premises requiring leaps of faith and inferences defying known canons, some of them may make us wonder if we know the full story quite yet. We know that governments have engaged in all kinds of dodgy plans. This leads to the question whether there really is anything general that one can, or should, say about them. In my view, there is. I would argue that:

⁷ But perhaps such pessimism is not warranted. After all, some philosophers, and also other scholars, have long expressed doubts whether terms like ‘truth,’ ‘validity,’ or ‘justification’ have any substantial content, and that attitude has surfaced in the public sphere.

⁸ A more detailed version of the argument in this section is in Räikkä and Ritola (forthcoming).

1. Conspiracies are typically revealed by epistemic authorities, helped by leakers.
2. In (W*^{NC})⁹, the epistemic authorities have earned their position and reputation through their epistemic efforts.
3. Conspiracy theories are explanations of social events that contradict the explanations of the epistemic authorities.

This lends support to:

4. Typically, a given conspiracy theory has the burden of proof.

Not everyone in the field finds this argument cogent, especially when they place it into their own context. The view purporting to say anything general about conspiracy theories has been termed ‘generalist.’ (Buenting and Taylor 2010), as opposed to ‘particularism,’ which avers that every conspiracy theory needs to be examined in their own right (see for example Pigden 2007, Dentith 2018).

I think there is mistake here. As Räikkä (2018) has noted, a generalist may well admit that every case needs to be judged on their own merits, but it does not follow that nothing that can be said about conspiracy theories in general. The general rational believability (or acceptability) is dependent on the general reliability of the relevant epistemic authorities. At least in Finland, the authorities work tolerably well.

The issue of general rational believability of the sources is directly related to the nature and extent of testimonial knowledge¹⁰. Our knowledge is deeply dependent on others, and it is almost automatic. If we cannot rely on others, the number of things we know drops drastically. The traditional, startling example is that we do not know even our own name, if we try to exclude from the things we know, the things we know based on testimonial knowledge. This deep epistemic dependency on others is present from the mundane pieces of knowledge to expert knowledge. In modern academia, also experts are dependent on other experts, and the specialization of scientific and scholarly knowledge is ever increasing. Naturally, it is not the case that we always know, because others know; error and deliberate misinformation can spread because of this dependency.

In undemocratic, closed, and badly corrupt societies people do not trust the media, or the official sources. But it is worth bearing in

⁹ By (W*^{NC}) I mean the actual world, Nordic countries. I need to stick to what I know.

¹⁰ For an introduction, see Stanford Encyclopaedia of Philosophy, s.v. ‘Epistemological Problems of Testimony.’

mind that in those cases, it is also common knowledge why the official sources cannot be trusted. And wherever one's surroundings stand on the Corruptions Perception Index¹¹ or something to that effect, we have reason to be critical of our epistemic authorities. But it is important to note that lists of errors by epistemic authorities are provided by those very same epistemic authorities. They are evidence of their general reliability.

Matthew Dentith has argued that

[i]t is possible that we live in a society which merely looks open [...] but that apparent openness might be the product of those very same conspiracies. [...] Our judgments about the prior probability of conspiracies in the past make claims of conspiracy [...] worthy of consideration [...] (Dentith 2018, p. 7)

David Coady has argued that

[i]t may be that in an ideal society official stories would carry an epistemic authority such that it would almost always be rational to believe them. But that is not our society, nor, I suspect, is it any society that has been or ever will be. (Coady 2007, p. 199)

To reiterate, the generalist need not hold that it is always rational to believe the official story. Even large-scale conspiracies are possible, and we have no reason for complacency about the official stories. Claiming anything beyond this logical possibility depends on what you have reason to believe. But it is essential to ask how do we get the prior probabilities about the conspiracies (i.e. established historical facts)? It seems we are at the mercy of the official stories. If we take the sceptical scenarios that Dentith and Coady play with seriously, we lose, not only knowledge of our names and personal origins, but our general knowledge of history and general knowledge of our surroundings. This includes the bulk of our knowledge, including the knowledge of there ever being conspiracies!

A third objection to the generalist position is that some conspiratorial scenarios "are too 'toxic' for our usual institutions of public information to disseminate to the public" (Basham 2018, p. 73). Yet, it seems that quite a few conspiracies about for example political

¹¹ Provided by Transparency International. The reports can be found at <https://www.transparency.org/research/cpi/overview>.

corruption, targeted killings, and questionable military campaigns, have been made public. 'Toxicity' does not seem to work all the time.

Still, I agree with a number of authors (for example Basham 1999, 2003, 2006; Hagen 2018) that we should not think that conspiracy theories are automatically false or always produced by questionable inference mechanisms. The issue of human rationality is complex, and though it seems that the failings of human inference cannot be explained away, they are not specific to conspiracy theories.

3.2 The expanding strategy

The expanding strategy refers to an argumentative/inferential move in which the defender of a particular conspiracy theorist argues or infers that a source providing information conflicting with the theory must be "in on it." The claim that this strategy is common among conspiracy theorists was discussed already in Keeley's seminal article (see 1999, p. 122). The idea there is that a given conspiracy theory can initially involve a small circle of actors, but as positive evidence for the theory fails to obtain, more and more actors are implicated by the theory, and eventually, this ever increasing scepticism makes the theory irrational.

We seem to have an interesting dilemma here. On the one hand, this kind of a move can protect a theory to such an extent that it is unfalsifiable, on the other hand, as many authors have noted, if an effective conspiracy is actually taking place, you should expect false evidence¹². Keeley's arguments have faced criticism of varying quality, but we cannot cover that here.¹³ The expanding strategy is not specific to defence of conspiracy theories: one can apply the same dubious strategy to dismissing evidence about a conspiracy. To me, the dilemma is about sailing between the Scylla of extreme scepticism and the Charybdis of naivety.

Say, then, that you disagree with your epistemic peer about some conspiracy theory. For example, you believe that the Holocaust did happen, and the six million dead –claim is roughly correct.¹⁴ Faced with this disagreement, what should you do? The epistemology of

¹² For example, "if powerful actors are trying to hide something it only stands to reason that confirming evidence will be hidden and red herrings will abound." (Uscinski 2018, p. 5)

¹³ See Räikkä 2009 for review and development.

¹⁴ The modern conspiracy theory on the Holocaust does not claim that there were no concentration camps, but that the 'official' number of people dead is a gross exaggeration. See for example http://content.time.com/time/specials/packages/article/0,28804,1860871_1860876_1861026,00.html.

disagreement –literature presents, roughly, two main positions.¹⁵ First, you can hold your view in the face of disagreement. After all, if you are in possession of the evidence and reviewed it carefully, you should hold your own. This known as the steadfast –position. Alternatively, you could think that since you are disagreeing with your peer, you have no reason to believe that you are more likely to be right than your peer. This could lead you to reduce your confidence in the claim, or even suspend judgment. This is known as the conciliatory –position. A partial motivation here is that you would need at least some reason to discount your peer’s opinion as we want to align our beliefs with our reasons, and peer’s opinion is a reason.

An essential part of the conciliatory position has been a principle called Independence, which states that when deciding what to do in a disagreement, the reasons because of which you discount your peer’s opinion must be independent of the disagreement. The motivation is that it would seem question-begging to reason: I have reasoned that p. My peer disagrees with me on whether p. Therefore, she must be wrong. Or, alternatively, to reason: I have reasoned that p. My peer has reasoned that not-p. Therefore, he is not my peer. Further, were one to repeat such inferences, there is no number of people disagreeing with you that you could not dismiss. But that is unintuitive.

Thomas Kelly (2013) has criticized the principle forcefully, and a central counterargument he brings to bear on it is the case of a Holocaust denier. He first formulates independence as:

(I): In evaluating the epistemic credentials of another person’s belief about P, in order to determine how (if at all) to modify one’s own belief about P, one should do so in a way that is independent of one’s assessment of those considerations that led one to initially believe as one does about P. (Kelly 2013, p. 40)

As Kelly (2013, p. 39) notes, in many cases, where the evidence-base is quite large, this is impossible: I am not aware of all the things that led me to believe some historical account, so how do I actually conduct the evaluation independent of it?

Suppose I defend my position by referring to history books, my visit to Auschwitz, documents containing, and films referring to, first-person accounts and so on, and my friend counters: “Sure, there is all that evidence, but it is actually misleading: the conspiracy about the

¹⁵ Very few would now defend these major positions as described, but the discussion proceeds as responses to, and developments of, them. For an introduction and references, see <https://plato.stanford.edu/entries/disagreement/>.

Holocaust has produced it. Surely you don't think they would try to pass a lie this big without a lot of misleading evidence? In situations like this, it is all the more important that you conduct your evaluation of our disagreement independently of those considerations that led you to believe that it did occur."

Kelly argues the problem with straightforwardly discounting everyone who disagrees with you is that it is dogmatic (2013, pp. 43-49). But the case of the extreme conspiracy theorist makes even worse use of the Independence –principle. It is not obvious she is being dogmatic: if there is such a conspiracy, you should expect misleading evidence. In the Holocaust –case, (I) seems unacceptable, but we lack an explanation of just what is wrong with it.

I propose that one aspect of the relevant evaluation of conspiracy theories is the amount of damage done to my overall evidential-base and belief-acquiring methods, outside the issue on which I disagree with my peer. This idea gives some credit to the Independence –principle, but not quite in the original sense. The idea is that if in order to believe the theory, I must discredit my beliefs and belief-acquisition methods overall, I should be epistemically entitled to ignore either the official version or the conspiracy theory (whichever case we happen to be dealing with).¹⁶

3.3 Conspiracies and Uniqueness of evidence

The discussion in the previous section assumed that if we could agree on the evidence on a given conspiracy theory, we could come to a rational agreement on it. This assumption is based on the idea that any given body of evidence can justify at most one rational attitude towards some proposition: either affirmation, rejection, or suspension of judgment. This assumption is called the uniqueness of evidence, and it, while quite intuitive in its own right, seems necessary for conciliationism¹⁷: if a given body of evidence makes several attitudes justified, the appeal of conciliating based on disagreement disappears, and we can both remain steadfast in our position. I will first try to present some considerations how we should understand uniqueness

¹⁶ This idea is very much in spirit of Christensen's (2011) account. He defends the Independence principle (2011, see especially pp. 15-16), and uses the difference between positive and negative undermining, which he defends by noting that the latter would require us to have non-question-begging response to the sceptic. But I think his account does not fully appreciate the seriousness of the charge of question-begging that both sides in a conspiracy debate can raise against each other. I discuss this more fully in "Disagreements in and about Conspiracy Theories" (forthcoming).

¹⁷ But see Christensen 2007.

and then provide a sketch of an argument showing that one argument against it is fallacious.

Define uniqueness as:

(U) For any given proposition and total body of evidence, some doxastic attitude is the one evidence makes rational (justifies) toward that proposition.

We need to first set some assumptions to rule out some uninteresting counterexamples to the principle. We must assume that if two arguers disagree on what the evidence justifies, and we are to examine uniqueness, the difference of opinion should not be the result of different: a) credences placed on the premises; b) background theories; or c) methods of justification or ways in which the evidence is processed.¹⁸

But there are many theories of justification. Can we apply uniqueness to all of them? For example, take the following (mock) definition of a causal account of justification:

(CJ) S is justified in believing that p if and only if the fact p is causally connected in an “appropriate” way with S’s believing p.

The difficulty here is that such an externalist theory grew out of the frustration to the persistence of the problem of scepticism. The idea was that it is the notion of justification that we need to get rid of, if we are to avoid scepticism. Still, bearing this complication in mind, note that if the proponents of (CJ) approved of (U), they would be keen to hold that it would not be possible for there to be two persons that had the exact same inputs to the causal processes, placed equal value to those inputs, applied the exact same causal processes, and ended up with different beliefs. Indeed, it seems that if that were to happen, we would wonder if causality was present. Be that as it may, we should note that uniqueness is stronger than (CJ), which does not rule out different outcomes.

Consider next an early, to-be-developed, version of reliabilism:

(R1) If S’s believing p at t results from a reliable cognitive belief-forming process (or set of processes), then S’s belief in p at t is justified. (Goldman 1979, p. 13)

It would seem natural to doubt the reliability of a cognitive belief-forming process (or set of processes), if two exactly similar processes

¹⁸ I discuss these assumptions, and the general argument, in more detail in a longer version of this paper that I am happy to provide upon request.

resulted in different results. But again, note that it could happen: for a process to be reliable, it would not have to be 100 % reliable.

But here lies a rub: the escape I just provided assumes that if identical process resulted in different results, we would doubt its reliability and try to explain it away by noting that it need not be 100 % reliable. But that assumes the idea we were supposed to be looking at: uniqueness. We want a theory of justification that avoids arbitrariness (i.e. different results from the exact same processes), but since the sceptical dilemma lies at the root of epistemological theorizing, we do not want to hold that the account of justification is infallible. Infallibility as a requirement for a theory of justification makes the theory vulnerable to the sceptic, who is keen to point out: "Well, you accept that we make mistakes. So, how can you be sure that you are not mistaken in things you consider knowledge?" The hinge proposition of much epistemological theorizing today is that we must not make the requirements of knowledge so stringent that no theory can pass the muster; there must be room for mistakes.

Now consider the idea of a mistake. If the exact same evidence (in the assumed sense), really could make different beliefs justified via identical processes, what would count as a mistake? I could not even posit the idea that less than 100% reliability, if I did not assume that there is chance of me being wrong. What do we mean by 'being wrong here?' Well, we mean exactly that I should have come to a specific conclusion, and the processes that do not come to that specific conclusion, are wrong. But why are they wrong? Well, because we assume that from that evidence, with those priors, and that process, I should have come to a different conclusion. But that is just the thing we were trying to study: that the evidence should mandate a specific conclusion, since it is unique, and when different results come about, we have to explain something. But we need not explain coming to different conclusions, if we just get rid of uniqueness. But now we have face the possibility that we have won the debate against the sceptic at the cost of losing the idea of justification; namely that justification is something that has the tendency to lead correct, i.e. not mistaken, results.

The two theories above belong to a group called externalism. The mortal enemy of externalism is internalism¹⁹, but when it comes to uniqueness, it seems both camps are equally uncommitted. Consider one classic formulation:

D2. S is justified in accepting p at t on the basis of system X of S at t if and only if p coheres with X of S at t. (Lehrer, 1988, p. 341)

¹⁹ Cf, for example, *Stanford Encyclopaedia of Philosophy*, s.v. 'Internalist vs. Externalist Theories of Justification.'

This formulation does not require uniqueness. The traditional critique of coherentism is that many things can cohere with the system, which is thought to be unintuitive. Since Keith Lehrer aims at undefeated justification, i.e. knowledge here, he introduces:

D3. *p* coheres with *X* of *S* at *t* if and only if all competitors of *p* are beaten or neutralized for *S* on *X* at *t*. (Lehrer, *ibid.*)

So, again, uniqueness is a separate, stronger requirement than the basic theory of justification. Whether the sceptical competitors can be beaten or neutralized is a different issue. But arguably, the sceptic has different priors.²⁰

Finally, a typical evidentialist theory holds that:

(EV) Doxastic attitude *D* toward proposition *p* is epistemically justified for *S* at *t* if and only if having *D* toward *p* fits the evidence *S* has at *t*. (Feldman and Conee 1985, p. 15)

We can see, again, that uniqueness is stronger than this theory of justification: different types of attitudes could fit the evidence. But we can certainly note an uneasiness in saying that having the exact same evidence, the exact same priors, and the exact same background theory does not rule out having different but equally fitting responses to the evidence. But (U) could added to (EV): the one attitude that the evidence made rational would be *the* fitting response.

Nathan Ballantyne and E.J. Coffman (2011) have presented a complicated argument against uniqueness, and this argument is based on the idea that since uniqueness implies evidentialism, but evidentialism does not imply uniqueness, uniqueness is stronger than evidentialism. As we just noted, this seems to be the case. Ballantyne and Coffman then use this to argue that (U) rules out almost all other theories of justification than evidentialism, which is absurd. But given that we have noticed that uniqueness is stronger than typical theories of justification, we have reason to be sceptical of the implication. If the fact that (U) is stronger than (EV) implies the account of justification given by (EV), then, in like manner, (U) implies all accounts of justification, which is absurd. What 'being stronger than' actually means is that if you buy into the given theory, uniqueness will have something to say about the results of that theory. Uniqueness is an assumption about the true

²⁰ I do not mean this as a general refutation of scepticism; I doubt it can be had. But this is one way of looking at the whole problem: the sceptic is overly worried about the possibility of us being completely wrong.

nature of evidence, not a theory of justification. If so, you might still have reason to debate with your disagreeing peer, or at least, search for differences in evidence, priors, or background theory.

4. CONCLUSION

In this essay, I first presented the topic of conspiracies, and then argued that in the context that this essay was written, it is reasonable to place the burden of proof on conspiracy theories. Next, I examined an interesting epistemic strategy that conspiracy theories bring to fore: the expanding strategy. Finally, I concluded with a discussion on the uniqueness of evidence.

The motivation for discussing the expanding was that it seemed like a violation of the principle of Independence, a principle that has been the topic of intense philosophical debate in the epistemology of disagreement. It seems to me that independence is an attractive principle that allows us to avoid some forms of begging the question. Yet, based on the discussion, it does not seem at all clear to me that it can be upheld. The question that made this discussion relevant in general is the idea that in order for us to debate rationally, we need some common starting points. But the expanding strategy, both in the case of the extreme sceptic and the naïve follower of official information, seems to threaten the possibility of there being common starting points in political discussions. I think discussing expanding with insights from the disagreement literature allows us to understand the epistemic situation better.

Finally, I discussed an argument to the effect that the evidence we bring to bear on disagreements might not be unique. This requirement, I argue, is essential to our cognitive efforts; to having justified beliefs in the first place. If we cannot assume that the evidence is unique, we seem to have little reason to even debate with our fellow citizens. I tried to show that while theories of justification typically do not take uniqueness as a requirement of being justified, the idea of justification seems threatened if we do away with it. Finally, I tried to show that one particular argument against is fallacious. This, of course, does not go far in proving that uniqueness should be accepted, but, as I argued, we seem lost without it too.

REFERENCES

- Ballantyne, N. & Coffman, E.J. (2011). Uniqueness, Evidence, and Rationality. *Philosophers' Imprint*, 11(8), 1-13.
Basham, L. (2018). Joining the Conspiracy, *Argumenta*, 3(2), 271-290.

- Buenting, J. Taylor, J. (2010). Conspiracy Theories and Fortuitous Data. *Philosophy of the Social Sciences*, 40(4), 567-578.
- Christensen, D. (2007). Epistemology of Disagreement: The Good News. *The Philosophical Review*, 116(2), 187-217.
- Christensen, D. (2011). Disagreement, Question-Begging and Epistemic Self-Criticism. *Philosophers' Imprint*, 11(6), 1-22.
- Dentith, M. (2016). When Inferring to a Conspiracy Theory might be the Best Explanation. *Social Epistemology*, 30(5-6), 572-591.
- Feldman, R. & Conee, E. (1985). Evidentialism. *Philosophical Studies*, 48(1), 15-34.
- Goldman, A. (1979). What is Justified Belief?. In G. Pappas (Ed.), *Justification and Knowledge* (pp. 1-25). Boston: D. Reidel.
- Keeley, B.L. (1999). Of Conspiracy Theories. *Journal of Philosophy*, 96(3), 109-126.
- Kelly, T. (2013). Disagreement and the Burdens of Judgment. In D. Christensen & J. Lackey (Eds.), *The Epistemology of Disagreement: New Essays* (pp. 31-53). Oxford: Oxford University Press.
- Lehrer, K. (1988). Metaknowledge: Undefeated Justification. *Synthese*, 74(3), 329-347.
- Pigden, C. R. (2007). Conspiracy Theories and the Conventional Wisdom. *Episteme*, 4(2), 219-232.
- Räikkä, J. (2018). Conspiracies and Conspiracy Theories: An Introduction. *Argumenta*, 3(2), 205-216.
- Popper, K. (1945). *Open Society and Its Enemies*. Princeton: Princeton University Press.
- Uscinski, J. (2018). The Study of Conspiracy Theories. *Argumenta*, 3(2), 233-245

Inference to the Best Metaphor

CRISTIÁN SANTIBÁÑEZ

Universidad Católica de la Santísima de Concepción, Chile
csantibanez@ucsc.cl

I propose that we produce some of the most important inferences in a deliberative context, when we reason metaphorically. Just as we select a particular hypothesis because it covers certain facts in the best way and then explains the events, we select a particular conceptual metaphor because it allows us to mediate the social or cultural differences between the interlocutors and in this way the framing of part of the future exchanges.

KEYWORDS: Argumentation, Conceptual Metaphors, Deliberation, Dual Process Theory, Inference, Reasoning.

1. INTRODUCTION

In this work the target is the speaker, the producer of a metaphor. It is not about the comprehension processes to understand metaphoric expressions. It is not about the place and role of a metaphor or a metaphoric expression in an argumentative chain (standpoint, or argument, or warrant).

The specific question of this piece is: why do we use, or need to use, a metaphoric expression in an argumentative activity? With the title of “Inference to the best metaphor”, I take Harman’s idea to propose an explanation why we use metaphoric expressions in argumentative contexts. Particularly, I propose that we produce some of the most important inferences in deliberative contexts (or practical reasoning), when we reason metaphorically.

In other words, we deliver inferences to the best metaphor. This would be the reason why we use metaphors. Just as we select a particular hypothesis because it covers certain facts in the best way and then explains the events, we select, consciously or unconsciously, a particular conceptual metaphor because it not only covers the facts in the best way, but also because it allows us to mediate, or construct common grounds for, the social or cultural differences between the interlocutors and in this way the frame of part of the future exchanges. More importantly, we produce inference to the best metaphor (IBM henceforth) because if the metaphoric frame is challenged, we look for specific correspondences as evidence. A challenged IBM favours the

argumentativeness. This work is about the benefits of arguing by means of metaphors.

2. METAPHORS AND REASONING

The topic on metaphors and reasoning only recently has attracted systematic attention (Thibodeau & Boroditsky, 2011, 2013; Steen, Reijnierse & Burgers, 2014). But the perspective is metaphoric understanding: how the audience reacts to (or processes) a metaphorical expression.

Thibodeau and Boroditsky's (2011) study, for example, emphasizes that metaphorical recognition systematically frames metaphorical production, this is, if we are exposed to a discourse that uses the conceptual metaphors "a virus infecting the city" or "a wild beast preying on a city", those who have heard or read the former ("virus infecting a city") produce expressions suggesting to investigate the source of the virus and to implement prevention measures to decrease the spread of the virus.

According to the results of Thibodeau and Boroditsky, the power of metaphor is covert: when given the opportunity to identify the most influential aspect of a crime report, participants ignore the metaphors; people quote or refer to crime statistics. For Thibodeau and Boroditsky (2011) even fleeting and seemingly unnoticed metaphors in natural language can instantiate complex knowledge structures and influence people's reasoning in a way that is similar to the role that schemas and scripts have been argued to play in reasoning and memory. For these authors, metaphorical frames can play a powerful role in reasoning because they implicitly instantiate a representation of the problem in a way that steers us to a particular solution. In other words, metaphors relieve the cognitive costs of thinking social issues.

3. METAPHORS AND ARGUMENTATION

The relationship between metaphors (or properly conceptual metaphors) and argumentation (or properly argumentative dialogues) has also received little attention (Ervas & Ojha, 2019; Macagno & Zavatta, 2014; Oswald & Rihs, 2014; Santibáñez, 2010; Wagemans, 2016; Walton & Hyra, 2018).

Ervas and Ojha (2019) emphasize that the revitalization of conventional metaphors in the premises are guided by the need to confirm participants' belief in the conclusion. The authors conclude that in order to produce an argument with a novel metaphorical meaning without being misinterpreted, participants prefer to make the metaphorical meaning explicit. According to Ervas and Ojha results show that it is easier to produce believable fallacies with conventional

metaphors than with novel metaphors, where the implicitness of the metaphorical meaning is abandoned in favour of the clarity and understandability of the argument.

Some ideas of Oswald and Rihs' (2014) analysis of extended metaphors come close to my proposal. Particularly, their conclusion that extended metaphors induce a constant process by which the receiver perceives its relevance and takes it as a reliable piece of information (e.g. correspondences between conceptual domains through a text). For these authors, the continuous use of a metaphorical frame in a controversial dialogical context contributes to belief fixation. The idea would be that every occurrence of an additional aspect of the source domain, to the extent that it is mapped onto the target domain in a plausible manner, may serve as a confirmation of the overall relevance of the initial metaphorical construal. For these reasons, extended use of metaphors is the result of the cumulative nature of discourse.

Oswald and Rihs' (2014) approach is, nonetheless, an echo of Blackmore's (1992) idea: the speaker's metaphors encourage the addressee to further process utterances to discover additional implicit contents. These are contents that "justify the speaker's utterance as the best means of representing his thoughts, and it is these implicatures which explain why even rather standardized examples of metaphor cannot be paraphrased without loss."

4. INFERENCE TO THE BEST EXPLANATION

The literature on inference to the best explanation (henceforth IBE) is, simply, massive and the approaches to the topic range from the idea that IBE and abduction are somehow the same (Thagard, 1978, 1981, 1988; Lipton, 2004: 56; Walton, 2004: 10), to that IBE is a special form of abduction (Gabbay & Woods, 2005: 270), or even that abduction is a special form of IBE (Schurz, 2007).

Harman (1965) himself points out that the Inference to the best explanation corresponds approximately to what others have called abduction, that is, the method of hypothesis. The core idea is that IBE is the inferential practice by which human beings go from the recognition that a hypothesis could explain certain evidence, to the truth of that hypothesis. For example, "when a detective puts the evidence together and decides that it must have been the butler, he is reasoning that no other explanation which accounts for all the facts is plausible enough or simple enough to be accepted" (Harman, 1965: 89). We proceed in this way daily, just like the detective, because we consider, consciously or unconsciously, that we are best justified when the hypothesis is more plausible, simpler, or less ad hoc. One of Harman's points of departure for proposing this description is the existence of the robust human experiential habit of assuming regularities and/or correlations in order

to explain a past event or to predict a future situation, such as the case of going from “All observed As are Bs” to “The next observed A will be B”.

5. IBE AND IBM

Just as we select a particular hypothesis because it covers certain facts in the best way and then explains the events, we select a particular conceptual metaphor because it not only covers the facts in the best way but also because it allows us to frame our future exchanges (Lakoff, 2006).

It is pointed out that IBE is fallible, when the inclusion of additional premises can transform the inference from correct to incorrect. As such IBM is fallible, when the inclusion of additional correspondences can weaken the metaphorical inference force from accurate to inaccurate. IBE as an inferential process is more natural, simple and consistent with the given data. IBM as an inferential process is more natural, simple and consistent with the understandability (given data?). The power of IBE is its quality in explaining. The power of IBM is its quality in defining how to conceive an issue.

6. SCHEMA-IMAGE

One of the points of departure of the cognitive view on metaphors (Johnson, 2017) is that there is a pragmatic principle of continuity between body and reason. The core idea is that we have not developed two separate logical and inferential systems. In this perspective, inferences are carried out via de corporeal logic of our sensorimotor capacities. Via the source-to-target domain mapping, the corresponding logical inferences are drawn in the target domain. According to Johnson (2017), there is a definite spatial or bodily logic of containment that arises in our experience with containers:

- A. An entity is either inside the container or outside it, but not both at once.
- B. If I place an object O within a physical container C and then put container C inside another container D, then O is in D.

For example, if we take the conceptual metaphor CATEGORIES ARE CONTAINERS, then the category “human” is contained in the category “animals”, which is contained in the category “living things”. So, the syllogism would be:

- A'. An entity either falls within a given category or falls outside it, but not both at once (e.g. Charles cannot be a man and not a

man at the same time, in the same place, and in the same manner).

B'. If an entity E is in one category C', and C' is in another category D', then that entity E is in the category D' (e.g. all men are mortal [C' is in D''] and Socrates is a man [E is in C'], therefore Socrates is mortal [E is in D'']).

CONCLUSIONS

I agree with Thibodeau and Boroditsky (2011) that it is a kind of bias to deny the influence of metaphors in our reasoning and arguments. The bias would be that we would be less rational to admit it.

In my view, IBM is a robust mechanism to fix beliefs in a distributed cognition, where biological human resources are deposited in artifacts and symbols. More importantly, in a deliberative context (e.g. political discourse), IBM is an invitation to react. From the point of view of the speaker, if challenged, the speaker's metaphors encourage the speaker himself to further make explicit the relevant mapping or other metaphoric expressions.

Certainly, there are many points that need further consideration. For example, as Gibbs & Ferreira (2011: 225) emphasize "Speakers may only intend a small part of what a conceptual metaphor makes available". Speakers may only be aware of a part of the correspondences of a conceptual metaphor. So, a difficult problem would be determining the awareness of mapping vs. degree of automaticity in production. This problem can be approached by taking some ideas from Dual Process Theory: as DPT predicts (Evans, 2010), an intuitive and automatic (system 1) way of thinking will effortlessly produce a behavior for familiar tasks because this behavior produces beneficial results. Seemingly, we would emit the best conventional metaphorical expressions to frame common issues because we have experienced beneficial results by doing it. In principle, no need for awareness of mappings. If the topic, the context and the participants demand it, we will search (system 2) for the best metaphorical construal to frame an issue looking for the beneficial results. In principle, we are to some degree aware of some important mappings. In both cases or scenarios, IBM may affect the speaker to look for robust specific correspondences (as evidence) on demand.

ACKNOWLEDGMENTS: This work is part of the Research Project Fondecyt 1170492.

REFERENCES

- Blackemore, D. (1992). *Understanding utterances*. Oxford: Blackwell.
- Evans, J. (2010). *Thinking Twice. Two Minds in One Brain*. New York: Oxford University Press.
- Ervas, F. & Ojha, A. (2019). Metaphor in argument production vs. Understanding. In B. Garssen, D. Godden, G. Mitchell & J. Wagemans (eds.), *Proceedings of the Ninth Conference of the International Society for the Study of Argumentation* (pp. 330-341). Amsterdam: Sic Sat.
- Gabbay, D. & Woods, J. (2005). *The Reach of Abduction. Insight and Trial. A Practical Logic of Cognitive Systems. Volume 2*. Amsterdam: Elsevier.
- Gibbs, R. & Ferreira, L. (2011). Do people infer the entailments of conceptual metaphors during verbal metaphors understanding? In M. Brdar, S. Gries & M. Zic. (eds.), *Cognitive Linguistics. Convergence and Expansion* (pp. 221-236). Amsterdam: John Benjamins.
- Harman, G. (1965). The Inference to the Best Explanation. *The Philosophical Review*, 74 (1), 88-95.
- Johnson, M. (2017). *Embodied Mind, Meaning, and Reason. How Our Bodies Give Rise to Understanding*. Chicago: The University of Chicago Press.
- Lakoff, G. (2006). *Thinking Points. Communicating Our American Values and Vision*. New York: Farrar, Straus and Giroux.
- Lipton, P. (2004). *Inference to the Best Explanation*. London: Routledge.
- Macagno, F. & Zavatta, B. (2014). Reconstructing metaphorical meaning. *Argumentation*, 28 (4), 453-488.
- Oswald, S. & Rihs, A. (2014). Metaphor as argument: Rhetorical and epistemic advantages of extended metaphors. *Argumentation*, 28 (2), 133-159.
- Santibáñez, C. (2010). Metaphors and argumentation: The case of Chilean parliamentary media participation. *Journal of Pragmatics*, 42 (4), 973-989.
- Schurz, G. (2007). Patterns of Abduction. *Synthese*, 164 (2), 201-234.
- Steen, G., Reijnders, G. & Burgers, C. (2014). When do natural language metaphors influence reasoning? A follow-up study to Thibodeau and Boroditsky (2013). *PLoS ONE*, 9, e113536.
- Thagard, P. (1978). The Best Explanation: Criteria for Theory Choice. *The Journal of Philosophy*, 75 (2), 76-92.
- Thagard, P. (1981). Peirce on Hypothesis and Abduction. In K. Ketner et al. (eds.), *Proceedings of the C.S. Peirce Bicentennial International Congress* (pp. 271-274). Lubbock: Texas Tech Press.
- Thagard, P. (1988). *Computational Philosophy of Science*. Cambridge: The MIT Press.
- Thibodeau, P. & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PLoS ONE*, 6, e16782.
- Thibodeau, P. & Boroditsky, L. (2013). Natural language metaphors covertly influence reasoning. *PLoS ONE*, 8, e52961.
- Wagemans, J. (2016). Analyzing metaphor in argumentative discourse. *RIFL*, 2, 79-94.
- Walton, D (2004). *Abductive Reasoning*. Tuscaloosa: The University of Alabama Press.
- Walton, D. & Hyra, C. (2018). Analogical Arguments in Persuasive and Deliberative Contexts. *Informal Logic*, 38 (2), 231-262.

Various Efforts of Enhancing Real World Online Discussions

ALEXANDER SCHNEIDER

Heinrich-Heine-University Düsseldorf, Germany
alexander.schneider@hhu.de

CHRISTIAN METER

Heinrich-Heine-University Düsseldorf, Germany
christian.meter@hhu.de

In this work we present a suite of software which enables gathering of natural language arguments from non-expert users of argumentation software without the use of NLP or other argument mining techniques. This is achieved by presenting the user with interfaces that prompt them to enter the data in a way in which it can be correctly added to an argument graph.

KEYWORDS: argumentation, online discussions, dialog-based discussions, web applications

1. INTRODUCTION

In this work we present various efforts that try to answer the question of how to gather structured argumentation graphs from natural language discussions of non-expert users.

Gathering arguments through argument mining from natural language is an ongoing research effort that made a lot of progress in the last years. Despite this, considerable challenges need to be solved before argument mining is at its peak. Because of this we present different ways of gathering argument data from natural language discussions.

We tackle the problem by designing interfaces and systems which allow the user to input arguments, while the data is automatically structured into an argument graph in the background. We made several efforts to design dialog-systems which make use of this approach to interact with everyday users that are not argumentation-experts in any way.

We should shut down University Park because criminals use University Park to sell drugs.

Antonia agrees that criminals use University Park to sell drugs. But she does not believe that this is a good argument for we should shut down University Park. She thinks that we should not give in to criminals.

What do you think about that?

- ☐ Wrong, her argument is false.
- ☐ Right, her argument is true.
- ☐ Right, her argument is true. But I do not believe that this is a counter-argument for my argument.
- ☐ Right, her argument is true and I do accept that this is a counter-argument for my argument. However, I have a much stronger argument for my point of view.
- ☐ Show me another argument.

Figure 1: Gathering feedback during a confrontation in D-BAS.

A typical user is presented with an argument and the request to react to that argument (see Figure 1). Participating users can then position themselves to that argument using statements introduced by other participants, thus strengthening the existing graph-structure or enter their own opinion. In that case the interface prompts them to input their argument in such a way that structured argument data is produced without further processing. This can be done by adding the new statement in the proper place in the argumentation graph, which the system can deduct from the selected choices of the user.

In this paper we describe three such interfaces, namely our *Dialog-Based Argumentation System* (D-BAS) (Krauthoff et al., 2018), *discuss* (Meter et al., 2017), and *Jebediah* (Meter et al., 2018). The interfaces differ in their approaches and use-cases. While D-BAS is a dedicated webservice for discussions which the user needs to visit, *discuss* allows the embedding of the interface into arbitrary websites. *Jebediah* enhances user experience by providing an agent for social networks with support for natural language processing. All these approaches share the same argumentation engine in their backend, which is accessible via D-BAS' *Application Programming Interface* (API) in the reference implementations.

The structured data created by the interfaces lends itself to reuse, and as a consequence we also present our *Extensible Discussion Entity Network* (EDEN) (Meter et al., 2018). EDEN is a reference implementation, which be used by discussion-providers to perform an automatic exchange of argumentation data. Examples of exchanged data are statements and arguments from the users, which can then be re-used in further discussions. We show that (automatic) reuse of argument data is possible and valuable.

As a last step in our pipeline we also provide a tool called *dabasco* (Neugebauer, 2018), which enables the transformation of the gathered data into instances of *Argumentation Frameworks* (AF) (Dung,

1995), *Abstract Dialectical Framework* (ADF) (Brewka & Woltran, 2010) and ASPIC+ (Caminada & Amgoud, 2007).

Thus, we present a complete pipeline of software projects which aid in the creation of natural language online discussions for non-expert internet-users, resulting in structured argumentation graphs that can be further used for analysis and other relevant processes. We reason that the pipeline presented in this paper is viable in conducting large-scale online discussions.

The rest of the paper is structured as follows: Section 2 describes dialog-based argumentation in general and D-BAS in particular. Following, Section 3 introduces the reuse of arguments and an implementation for networking several dialog-based argumentation systems. In Section 4 an alternative interface for integration of dialog-based argumentation systems into arbitrary web content is discussed. A social-agent-based interface and miscellaneous ways of exporting the collected data into other discussion frameworks are presented in Section 5. In closing, we discuss related work in Section 6 and end with our conclusions and future work in Section 7.

2. DIALOG-BASED ONLINE ARGUMENTATION

A lot of research in the argumentation community focuses on argument mining from natural language texts. Most argument mining research is done with the goal of creating a machine understandable corpus of arguments, which can be processed and used by algorithms. With that same goal in mind, we want to present a different approach. Instead of letting human users debate with free text, e.g. in forums, and trying to mine the arguments after the fact, we want to engage them in a dialog-like exchange. This exchange still lets the users use natural language but presents them with certain prompts at the same time. This compels the user to enter their thoughts in a structured manner, yielding arguments which can be added to an argumentation graph instantly.

2.1 *The Idea Behind Dialog-Based Online Argumentation*

Dialog-based argumentation was introduced in detail by Krauthoff et al. (2016, 2018) and is best described as a multi-user dialog with a single system. Each user is confronted with an argument for some topic, that was not generated by the system but was entered by other users. Therefore, the user is basically engaged in a time-shifted dialog with other users. The main difference to “traditional” online discussions like forums is that the user is at all times being presented with a single argument, instead of e.g. a list. After the user reacts to the presented argument, a next argument made by other participants is chosen based

on the user's reaction. The reaction is then stored to be used in future interactions with the system.

Let us take a look at an example: The system contains a discussion with the topic "We should renovate the city's library". Now, the system could present the interested user with several options, which confront the user with arguments *in favor of* renovating the city's library or with arguments *against* renovating the library because, for example, it costs too much money. The user in turn can react to those arguments by either choosing counter- and supporting arguments that other users already made, and the user feels are compelling, or by entering their own thoughts. This step is the crucial one which prompts the user to enter their argument in a structured manner as presented in Figure 4. Since the user is guided through a specially crafted menu, the system knows whether to input the user's statement as an attack or support on a certain other statement, or if it is e.g. an undercut for some argument.

2.2 User-Focused Measures

The type of argument gathering, that we present with dialog-based discussion, relies heavily on the correct use of the system by the users. This leads us to focus on interface measures, which help the participants to navigate the system without issues.

Let us say a user is interested in the topic of whether to buy a dog or a cat. After the user expresses their interest in the topic, the system asks the user about what they want to debate in detail. Those options are for example "We should get a dog", "We should get a cat" or "We should get another pet". When the user selects the position they are interested in, they are prompted to state whether they are in favor or opposed to that option (or have no opinion but want to see some arguments for that option). This is done, so the system knows whether the user interactions to come should be tallied as attacks or supports of certain arguments. Furthermore, it enables the system to confront the user with fitting arguments from its database.

Anytime the user formulates their own arguments instead of reusing others, the system scans for similar arguments already made and presents them to the user. They can then choose to use one of the already present arguments to keep duplicates to a minimum. The dialog continues until the user does not want to have a discussion anymore, or until they reach a point in the discussion graph where there is no more attacking or supporting arguments left.

Duplicate, malicious or grammatically unsound arguments still make it into the system, since its main input source are typical humans.

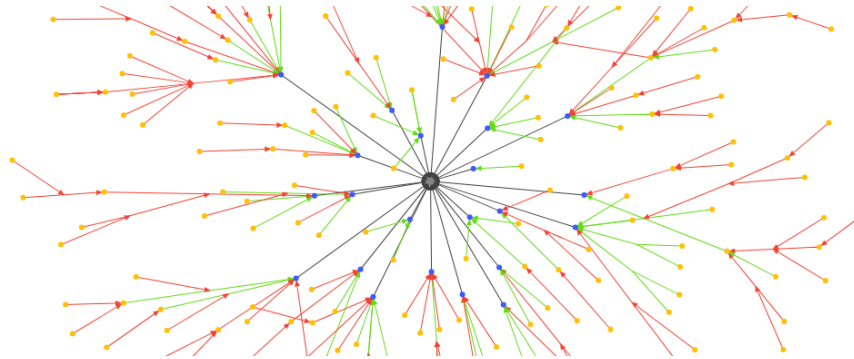


Figure 2: The graph resulting from discussions through D-BAS, discuss and Jebediah. Depicted is an instance from a real-world discussion. Colors: grey: issue, blue: positions, yellow: statements, green: supports, red: attacks.

Those arguments can be moderated to make the experience a pleasant and engaging one for the users. Instead of using traditional moderators, the system implements the power of the masses. This has been included in D-BAS as a *decentralized moderation system* (Krauthoff et al., 2018). Users can e.g. mark duplicates or arguments violating the community's policies. Experienced users can then visit special randomized moderation queues, where they are presented with some of the marked arguments and can democratically vote whether to take action against those. Possible actions are for example "delete argument", "reformat argument" or "merge duplicates". If enough votes are tallied for a single option, it is executed.

2.3 Field Experiences

The dialog-based argumentation system D-BAS is online and free to use¹. Besides experiences gathered from running the service, there also have been lessons learned from a formal evaluation through a field-study (Krauthoff et al., 2017). The study took place over 19 days and had 318 unique participants that visited the corresponding website. In this study the topic was how the computer science faculty could improve the bachelor's courses despite student numbers growing rapidly. All computer science students were invited to participate, and the faculty promised to use the results as a base for future decisions.

During the experiment, more than 250 arguments have been created, which seems to suggest that users untrained in argumentation techniques are able to create a complex argument graph with the help of

¹ <https://dbas.cs.hhu.de>

dialog-based argumentation. Parts of the resulting graph can be seen in Figure 2 and the associated data can be obtained online.

2.4 Application Programming Interfaces

D-BAS has two fully documented² and usable API options built-in to export the contents of a discussion and to allow third party applications to access the *Dialogue Game Execution Platform* (DGEP) (Bex et al., 2014) parts.

The first endpoint provides authentication, authorization and the execution of discrete steps in the discussion. Applications can send requests to this endpoint to tell D-BAS about their current status of the discussion which then produces a response containing the next options and possible next discussion actions. Also sample text-responses are returned, which can then be used.

Data retrieval from our databases can be achieved using the second endpoint, which provides a GraphQL (The GraphQL Foundation, 2019) API. This way people interested in the data can write their own queries to our databases to retrieve the public information from the hosted discussions.

3. NETWORKED ARGUMENTS AS A RESOURCE

Through the use of dialog-based argumentation, people are able to create a wealth of arguments by following a dialog. But there are also scenarios where D-BAS has disadvantages. If we assume that, for example, several media outlets use dialog-based argumentation instead of simple list-like comments under their publications, each of them could run their own instances of dialog-based argumentation software. Now, every user that wants to debate the same or a similar topic at different media outlets, is confronted with repeating arguments they are already familiar with. This would almost certainly happen due to the nature of how dialog-based argumentation is conducted. Furthermore, arguments made at one instance will never be seen on another, no matter how insightful or well worked out they may be. This section presents our thoughts on how to tackle these and related challenges.

3.1 Distributing and Versioning Arguments

We call every host, from the before-mentioned scenario, running their own dialog-based argumentation software, an *aggregator*. To put it in another way: an aggregator is an entity providing content and the space

² <https://dbas.cs.hhu.de/docs>

to discuss it. To allow distribution of arguments, every aggregator can join a distribution network. Aggregators may have differing policies about which arguments are valid according to some rules or community standards. Hence, flooding the arguments to all aggregators in the network is unwise, because not all instances have the same policies and would be willing to receive certain arguments. Moreover, aggregators possibly want to keep the intellectual rights on arguments devised on their platform. Thus, every argument needs to reference which aggregator is the authoritative instance for it. This means, that the arguments stay property of the differing aggregators, but still can comprise a single argumentation graph spanning over different physical and logical entities participating in the argument network. To allow other participants to propose changes to arguments, that they are not authoritative of, we need to introduce versioning. As presented by Meter, Schneider and Mauve (2018) one can use a decentralized version-tree which is already known for versioning source-code. This means, that every argument has a pointer to its predecessor if one exists. Any changes can be proposed at once without violating or changing the original argument by creating a changed version which points to the original as its predecessor. The authoritative aggregator can decide whether to accept any of the proposed updates and incorporate them into the official version. But even in that case, there will be a new version from the authoritative source, since all arguments are created immutable.

3.2 EDEN: Extensible Discussion Entity Network

An exemplary implementation of a distributed argumentation network powered by aggregators is EDEN which was presented in detail by Meter et al. (2018). EDEN was developed in Clojure, a functional language on the JVM. Furthermore, we pursued a modular approach with EDENs architecture, which splits it up into four distinct modules – interface, discussion platform, database and aggregator core – which can be interchanged as long as the new module adheres to the proposed interfaces between the major parts.

The interface is tasked with guiding the user through the dialog-based argumentation. A database stores and persists the locally needed arguments. It can also provide features like semantic search on the arguments. The discussion platform is the piece of software that provides the internal logic on how to conduct the dialog-based argumentation, also known as DGEP. In the default case EDEN utilizes D-BAS as a DGEP. An aggregator core coordinates the flow of arguments between the different modules as well as between aggregators.

Communication between aggregators is handled in two parts. First, there is a REST API providing aggregators with the ability to actively query for discussion entities like arguments and their interrelations. As a second option a publish/subscribe queue exists, which automatically updates entities from known aggregators. For example, if aggregator *B* requests some argument *X* on the topic of dogs from aggregator *A*, they also subscribe to the corresponding queues. When an update for *X* is available, *B* automatically gets informed about the update by *A* via the queue. Different update forms can be used. Instead of updates on queried arguments, *B* could receive notifications every time there is a new argument on the topic of dogs, to broaden its repertoire

4. DISCUSS: EMBEDDING DIALOG-BASED ARGUMENTATION INTO WEB-CONTEXTS

One of the first applications using the API of D-BAS, is *discuss* (Meter et al., 2017). *discuss* provides a minimal discussion interface to interact in the same flow as we have seen it in D-BAS, with the distinction, that it can be embedded in every web-context utilizing a JavaScript environment. This is intended to be used, for example, in online newspaper articles, which ask the readers to start a discussion in the comment sections. But since comment sections do not provide any structure, this approach could bring a significant improvement, because of the structural manner how the arguments of the users are being gathered.

Create an argument with a text reference

You can refer position to a text passage here. Please fill in the following fields.

I think that...

... that

... because

“ ” ✕

10 CHARACTERS REMAINING

Home
Create Argument
My Arguments
Options
Logout

Figure 3: discuss: Create a new argument with a reference to a passage in the author’s article.

Without having to leave the current scope, discuss provides (1) direct interaction with the author’s arguments, (2) jumping into the discussions, where other participants interacted with the article, (3) enabling discussions in our proposed dialog-based flow (see Subsection 2.1) and (4) connect to the EDEN network.

4.1 Interacting with the Author’s Arguments

One of the core functions of discuss is to directly interact with the author’s article. Selecting an interesting part of a text passage opens up a dialog, where the reader can create a new argument with the selected text as a *reference* (see Figure 3). Internally, the creation of an argument in this way is the same procedure as adding a new position in D-BAS, which introduces a sub discussion in the context of the discussion topic.

4.2 Jumping into the Discussion

Interactions with the article, which created a new argument with a reference to parts of the article, are highlighted so that the user sees an interactive element on the website (see Figure 4). These references provide an entrypoint to the discussion, where the user’s argument has been used. Also, other arguments, which referenced the same text passages, are listed and users can decide where they want to jump into the discussion.

Currently, the city council discusses to close the University Park, because of its high running expenses of about \$100.000 per year. But apparently there is an anonymous investor ensuring to pay the running costs for at least the next five years 🗨️. Thanks to this anonymous person, the city does not lose a beautiful park, but this again fires up the discussion about possible savings for the future.

Figure 4: Text passage from an article, which has been used in an argument. A click on it opens the interface to jump into the discussion

4.3 Dialog-Based Discussion Flow

We omit the selection of the initial positions in discuss, because we encourage to directly jump into the discussion via a reference in the text, i.e. hook into a pre-existing argument from a user, or by selecting a text-passage, i.e. create a new argument referring to the text. After the initial step, discuss presents the classical discussion flow which we have already seen in D-BAS (see Subsection 2.1). Specifically, this means that we conduct a dialog with the users and present those arguments, which have been posted about the argument from the article.

4.4 EDEN Integration

Besides the described functions, discuss can be used to connect to the EDEN network (see Subsection 3.2). D-BAS is then solely used as an DGEP for the steps in the discussions, whereas the arguments are being fetched from EDEN. This mechanism allows to retrieve and collect arguments from different locations and discussions, which can then be used in the current article's discussion.

5. EXPERIENCES WITH AUXILIARY APPROACHES

Based on the presented tools, we felt the need for auxiliary applications. One is *Jebediah*, an alternative interface into dialog-based online discussions enabling users to discuss matters through chatbots and voice assistants. Furthermore, we present *dabasco*, which allows the data generated through D-BAS and its applications to be converted to other discussion frameworks for further use.

5.1 Jebediah

A vast part of online discussions takes place on social media platforms. Jebediah (Meter et al., 2018) is an interface which enables users of those platforms to take part in dialog-based online argumentation through chat-bots and voice assistants. Classifying the user's input is realized

with the help of Google’s Dialogflow platform (Google Ireland Limited, 2019), which is an Artificial Intelligence processor that tries to match the natural language input against predefined and pre-trained rules. The matching-process has the goal to produce structured data and the resulting data is being sent to a dialog-based argumentation software, like D-BAS. It returns a response, which is then again formatted and forwarded to the user through the chat-bot (see Figure 5). This is still a highly experimental feature, which works most of the time but certainly can be further improved upon. Nonetheless, it would be interesting future work to test how users feel when discussing topics with a bot instead of a text-interface.

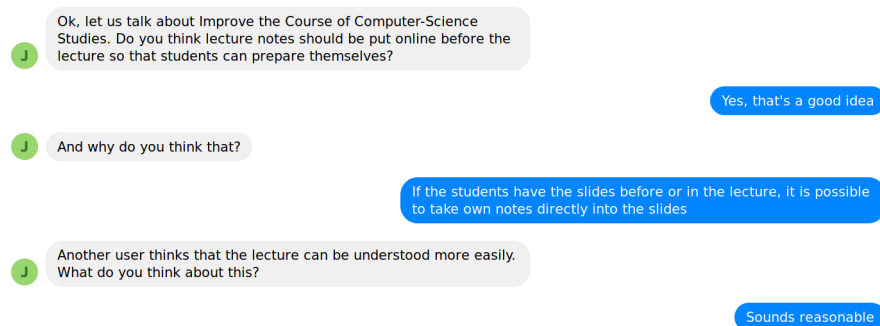


Figure 5: Left side: Dynamically produced text messages from Jebediah, right side the user’s answers in the Facebook Messenger.

5.2 *dabasco*

The last step in our pipeline is the export of the generated data. Exports are useful to utilize collected argument data for further analysis. Building on the fact that some established tools by the community expect certain formats, Neugebauer developed an export interface called *dabasco* (Neugebauer, 2018). This way it is possible to export AF, ADF, and ASPIC+ data which was converted from D-BAS’ data structure. *dabasco* uses D-BAS’ API and provides the first 3rd party application interacting with our software stack.

6. RELATED WORK

Tools for facilitating online argumentation have been described and developed before. The set of tools that is most like the proposed pipeline is the argument web (Rahwan et al., 2007). We build on similar ideas of a unified structured web of arguments and are not striving to compete with the argument web but to be compatible to magnify the extend of the argument network. AIFdb, developed by Lawrence et al.

(2012), is in spirit akin to EDEN regarding collecting arguments from differing sources, but differs in aspects of centralization and the kind of arguments collected. Other approaches at structuring arguments, include Carneades (Gordon & Walton, 2006), Deliberatorium (Klein & Landoli, 2008) or OVA as introduced by Snaith et al. (2010). The difference to is that none of those are based on dialog-like argumentation. Most of these tools focus on the whole discussion, whereas our smallest entity is the statement, which could be put together to an argument and the put into context, e.g. of a discussion.

7. CONCLUSION

In this paper we presented a complete pipeline for gathering, sharing and exporting user- generated arguments. We introduced D-BAS, a system that conducts discussions by simulating a dialog with other users. A field-study verified that this approach yields a structured argumentation graph and even untrained users were able to use our software in a productive way. Moreover, we presented discuss, which enables arbitrary websites to integrate a D-BAS-style discussion and Jebediah, which does the same for artificial assistants. To share the generated arguments between instances of D-BAS, we use EDEN, which provides the ability to decentralize an argumentation network. Lastly, dabasco allows the export of D-BAS arguments to different argumentation frameworks, which can be used for further calculations.

This paper showed that a pipeline for gathering structured argumentation from natural language without argument mining is possible and how such a pipeline may be structured.

For future work we plan to conduct field experiments that make use of the complete pipeline to test its efficiency. We furthermore are developing tools that harness the dialog-based stack to conduct discussions with the goal of finding and voting on solutions for e.g. the budgetary allocation of a city.

REFERENCES

- Bex, F., Lawrence, J., & Reed, C. (2014, January). Generalising argument dialogue with the Dialogue Game Execution Platform. *Frontiers in Artificial Intelligence and Applications* 266, pp. 141-152.
- Brewka, G., & Woltran, S. (2010). Abstract Dialectical Frameworks. *Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning*.
- Caminada, M., & Amgoud, L. (2007). On the evaluation of argumentation formalisms. *Artificial Intelligence*, vol. 171, pp. 286-310.

- Dung, P. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, vol 77, no. 2, pp. 321-357.
- Google Ireland Limited. (2019, June). Dialogflow. Retrieved from <https://dialogflow.com/>
- Gordon, T., & Walton, D. (2006). The Carneades Argumentation Framework - Using Presumptions and Exceptions to Model Critical Questions. *Frontiers in Artificial Intelligence and Applications*, vol. 144: Computational Models of Argument, pp. 195-207.
- Klein, M., & Iandoli, L. (2008, February 20). Supporting Collaborative Deliberation Using a Large-Scale Argumentation System: The MIT Collaboratorium. MIT Sloan Research Paper No. 4691-08.
- Krauthoff, T., Baurmann, M., Betz, G., & Mauve, M. (2016). Dialog-Based Online Argumentation. *Computational Models of Argument*, pp. 33-40.
- Krauthoff, T., Meter, C., & Mauve, M. (2017). Dialog-Based Online Argumentation: Findings from a Field Experiment. *Proceedings of the 1st Workshop on Advances in Argumentation in Artificial Intelligence*, vol. 2012, CEUR-WS, pp. 85-99.
- Krauthoff, T., Meter, C., Betz, G., Baurmann, M., & Mauve, M. (2018, September). D-BAS – A Dialog-Based Online Argumentation System. *Computational Models of Argument*, pp. 325-336.
- Lawrence, J., Bex, F., Reed, C., & Snaith, M. (2012). AIFdb: Infrastructure for the Argument Web. *Frontiers in Artificial Intelligence and Applications*, vol. 144: Computational Models of Argument, pp. 515-516.
- Meter, C., Ebbinghaus, B., & Mauve, M. (2018, September). Jebediah – Arguing With a Social Bot. *Computational Models of Argument*, pp. 467-468.
- Meter, C., Krauthoff, T., & Mauve, M. (2017, July). discuss: Embedding dialog-based Discussions into Websites. *Human Computer Interaction International: Proceedings of HCI*, pp. 449-460.
- Meter, C., Schneider, A., & Mauve, M. (2018, September). EDEN: Extensible Discussion Entity Network. *Computational Models of Argument*, pp. 257-268.
- Neugebauer, D. (2018, September). DABASCO: Generating AF, ADF, and ASPIC instances from Real-World Discussions. *Proceedings of the 2018 Conference on Computational Models of Argument*, pp. 469-470.
- Rahwan, I., Zablith, F., & Reed, C. (2007). Laying the foundations for a World Wide Argument Web. *Artificial Intelligence*, vol. 171, no. 10-15, pp. 897-921.
- Snaith, M., Devereux, J., Lawrence, J., & Reed, C. (2010, August). Pipelining Argumentation Technologies. *Proceedings of the 2010 conference on Computational Models of Argument*, pp. 447-453.
- The GraphQL Foundation. (2019, June). GraphQL - A query language for your API. Retrieved from <https://graphql.org/>

Beyond Randomized Clinical Trials: Emerging Innovations in Reasoning About Health

JODI SCHNEIDER

School of Information Sciences, University of Illinois at Urbana-Champaign, USA
jodi@illinois.edu

SALLY JACKSON

Department of Communication, University of Illinois at Urbana-Champaign, USA
sallyj@illinois.edu

Specialized fields may at any time invent new inference rules—that is, new warrants—to improve on their stock of resources for drawing and defending conclusions. One such invented warrant, Randomized Clinical Trial, is widely regarded as the gold standard for making inferences about causal relationships between medical treatments and patient outcomes. Tensions that arise from the competing perspectives of scientists, clinicians, and patients have recently led to reconsideration of RCT and to emergence of alternative research strategies, notably ‘pragmatic trials’ and ‘N-of-1 trials’.

KEYWORDS: field-specific reasoning, medical reasoning, N-of-1 Clinical Trial, Pragmatic Clinical Trial, Randomized Clinical Trial, warrants, warrant-establishing arguments

1. INTRODUCTION

Toulmin (1958) pointed out the possibility that specialized fields may at any time invent new inference rules—that is, new warrants—to improve on their stock of resources for drawing and defending conclusions. This appears to be happening at a very rapid pace in the field of health science, where several waves of innovation have occurred over the past century or more. Jackson and Schneider (2018) analyzed one recent innovation, a form of evidence aggregation known as a Cochrane Review. Although Cochrane Review functions argumentatively as a generalized warrant, it has special features not normally attached to warrants, including technical components invented specifically to support the use of the

warrant in reasoning within the field. We introduced the term “warranting devices” for a class of such innovations that involve an inference rule packaged with its technical components in such a way that any use of the rule includes tacit assurance that it generates dependable conclusions. A warranting device, then, is a specialized inference rule, invented within a field for some particular argumentative purpose, and backed by a set of assurances that may be partly material, partly procedural, and even partly institutional.

In subsequent work, Schneider and Jackson (2018b) examined another warranting device known as the Randomized Clinical Trial (RCT), widely regarded as the gold standard for making inferences about causal relationships between medical treatments and patient outcomes. Still controversial through the early twentieth century, RCT achieved broad acceptance within the field as a result of warrant-establishing arguments circulating in the medical literature starting in the 1950s (Schneider & Jackson, 2018a). In this paper, we examine several less well-established movements within health science (notably ‘pragmatic trials’ and ‘N-of-1 trials’) that seek to go beyond RCT as a basis for reasoning about treatments. We consider how early decisions about the design of the warranting device (notably, a focus on group averages as central to inference about cause and effect) brought about undeniable improvements in reasoning while also sowing seeds for later dissatisfaction with how results were translated into clinical practice.

Although any proposed warranting device may be established through successful demonstration that it can produce dependable conclusions, these devices are by their very nature changeable, either wholly or in part. A device may become stabilized within the reasoning practices of a field at one point in time, then de-stabilized at a later point in time, because new vulnerabilities in the device are discovered, because some new device pushes an older one toward obsolescence, or because the arguments generated by the device meet new forms of criticism in new discourse contexts. Warrant-establishing argument is never completely conclusive; disagreement over the acceptability of an invented warrant can always be re-opened.

In this paper, we explore the arguments that have helped to re-open debate over RCT, exploring the tensions that arise from the competing perspectives of scientists, clinicians, and patients.

2. CLINICAL TRIALS

Clinical trials have become, or are quickly becoming, a worldwide standard for generating evidence of the effects of proposed treatments. The feasibility of clinical trials depends on material and institutional resources. For instance, they are affected by the health care systems in a

locality (e.g., the logistics of recruiting patients and managing a controlled administration of treatments). They may also be subject to different restrictions in different national jurisdictions. Both the feasibility of clinical trials and the quality of evidence resulting from them can be affected by societal conditions that are outside the control of scientists, requiring well-organized efforts to create conditions more supportive of clinical trials (e.g., for the EU, making cross-national recruitment feasible; see Demotes-Mainard, & Kubiak, 2011). Clinical trials have economic value, and globalization of the practice is thus partly driven by the pharmaceutical industry (Thiers, Sinskey, & Berndt, 2008).

While the basic logic of clinical trials is global in reach, national or regional institutional context is important to understanding how clinical trialing as a practice has developed. In the US, clinical trials often depend upon hospitals that have a research mission, especially university hospitals. Funds needed for independent (non-industry) research are controlled by the US National Institutes of Health. The US Federal Drug Administration governs approval of new pharmaceuticals and has played a significant role in the institutionalization of the phase structure of clinical trialing (described below). Finally, journal editors can exert pressure on how research is conducted by limiting publication opportunities based on compliance with both scientific and ethical standards (as in De Angelis et al., 2004, and Taichman et al., 2017). These institutional actors, pushing toward their own goals, and sometimes pushing back against one another, have shaped the structure of clinical trials in the US.

Clinical trials are defined by the NIH as research studies “in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes” (U.S. National Institutes of Health, 2014). They are experiments on human subjects whose independent variables are potential treatments and whose dependent variables are aspects of health or well-being. The logic of clinical trials is apparent from the diagram in Figure 1, showing random assignment of a large number of patients to contrasting forms of treatment. Inferences about whether and how the treatments differ in effects are delegated to tests of statistical significance and quantitative measures of effect size.

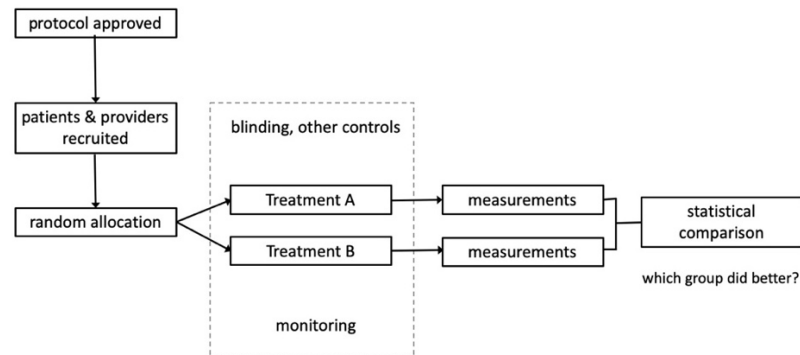


Figure 1 – A simple Randomized Clinical Trial (Schneider & Jackson, 2018a).

Conducting experimental research on human subjects in *phases* allows for (and often requires) evaluation of the safety of a treatment for healthy patients prior to evaluation of the efficacy of the treatment for sick patients. In the highly regulated world of pharmaceuticals, conducting research in phases has become institutionalized to such an extent as to permit explicit codification by the US Federal Drug Administration (Office of the Commissioner, 2019). Clinical research (on human subjects) is expected to begin with a demonstration that the drug can be safely given to humans. For new drugs, they must first have been tested on non-human animals (Center for Drug Evaluation and Research, 2019). Phase I trials recruit healthy subjects, typically not very many, and may involve such design features as dose escalation over the course of the trial. Assuming that a safe dose level is found, the drug may be used in a Phase II trial on volunteers from the relevant patient population, allowing not only continued assessment of safety for this patient population but also assessment of efficacy. Phase III trials are larger in size (number of patients) and longer in duration, to allow for ‘small’ effects to be detected, especially any adverse effects that may not be noticed in a smaller sample or over a shorter period of time. A drug that assembles evidence of safety and efficacy over these three phases is a good candidate for FDA approval. From an argumentative perspective, FDA approval encapsulates claims of safety and *potential* benefit for patients to whom the drug may be administered. Post-approval clinical trials are known in the FDA world as Phase IV trials and have aims similar to Phase III.

The distinction between Phase I and the other phases is particularly significant for drug treatments and certain other interventions: Phase I trials recruit healthy patients, not those with the condition that the intervention is expected to treat, so they do not normally provide much reason to believe that an intervention will be beneficial for treating that condition. From an argumentative

perspective, promising results from a Phase I trials do not even provide evidence that the treatment is *safe* for patients with that condition. They do, however, provide evidence on safety that can allow prospective volunteers for Phase II trials to consider their own risk realistically. Phased trialing adds considerable nuance to what claims are supported by RCTs; an RCT may establish a safe dose level, or a difference between one treatment and another, or a certain “success” rate in patient care, and although all of these are labelled ‘effects’ of treatments in their respective studies, they are not all the same. It takes a very long time to get through all of the work of a phased trial sequence, and at any point in time, the kind of claims that are actually warranted vary by which phase has or has not been passed.

Understanding how thoroughly intertwined clinical trials are with institutional context is very important to understanding how they can warrant inferences about treatment effects—and especially important for understanding why publics push back against them when these inferences become part of reasoning about actual treatment of actual patients.

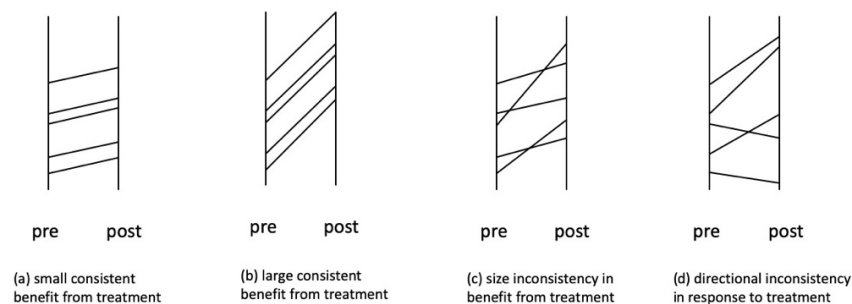


Figure 2 – Treatment effect at individual level expressed as change from pre-test to post-test, with each line representing one patient. Lines sloped upward represent benefit from treatment; lines sloped downwards represent worsened condition after treatment.

Clinical trials answer very well-defined questions that are relevant and important to decisions about how to treat patients, but practitioners and researchers alike know that successful results in Phases I to III do not assure that the drug or other treatment evaluated will be successful for all patients under all conditions. In fact, many of these sequences produce evidence that treatment effects vary widely from patient to patient. An average benefit may appear when some patients benefit while others do not, or when some benefit while others are actually harmed. Statistically, this situation is known as ‘person by treatment interaction,’ but it is not always visible (or estimable) within a

standard clinical trial. To estimate person by treatment interaction, a researcher must observe what happens to each individual, both with and without the proposed treatment (for example, in a pretest/posttest design). In Figure 2, each panel shows the effect of treatment as change from a pretest measurement (without the treatment) to a posttest measurement (with the treatment). Each patient's pretest and posttest measurements are represented by a line drawn between two vertical axes. Lines with positive slope (rising to the right) represent patients who benefitted; lines with negative slope (falling to the right) represent patients who did worse after being treated. Variability in the slopes of the lines represent person by treatment interaction. All of these cases produce an average benefit; all four configurations could produce a statistically significant benefit for the treatment (especially if only a small proportion of patients do worse with the treatment than without). For patients and their care providers, this means that a treatment that is beneficial on average may or may not be beneficial for any one individual. Likewise, when one treatment is shown to be better than another on average, it may still be true that the "less effective" treatment is best for some patients.

So even after Phase III, there can remain a large gap between what is established through this trial sequence and what a reasonable physician or patient would want to know before choosing to administer or to accept the treatment, either as a standard option or—especially—as a specific choice for an individual patient. As rules and preferences are imposed over time by funders, regulators, and publishing gatekeepers, this gap can widen—or narrow.

But the gap has become more noticeable over time. An under-appreciated fact is that the "logic" behind a particular innovation in inference, even when made quite explicit, cannot always be fully evaluated without applying it to the task of drawing conclusions. After the initial successful defense of RCT for drawing conclusions about medical treatments, there was great optimism about its potential and great momentum behind exploiting this potential. But as medical practice has become more infused with evidence from RCTs, what seemed like unproblematic reasoning has turned out to have unexpected limitations. Hundreds of Cochrane Reviews framed by practical questions about care locate *zero* papers reporting data worth aggregating.¹ Evidence worth aggregating based on each review's pre-specified criteria is not always

¹ Roughly 9% of Cochrane reviews are empty. This ratio seems relatively constant over time: As of August 15, 2010 Yaffe and colleagues (2012) found 8.7% empty reviews (376 of 4320 reviews), while we determined that as of January 3, 2018, 9.2% (659 of 7156) published Cochrane reviews in the Cochrane Library were empty.

forthcoming, either: some reviews remain empty for ten years or more, even after repeated attempts to locate relevant evidence (Yaffe, Montgomery, Hopewell & Shepard, 2012). This suggests gaps between the answers health care practitioners want and the evidence available for synthesis from RCTs and other methods. For example, what we know scientifically about possible treatments for a health condition is dependent in part on what it is allowable to study, in part on what is prioritized by funding sources, and in part on what scientists themselves find interesting. The lack of scientific evidence for something is often a direct consequence of institutional actors having no interest in it.

3. PRAGMATIC TRIALS

RCT “technology” might have developed quite differently than it actually has—which is to say that its core ideas could have been elaborated in multiple different ways. Bradford Hill’s defense of controlled clinical trials in the 1950s, analyzed by Schneider and Jackson (2018a), left many avenues of development open—not just the avenue that has resulted in NIH’s three (or four if counting post- approval Phase IV) distinct trial phases.

Early proponents of alternative technological directions included Daniel Schwartz and Joseph Lellouch, whose 1967 article titled “Explanatory and Pragmatic Attitudes in Therapeutic Trials” took decades to attract a large enough following to get ‘pragmatic trials’ broadly acknowledged as a fourth phase. Schwartz and Lellouch argued that a basic inferential strategy of comparing outcomes obtained with contrasting treatments could be undertaken with either a purely epistemic aim as in basic science (to explain something) or with a pragmatic, choice-oriented aim (evaluating a course of treatment or choosing a treatment policy). Both aims can be served by a standard experimental design (shown earlier in Figure 1): People are randomly allocated to one of two alternative treatments, Treatment A or Treatment B; measurements are taken (and statistically compared) on whatever physical or mental state Treatments A and B are expected to improve.

Schwartz and Lellouch pointed out that despite commonality of this structure, designing a trial to satisfy explanatory aims is very different from designing a trial to satisfy pragmatic aims. We will not review all of the nuance of their argument but simply summarize three issues that clearly differentiate explanatory and pragmatic aims: how to form comparison groups, how to conceptualize treatments, and how to select meaningful outcomes.

3.1 Comparison groups

That comparison groups should be formed at random from a common pool is not disputed by Schwartz and Lellouch. Their concerns are with how the common pool is developed, and with what happens when individuals from this common pool drop out after random assignment to a treatment. They argue that in such cases, statistical analysis may be conducted either on the premise that the dropouts are simply people for whom the treatment was unsuitable (that is, people who have nothing to tell us about the potential efficacy of the treatment), or on the premise that the treatment is problematic in some way (by virtue of failing for some of those it aims to benefit). As they put it, “in the first [explanatory] case the class of patient is defined to fit the predetermined treatments, while in the second [pragmatic] the treatments are defined to fit the predetermined class of patients” (p. 643).

3.2 Treatments

When two proposed treatments are to be compared, it will normally be the case that each considered individually is a complex assembly of components, including the form in which the treatment would most conveniently be administered, the time over which it would typically be administered, the setting in which it would ideally be administered, and much more. The explanatory attitude strives toward a contrast in which as many of these components as possible are equalized between the treatments to be compared, while a pragmatic attitude strives for a contrast between the optimal arrangement for each of the treatments. Conducting the comparison between two (artificially) equalized treatments invites the possibility that neither treatment works up to its potential. Conducting the comparison between two optimized treatments allows for all manner of confusion over *exactly what* makes the better of the two treatments better.

Suppose, for example, two different substances have been approved for treating a skin condition, one of which can only be successfully formulated as a gel and the other of which can be formulated either as a gel or a cream. In comparing the two clinically, an explanatory mentality would favor simply comparing the two treatments administered as gels, while a pragmatic attitude would compare the first treatment as gel with the preferred version of the second treatment (ability to deliver as cream being considered an actual advantage of the second treatment rather than a pesky confound). Comparing Treatment A (substance 1 in a gel) and Treatment B (substance 2 in a cream) looks, from an explanatory mentality, like a clear case of confounding two possible causes; from a pragmatic mentality, it looks like a

straightforward comparison of two actual treatments a patient might receive.

3.3 Outcomes

Schwartz and Lellouch point out that a pragmatic attitude prefers outcome measures that are close to what a patient and clinician are trying to accomplish with a course of treatment: a feeling of well-being, a remission of pain, a return to normal activity, an extension of life, or something similar. Some of these outcomes (death, for example) may be inconvenient or unethical in research, and others (anything involving patient self-assessment) have known validity problems. Explanatory clinical trials quite commonly use more convenient outcome measures that are known to correlate highly with the actual outcome of interest. For example, blood cholesterol levels are commonly used to assess preventive treatment for cardiovascular disease instead of tracking actual cardiovascular events such as heart attacks and strokes. The advantages of this kind of outcome measurement are obvious, but so are the limitations: A correlate of a disease may not be in any sense a cause of the disease, requiring (at some point) further evidence of effectiveness.

Schwartz and Lellouch were among the earliest to argue that explanatory trials would inevitably fall short of what would be needed to support clinical decision-making. Conclusions drawn from explanatory trials have superficial plausibility as means-end premises for practical reasoning about clinical decisions, as shown in Figure 3. Schwartz and Lellouch's arguments expose a serious threat to the validity of the conclusion: the means-end premise is plausible only if much too much is assumed about a demonstration of efficacy (specifically, that T's average efficacy justifies its use in every case, and that this is so irrespective of other possible treatments that may also be efficacious).

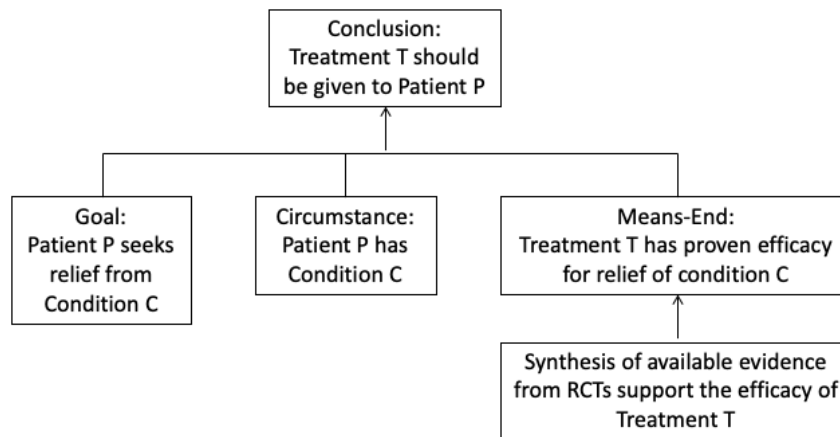


Figure 3 – Practical reasoning diagram, with Means-End premise drawn from explanatory RCTs.

Closing the gap between what is established by explanatory trials and what is needed for clinical care is a persistent theme in the many proposals for expanded use of pragmatic trials. A conservative approach to this is to simply add a Phase IV: Once efficacy is established through tightly controlled explanatory trials, go on to establishing effectiveness using more flexible and generalizable pragmatic trials. This is clearly not what Schwartz and Lellouch (1967) had in mind—but it is the obvious way to avoid starting over from scratch.

4. N-OF-1 TRIALS

N-of-1 trials, also called single patient trials, are RCTs that compare the effectiveness of two or more treatments on a single person. They were imported into medical science from experimental psychology (Guyatt, 2016). Their defining feature is that they produce meaningful conclusions for each individual patient; when repeated over many patients they may also support generalization, including generalizations about treatment variability of the kind shown earlier in Figure 2. Some advocates of evidence-based practice see N-of-1 trials as the highest form of evidence—as the top of an evidence pyramid of individual study designs (Guyatt, Rennie, Meade, Cook & American Medical Association, 2015, p. 11) or as one of the highest forms of evidence on treatment benefits and treatment harms, alongside systematic reviews (Howock et al., 2011).

A resurgence of interest in N-of-1 trials has been connected to their usefulness in clinical investigation (Guyatt, 2016), where N-of-1 trials offer potential benefits in comparison to other approaches. They are inexpensive compared to conventional RCTs enrolling many patients

(Shamseer et al., 2015, p. 43). They can provide timely results to each individual patient, and a series of similar N-of-1 trials can be aggregated to estimate population level effects (Nikles et al., 2011, p. 479).

A particular advantage of N-of-1 trials is their closeness and relevance to clinical care, “making research more like practice and practice more like research” (Kravitz et al., 2014, pp. 7–8). By contrast, there are multiple limitations in applying RCTs to routine clinical care. One challenge is in generalizing from research populations to patient populations: “Patients recruited into RCTs differ from those who are eligible but not recruited in terms of age, sex, race, severity of disease, educational status, social class, and place of residence” (Rothwell, 2005, p. 86). In the past, researchers had more freedom to restrict eligibility for what they thought of as design reasons, so an additional complication is that older research may be based on narrow categories of patients such as white men between 20 and 40 years old. Such arbitrary restrictions on eligibility conditions are now more carefully scrutinized by oversight agencies. The past literature base of RCTs is particularly likely to exclude women, the elderly, and patients with comorbidities (Rothwell, 2005). Another challenge, as noted earlier, is that an average benefit for a treatment is no guarantee of consistent benefit at the individual level. While RCTs provide population-level estimates of the efficacy, they do not indicate which course of treatment is best for a given patient.

Answering these challenges, N-of-1 trials give the most direct evidence possible for what works best for the individual patient—at least when it is in fact possible for all options to be tried by the same patient. Not every condition is suitable for comparative N-of-1 trials. They are best applied to chronic conditions that are relatively stable, where the treatment has a fast onset (and ideally a short half-life; Nikles et al., 2011, p. 473). As presently conceived, N-of-1 trials are not suitable for areas such as surgery, where an irreversible treatment may be given, or critical care/emergency medicine, where a patient being stabilized cannot serve as their own control but rather should be compared with other patients receiving a different treatment.

CONCLUSION

In our prior work we have focused on new inference methods—new ways to draw conclusions that are either better than old ways of drawing conclusions, or that allow us to draw entirely new kinds of conclusions. The central conceptual advance has been the idea of a warranting device—a proposed inference rule that generates conclusions whose quality is partly dependent on various kinds of assurances provided by the community that deploys the device. We are not prepared to say whether pragmatic trials and N-of-1 trials are new warranting devices,

mainly because the work of building out these assurances has not yet been done—as it has been for RCTs and for Cochrane Reviews.

From this study, we learn that these new inference methods will often have limitations that are exposed only in argumentative practice. The normal output of RCT is a carefully qualified claim about the average effect of a medical treatment when given to patients like those observed. Despite their obvious epistemic strengths, RCTs commonly provide evidence for conclusions that are still an inferential step away from the clinically relevant decision: whether a particular treatment should be given to a particular patient. Further inference is required, beyond what RCT itself warrants, to get to the claim that the treatment should be given to a particular patient. That gap does not become apparent until the scientific result moves from the upstream context of explanatory research to the downstream context of practical reasoning about health care.

Both pragmatic trials and N-of-1 trials aim to address this inferential gap. None of the arguments in favor of pragmatic trials or N-of-1 trials are arguments against RCT. On the contrary, both are infused with the spirit of experimenting and committed to extending RCT further and faster. But as may be intuitively clear, both of these innovations have potential to change the way we look at RCT.

ACKNOWLEDGEMENTS: Thanks to: the James Lind Library for providing useful reference material on the history of evidence-based medicine; Kiel Gilleade for help with images; Neil Smalheiser for drawing our attention to empty systematic reviews; Tzu-Kun ‘Esther’ Hsiao for analysis of empty reviews; Cochrane for providing bulk access to data; and Quynh Pham and Jessica Ancker for their AMIA 2016 preconference “Beyond the RCT: Evaluating Innovations in the Learning Health System.”

REFERENCES

- Center for Drug Evaluation and Research. (2019, August 9). Investigational New Drug (IND) Application. Retrieved August 28, 2019, from U.S. Food and Drug Administration website: <http://www.fda.gov/drugs/types-applications/investigational-new-drug-ind-application>
- De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A. J. P. M., Schroeder, T. V., Sox, H. C., & Weyden, M. B. V. D. (2004). Clinical trial registration: A statement from the International Committee of Medical Journal Editors. *New England Journal of Medicine*, 351(12), 1250–1251. <https://doi.org/10.1056/NEJMe048225>
- Demotes-Mainard, J., & Kubiak, C. (2011). A European perspective – the European clinical research infrastructures network. *Annals of Oncology*, 22(Suppl 7), vii44–vii49. <https://doi.org/10.1093/annonc/mdr425>
- Guyatt, G. (2016). N of 1 randomized trials: A commentary. *Journal of Clinical Epidemiology*, 76, 4–5. <https://doi.org/10.1016/j.jclinepi.2015.09.020>
- Guyatt, G., Rennie, D., Meade, M., Cook, D., eds. (2015). *Users' guides to the medical literature: A manual for evidence-based clinical practice*. 3rd edition. American Medical Association & McGraw-Hill Education.
- Howock, J., Chalmers, I., Glasziou, P., Greenhalgh, T., Heneghan, C., Liberati, A., Moschetti, I., Phillips, B., & Thornton, H. (2011). *Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence*. Retrieved December 16, 2019, from <https://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf>
- Jackson, S., & Schneider, J. (2018). Cochrane Review as a “warranting device” for reasoning about health. *Argumentation*, 32(2), 241–272. <https://doi.org/10.1007/s10503-017-9440-z>
- Kravitz R.L., Duan N., eds, & the DECIDE Methods Center N-of-1 Guidance Panel (Duan N., Eslick I., Gabler N.B., Kaplan H.C., Kravitz R.L., Larson E.B., Pace W.D., Schmid C.H., Sim I., & Vohra S.). Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014. <https://effectivehealthcare.ahrq.gov/topics/n-1-trials/research-2014-5/>
- Nikles, J., Mitchell, G. K., Schluter, P., Good, P., Hardy, J., Rowett, D., Shelby-James, T., Vohra, Su., & Currow, D. (2011). Aggregating single patient (n-of-1) trials in populations where recruitment and retention was difficult: The case of palliative care. *Journal of Clinical Epidemiology*, 64(5), 471–480. <https://doi.org/10.1016/j.jclinepi.2010.05.009>
- Office of the Commissioner. (2019, April 18). The Drug Development Process, Step 3: Clinical Research. Retrieved July 12, 2019, from U.S. Food and Drug Administration website: <http://www.fda.gov/patients/drug-development-process/step-3-clinical-research>
- Rothwell, P. M. (2005). External validity of randomised controlled trials: “To whom do the results of this trial apply?” *The Lancet*, 365(9453), 82–93. [https://doi.org/10.1016/S0140-6736\(04\)17670-8](https://doi.org/10.1016/S0140-6736(04)17670-8)

- Schneider, J., & Jackson, S. (2018a). Innovations in reasoning about health: The case of the Randomized Controlled Trial. In B. Garssen, D. Godden, & G. R. Mitchell (Eds.), *Proceedings of the Ninth Conference of the International Society for the Study of Argumentation, Amsterdam, Netherlands July 2018* (pp. 1042–1050). Retrieved from http://cf.hum.uva.nl/issa/ISSA_2018_proceedings.pdf
- Schneider, J., & Jackson, S. (2018b). Modeling the invention of a new inference rule: The case of ‘Randomized Clinical Trial’ as an argument scheme for medical science. *Argument & Computation*, 9(2), 77–89. <https://doi.org/10.3233/AAC-180036>
- Schwartz, D., & Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Diseases*, 20(8), 637–648.
- Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Nikles, J., Tate, R., Johnston, B.C., Zucker D., Shadish, W.R., Kravitz, R., Guyatt, G., Altman, Douglas G., Moher, D., Vohra, S., & the CENT group. (2015). CONSORT extension for reporting N-of-1 trials (CENT) 2015: Explanation and elaboration. *BMJ*, 350, h1793. <https://doi.org/10.1136/bmj.h1793>
- Taichman, D. B., Sahni, P., Pinborg, A., Peiperl, L., Laine, C., James, A., Hong, S-T., Haileamlak, A., Gollogly, L., Godlee, F., Frizelle, F.A., Florenzano, F., Drazen, J.M., Bauchner, H. Baethge, C., & Backus, J. (2017). Data sharing statements for clinical trials: A requirement of the International Committee of Medical Journal Editors. *JAMA*, 317(24), 2491–2492. <https://doi.org/10.1001/jama.2017.6514>
- Thiers, F. A., Sinskey, A. J., & Berndt, E. R. (2008). Trends in the globalization of clinical trials. *Nature Reviews Drug Discovery*, 7(1), 13–14. <https://doi.org/10.1038/nrd2441>
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge [England]: Cambridge University Press.
- U.S. National Institutes of Health. (2014). *Notice of Revised NIH Definition of Clinical Trial* (No. NOT-OD-15-015). Retrieved from <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-015.html>
- Yaffe, J., Montgomery, P., Hopewell, S., & Shepard, L. D. (2012). Empty reviews: A description and consideration of Cochrane systematic reviews with no included studies. *PLOS ONE*, 7(5), e36626. <https://doi.org/10.1371/journal.pone.0036626>

Multimodal argumentation and dissent – a perspective of multimodal critical discourse analysis

ANDREA SABINE SEDLACZEK

University of Vienna, Austria

andrea.sabine.sedlacek@univie.ac.at

In this paper, I advance a multimodal perspective on argumentative practices to investigate the way dissent can manifest in audio-visual documentary film texts, especially focusing on the role of music and sound in addition to language and images. This perspective combines aspects from critical discourse analysis, multimodal studies and the semiotic theory of Charles S. Peirce. It proposes a comprehensive analytical framework that investigates the material, referential and interpretative aspects of multimodal argumentative practices.

KEYWORDS: multimodal argumentation, dissent, critical discourse studies, multimodality, semiotics, Peirce, audio-visual texts, documentary film

1. INTRODUCTION

The conference theme “reason to dissent” serves as the starting point for this paper. In it, I try to reflect on the relationship between dissent and the study of visual and multimodal argumentation. I conceive of dissent in a broad sense, encompassing criticism, disagreement, objections, denial, refutation and controversies. My main assumption is that dissent has been at the heart of many of the discussions about visual and multimodal argumentation in the last 20 years in at least two ways: dissent *about* the object of study and dissent *as* the object of study of visual and multimodal argumentation.

From the outset, there have been dissenting views about the possibility, the actuality and the character of “visual” argumentation. Thus, the idea of visual argumentation has met with dissent or objections from early on. In this controversy, a shift can be seen in recent years from a focus on “visual” argumentation to “multimodal” argumentation. This shift has promised to overcome the common juxtaposition between “verbal” and “visual” argument that lies at the

heart of a lot of the previous controversies around the issue. I ascribe to this shift and will show in this paper that a truly multimodal approach to argumentation acknowledges that all communicative acts – and thus all argumentative practices – are necessarily multimodal. Argumentative meaning is thus created through the complex interplay of different modes, including spoken and written language, static and dynamic images, music and sound.

Specifically, I will advance a perspective on multimodal argumentation that comes from the tradition of multimodal critical discourse analysis, combining aspects from critical discourse analysis, multimodal studies and the semiotic theory of Charles S. Peirce. As part of this approach, I propose a comprehensive analytical framework that investigates the material, referential and interpretative aspects of multimodal argumentative practices (Sedlaczek, 2018).

I will then use this framework to investigate dissent as an object of study. In the research on visual and multimodal argumentation, the question whether dissent in the form of negation, denial and refutation can be expressed visually, has been long discussed (and has also met with dissenting views). Many previous contributions on this question of dissent have confined themselves to the visual or the interaction of the visual and verbal mode in both static as well as dynamic text genres. In this paper, I will show the relevance of a more comprehensive multimodal approach that explores the implications of adding other modes beside the verbal and visual for creating dissent. I will investigate the way dissent can manifest in multimodal, audio-visual texts, and I will especially focus on the way music and sound – in addition to language and images – can contribute to expressing dissent. I will illustrate this discussion with an example from a documentary film about climate change – a discursive context, where dissent and controversy is very relevant as well.

2. DISSENT *ABOUT* THE OBJECT OF STUDY: FROM VISUAL TO MULTIMODAL ARGUMENTATION

The interest in the study of “visual” argumentation has already a long history within argumentation theory (Birdsell & Groarke, 1996, 2007; Groarke, Palczewski, & Godden, 2016). From the outset, dissenting views were voiced that discussed questions such as whether visual arguments are possible and whether they exist (Blair, 1996; Champagne & Pietarinen, 2019; Fleming, 1996; Johnson, 2003; Patterson, 2010).

At the same time, the research on visual argumentation increasingly broadened to investigate argumentative practices in a wide range of communicative forms and media genres. These studies increasingly looked beyond the visual on its own to consider the

interaction between the verbal and the visual and other communicative modes, such as music, gestures or sound. Thus, the discussions increasingly shifted from “visual” to “multimodal” argumentation (Kjeldsen, 2015; Tseronis & Forceville, 2017b). In these newer discussions, two different approaches can be discerned: The first approach displays a – what I call – ‘additive’ perspective towards the role of different modes in argumentation. Starting from the perceived primacy of ‘verbal arguments’, other possible forms of argument – starting with the visual – are explored and added to the analytical focus. Groarke (2015, p. 151) for example suggests to look at verbal, visual, olfactory, tactile, savoury and auditory “modes of arguing” and how they interact.

In contrast to this additive perspective on different modes in argumentation, several contributions in recent years argue for a more inclusive and holistic perspective on multimodal argumentation. They want to overcome the conceptual distinction between “verbal”, “visual” and other possible forms of argument and argue that meaning is always context-dependent and multimodal – thus, there is no (pure) visual or verbal argumentation, but only ever “multimodal argumentation” (Žagar, 2016). Popa (2016) for example calls it a “category mistake” to talk about “visual argument” or “verbal argument” as the communicative or argumentative act is not determined by its reproduction in particular modes. Tseronis (2018) similarly pertains that “verbal” and “visual” should not be seen as categories of different arguments but that argumentative meaning is produced through the intricate interaction of different modes in a situated communicative context. Moreover, the argumentative meaning created by various modes does not only evolve out of the propositional meaning embedded in words and images, but is also influenced by choices in form and style. Tseronis thus calls for a multimodal approach to argumentation that takes the functional and context-dependent construction of meaning through the intricate interaction of different modes and their specificities into account.

My own framework takes a similar multimodal perspective on argumentation. It is informed by approaches to argumentation theory, critical discourse analysis (Reisigl, 2014; Reisigl & Wodak, 2016), multimodality (Bateman, Wildfeuer, & Hiippala, 2017; Kress, 2010; Wildfeuer, 2015) and semiotics in the tradition of Charles Sanders Peirce.

The framework presupposes a pragmatic, communicative and cognitive view on argumentation. Argumentation is situated in an interactional, communicative and discursive context, in which contentious claims are challenged and justified between various social actors who hold different points of view. Its basic purpose is to convince

or persuade the opposing party or recipient by bringing forth arguments in support of the claim (Reisigl, 2014, p. 70). This argumentative discussion is carried out in various stages, forming an intricate net of communicative acts. Argumentation is based on an abstract cognitive pattern of problem-solving that entails a functional relation between the central elements of an argumentation, i.e. the claim, the arguments or premises supporting this claim and a warrant or conclusion rule that links the arguments with the claim or conclusion (Reisigl, 2014, p. 70).

The cognitive and communicative nature of argumentation entails that an argumentative function can be established independently of the specific communicative resources used. Thus, argumentation does not have to be exclusively verbal. At the same time, the communicative choices made in the communicative acts – including the use of different modes – matter, as they influence the interpretation.

Here, I make reference to the semiotic theory of Charles Sanders Peirce (1931–1958) to suggest a comprehensive framework for the analysis of multimodal argumentation. Peirce advances a triadic sign model that builds on a triadic relation between the *sign* itself, the *object* the sign refers to and the *interpretant* as the effect of the sign on an interpreter. On each of these three sign poles, Peirce makes further triadic distinctions. Taking these concepts, one can thus investigate the *material*, *referential* and *interpretative* aspects of sign processes respectively.

In terms of the *material* quality of the sign itself, Peirce distinguishes between potential, actual and habitual signs, which he names *qualisigns/tone*, *sinsigns/token* and *legisigns/type* (CP 2.244–246; 4.537).¹

In terms of the *referential* aspects of sign processes, i.e. the relationship between sign and object, Peirce distinguishes a relationship of similarity, contiguity and convention, i.e. *iconic*, *indexical* and *symbolic* signs (CP 2.247–249).

In terms of the *interpretative* aspects of sign processes, Peirce makes several distinctions. Particularly relevant for the investigation of multimodal argumentation is his distinction of feelings, actions and thoughts as types of *interpretants* or sign effects, which he calls *emotional*, *energetic* and *logical interpretants* respectively (CP 5.575–476). In the case of a logical effect, Peirce further distinguishes between *rhematic*, *dicentric* and *argumentative* signs, which correspond to terms, propositions and inferences (CP 2.250–253).

¹ In line with accepted Peirce scholarship, references to the Collected Papers of Charles Sanders Peirce (Peirce, 1931–1958) are made with the abbreviation CP, followed by the numbers of the volume and paragraph.

On the basis of this triadic sign model, I suggest that a thorough analysis of multimodal argumentative practices has to consider three aspects: the material properties of the semiotic modes used, the referential meaning created by them and their interpretative effects (for a more detailed discussion of this semiotic framework for multimodal argumentation see Sedlaczek, 2018).

3. DISSENT AS THE OBJECT OF STUDY OF VISUAL AND MULTIMODAL ARGUMENTATION

In the following section, I will take the analytic framework presented in the last section as a foundation to look at the way dissent can manifest in multimodal texts. Dissent as an object of study of visual and multimodal argumentation has a long history. Whether images can express negation or denial and whether images can refute arguments has been a contentious issue from early on (Fleming, 1996, p. 17f.). A holistic multimodal perspective that sees argumentation as a communicative and cognitive phenomenon acknowledges that the argument does not lie in the picture itself. Rather, the argumentative function is established by the discourse participants in a particular discursive context. Thus, the question would have to be rephrased as “Can discourse participants use images to negate or refute arguments?” This question can be answered in the affirmative.

Previous contributions have offered some promising explorations of possible strategies of visual refutation, denial or negation. In their seminal paper, Lake and Pickering (1998) identified three strategies of visual refutation in three documentary films about the controversial issue of abortion. The three films explicitly reference one another and thus enter into a critical discussion, putting forth arguments and counter-arguments or rebuttals. The three strategies of (visual) refutation are: *Dissection*, in which an image used by the opposing side is verbally dissected in order to refute it. *Substitution*, where images used by the opponent are replaced by different images that show a contrasting reality. And *transformation*, where an image used by the opponent is recontextualised in a new visual frame that leads to contrasting interpretations and associations.

A similar investigation of two documentary films with opposing standpoints (on the effects of fast food on health) is conducted by Bloomfield and Sangalang (2014). They focus on the strategy of *juxtaposition* through the use of before and after images. This strategy involves an (implied) claim about causality and argues enthymematically – leaving the audience to reconstruct a premise. The authors also explore how the synecdochic function of images can be

exploited as an argument of analogy, arguing that the images are not just depicting a singular reality but are generalizable.

The strategy of *erasure* for visual denial in static ad images is explored by Oversteegen and Schilperoord (2014). They investigate erasure as a type of visual anomaly or incongruity (along with insertion, substitution and distortion). In erasure, elements that are expected to be present in a particular visual representation – by being a natural part of a cognitive schema of an object, a scene or an event – are erased and thus an interpretation of negation is evoked. In a variation of this strategy of erasure, two almost identical images are juxtaposed, and a particular element is visibly omitted in one of the two images.

Visual incongruities are also explored by Tseronis and Forceville (2017a). They refine the proposed categories of visual refutation by Lake and Pickering (1998) by differentiating the way the substitution or transformation of visual and/or verbal elements are creating an incongruity between the message of the original text and the message of the new text and thus function as an objection or rebuttal. In an investigation of static ad images from the genre of subvertisements, they identify four ways of manipulating image-text-relations for creating dissent: verbal-visual incongruity by substitution of the visual; verbal-visual incongruity by substitution of the verbal; verbal-visual manipulation conveying incongruity by transformation of the verbal and visual; and visual manipulation by transformation of the visual.

Tseronis and Forceville (2017a) advance a comprehensive multimodal approach to the question of refutation and dissent that takes a pragmatic view on communication, treating multimodal entities, such as image-text-combinations, as communicate acts and stressing that images are complex entities that involve various choices regarding composition, colour, perspective etc. that convey meanings in addition to the meaning of the whole image. As they are concerned with static image-text-genres, they however miss other modes that are relevant in audio-visual texts.

Conversely, the two papers that looked at the audio-visual genre of documentary films did not consider the relevance of other modes beside the verbal and the visual, such as music and sound, for refutation. In the following section, I want to explore how dissent or refutation can be expressed in documentary films using language, images, music and sounds. This will be done from the perspective of the semiotic framework presented above, investigating the material, referential and interpretative aspects of argumentative practices.

4. DISSENT IN DOCUMENTARY FILMS ABOUT CLIMATE CHANGE: A PERSPECTIVE OF MULTIMODAL ARGUMENTATION

I will investigate dissent in multimodal argumentation with the example of a documentary film about climate change. Similar to the previous studies of Lake and Pickering (1998) as well as Bloomfield and Sangalang (2014), this documentary film is explicitly presenting a dissenting view. The film is called *The Great Global Warming Swindle*. It is a television documentary from the British director Martin Durkin that was broadcast on the British channel 4 in 2007 and distributed internationally on DVD the following year (Durkin, 2008). It displays a sceptical position on the issue of anthropogenic climate change and tries to refute the common scientific consensus on this issue. Specifically, the film denies that climate change is caused by anthropogenic greenhouse gas emissions but pertains that climate change has natural causes and that it will not have severe consequences – thus there is no need for the commonly proposed climate change mitigation measures. In particular, the film was meant as an opposing view to Al Gore's film *An Inconvenient Truth* that had been issued the previous year.

In the following discussion, I will focus on the multimodal strategies of dissent and refutation in this documentary film.² The film shows a clear focus on dissent and heavily relies on an evaluative opposition between the consensus position to be rebuked and the own sceptical position of the film. In this dissent, the strategies of *transformation* and *substitution* suggested by Lake and Pickering (1998) can be identified. These are applied audio-visually, using language (voice-over together with prosodic features³), image (film images together with choices in cinematography and montage), music and ambient sound. Two short segments – see the transcripts in table 1 and table 2 – exemplify these strategies and serve as the basis for my discussion.

² A comprehensive analysis of *The Great Global Warming Swindle*, including a comparison to *An Inconvenient Truth*, was done by Sedlaczek (2012).

³ Kišiček (2016) has previously explored the role of prosodic features for argumentation.

Cuts	Images	Sounds	Voice-over	Music
2 sec.	the rough, dark-grey sea (medium shot)	sound of waves crashing		trumpet fanfare, dramatic music
1 sec.	a printing press is rapidly churning out newspapers (close-up, zoom-in)		"Each day news reports	...
2 sec.	a black tornado is raging against an eerily lighted violet sky (medium shot, shaky zoom-in)	sound of storm	about man-made global warming	...
2 sec.	a printing press is rapidly churning out newspapers (close-up, zoom-out)	...	grow more fantastically	...
1 sec.	palm trees are violently shaking in a storm (close-up, shaky camera)	sound of storm	apocalyptic.	...
2 sec.	the calm blue sea with a ship on the horizon (long shot, steady camera)	sound of waves	And	easy-going music with high rhythmic pattern
3 sec.	a human figure standing on a sandy beach, looking out at the calm sea, bathed in yellow sunlight (long shot, steady camera)	...	yet a number of senior climate scientists now	...
3 sec.	gentle waves on the beach, glinting in the	...	say the theory simply	...

	yellow sunlight (close-up, steady camera)		doesn't make sense."	
--	---	--	----------------------------	--

Table 1 – Multimodal dissent in *The Great Global Warming Swindle*, example 1, shots 77–84 (02:25–02:41) (Durkin, 2008)

Cuts	Images	Sounds	Voice-over	Music
2 sec.	a black tornado is raging against an eerily lighted violet sky (long shot, shaky zoom-out)			high trumpet fanfare, dramatic music
1 sec.	ships moored in a harbour are shaken by the rough sea (medium shot, shaky camera)	sounds of waves crashing	((deeper, dramatic voice)) "We are	...
1 sec.	two red pickup cars are submerged in violent dirty flood water (medium shot, shaky camera)	sounds of flood water	told that	...
1 sec.	a house is surrounded by dirty flood water (long shot, shaky pan)		we should be	...
2 sec.	palm trees are violently shaking in a storm (medium shot, shaky camera)	sounds of wind (piano)	worried, because the Earth's climate	...
1 sec.	palm trees and power lines are violently		is changing."	

	shaking in a storm (long shot, shaky camera)			
0,5 sec.	rain drops fall on a muddy ground (close-up, steady camera)			... finale of dramatic music
1,5 sec.	two human figures in rain coats and with an umbrella walk between trees (long shot, steady camera)		((change to higher tone of voice)) "But the Earth's climate is always changing.	... dramatic music fades out
2 sec.	a woman in a bikini is sitting on a sandy beach in front of a calm blue sea, with a cargo ship on the horizon (long shot, steady camera)			easy-going, whimsical music with fluctuating melodic phrases
2 sec.	a woman in a bikini is sitting on a sandy beach, applying sun lotion (medium shot, steady camera)		In Earth's long history	...
1 sec.	a palm tree in front of the clear, yellow-blue sky (medium shot, steady zoom-in)		there have been countless periods	...
1 sec.	two ships are moving on the horizon of a		when it was much warmer	...

	calm blue sea (long shot, steady camera)			
wipe, 2 sec.	a human figure in thick clothes walks in a snowstorm (medium shot, steady camera)	sounds of wind (piano)	and much cooler	...
1 sec.	a snowy mountain range (long shot, steady zoom-out)		than it is today.	...
wipe, 4 sec.	a calm sea in front of a yellow-orange sky (long shot, steady pan)		The climate has always changed, and changed without any help from us humans."	...

Table 2 – Multimodal dissent in *The Great Global Warming Swindle*, example 2, shots 144–158 (06:57–07:20) (Durkin, 2008)

The strategies of transformation and substitution are apparent in the recurring audio-visual contrasting between the two discourse positions: Through words, images, music and sound the consensus position is evoked and put into a new frame that discredits it as alarmist (*transformation*). This representation of the consensus position is then substituted by opposing words, images, music and sound that promotes the sceptical position as being more rational (*substitution*).

The way dissent is created through transformation and substitution can be more closely uncovered by investigating the material, referential and interpretative aspects of this multimodal text, following the framework I presented. The three elements of the framework should not be understood as separate analytic steps but rather as three dimensions that have to be simultaneously considered in accounting for each semiotic choice of the text in an iterative process.

The consensus position is represented through images and corresponding sounds of extreme weather events or natural disasters, especially storms and floods, accompanied by dramatic music. In contrast, the sceptical position is represented by images and sounds of sandy beaches and the calm sea, often at sunset, accompanied by easy-

going or lively music. Both sets of images consist of generic stock video material. The video material was not filmed for the documentary itself but is obtained from video databases. The generic quality of the images means that they do not represent specific events – in Peirce’s terms, they are *iconic types* that have a *rhematic interpretant*, i.e. a wide meaning potential. Ambient sounds accompanying the video material as *indexical* markers of authenticity are likewise confined to generic natural sounds. They are used sparingly and are clearly backgrounded in favour of the music. The generic film images only gain their referential meaning through the context of film editing. In line with documentary film theory, the documentary displays an “*evidentiary editing*” that is accompanied by a “voice-of-God narration” (Nichols, 2001, pp. 105-107). Thus, clear propositional and argumentative interpretations are evoked: The multimodal text audio-visually (images and sounds) and verbally (voice-over) asserts the claims of the consensus position to be rebuked and the dissenting sceptical position promoted:

The consensus position asserts that anthropogenic climate change leads to apocalyptic consequences (images of extreme weather events; “we should be worried, because the Earth’s climate is changing”). This standpoint is vaguely ascribed to media reporting (“news reports about man-made global warming” / “We are being told”) – visually emphasised through the intercutting of the images of extreme weather events with images of printing presses. Such a vague positioning of the source of the contentious standpoint to be rebuked can be seen as a strategy to avoid being accused of a straw man fallacy that exaggerates the real consensus position in order to dismiss it as alarmist. The sceptical position, in contrast, claims that a warming climate is due to natural, not anthropogenic causes and is leading to desirable consequences (evoked by the images of sandy beaches and the calm sea that invite leisure activities). This standpoint is backed with an argument of authority (“a number of senior climate scientists”) and an argument of history (“the Earth’s climate is always changing”).

Language, images and ambient sounds thus collaborate in creating propositional assertions of argumentative claims. At the same time, the asserted positions are evaluated – again verbally as well as through the visual and musical frame created: The material quality of the images of extreme weather events is generally poor. The images mostly consist of grainy video material that has an amateurish look to it, with shaky hand camera as well as cold, dark and dirty colours. Thus, the images ascribed to the consensus position are presented as unprofessional, not to be taken seriously (verbally expressed as “fantastically apocalyptic”). In contrast to the visual and musical frame accompanying the consensus position, the images of sandy beaches and

the sea consist of high-quality, steady video material, dominated by warm and light colours. They are thus presented as more authoritative and trustworthy.

The music is likewise contributing to the evaluation of the two opposing positions. Music leads to *emotional interpretants*. It wants to evoke certain feelings in the audience. The dramatic music accompanying the images ascribed to the consensus position is characterised by a monotone, 'brooding' quality, overlaid with trumpet fanfares. It exaggerates the 'apocalyptic' message in a way that the audience should feel that the argument of the consensus position is ridiculous. In contrast, the music accompanying the assertion of the sceptical position has a playful or whimsical quality, characterised by distinct musical or rhythmical patterns. It is both reassuring the positive evaluation of climate change as well as reinforcing an attitude of scepticism and dissent. These opposing evaluations are also mirrored by the tone of voice chosen by the voice-over: The narrator uses a deeper, more "dramatic" and exaggerated voice when speaking about the consensus position and a higher, neutral and sometimes patronising voice, when bringing forth the refuting argument.

By considering the material qualities of the semiotic modes used, their possibilities for creating referential meanings as well as the emotional, propositional and argumentative interpretations evoked through their interplay, the above discussion tried to offer a close look on the way transformation and substitution are employed as strategies of refutation and dissent in the documentary film *The Great Global Warming Swindle*.

5. CONCLUSION

In this paper, I investigated the strategies of dissent and refutation in a documentary film about climate change. I proposed an analytic framework for multimodal argumentation that investigates the material, referential and interpretative aspects of multimodal texts in a systematic way, taking into account language, images, music and sound. I argued that such a holistic framework can help to explore the argumentative functions of the different modes and how arguments, including dissenting or refuting views, are constructed in multimodal texts.

The framework incorporates insights from argumentation theory, critical discourse studies, the study of multimodality and the semiotics of C. S. Peirce. The holistic multimodal perspective on argumentation promoted by such an integration of approaches thus also sought to address the dissent experienced in the field of visual and multimodal argumentation, by offering a comprehensive account on the

way different modes with their distinct material qualities, their varying possibilities for representation and their potential effects combine to create argumentative meaning.

REFERENCES

- Bateman, J., Wildfeuer, J., & Hiippala, T. (2017). *Multimodality: Foundations, research and analysis – a problem-oriented introduction*. Berlin: De Gruyter Mouton.
- Birdsell, D. S., & Groarke, L. (1996). Toward a theory of visual argument. *Argumentation & Advocacy*, 33(1), 1–10
- Birdsell, D. S., & Groarke, L. (2007). Outlines of a theory of visual argument. *Argumentation & Advocacy*, 43(3–4), 103–113
- Blair, J. A. (1996). The possibility and actuality of visual arguments. *Argumentation & Advocacy*, 33(1), 23–39
- Bloomfield, E. F., & Sangalang, A. (2014). Juxtaposition as visual argument: Health Rhetoric in *Super Size Me* and *Fat Head*. *Argumentation & Advocacy*, 50(3), 141–156
- Champagne, M., & Pietarinen, A. V. (2019). Why images cannot be arguments, but moving ones might. *Argumentation*. doi: 10.1007/s10503-019-09484-0
- Durkin, M. (Writer). (2008). *The Great Global Warming Swindle* [DVD]. Germany: Sunfilm Entertainment.
- Fleming, D. (1996). Can pictures be arguments? *Argumentation & Advocacy*, 33(1), 11–22
- Groarke, L. (2015). Going multimodal: What is a mode of arguing and why does it matter? *Argumentation*, 29(2), 133–155. doi: 10.1007/s10503-014-9336-0
- Groarke, L., Palczewski, C. H., & Godden, D. (2016). Navigating the visual turn in argument. *Argumentation & Advocacy*, 52(4), 217–235
- Johnson, R. H. (2003). Why 'visual arguments' aren't arguments. *OSSA Conference Archive, Paper 49*,
- Kišiček, G. (2016). Prosodic features in the analysis of multimodal argumentation. In D. Mohammed & M. Lewiński (Eds.), *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation, Lisbon 2015* (Vol. II, pp. 629–642). London: College Publications.
- Kjeldsen, J. E. (2015). The study of visual and multimodal argumentation. *Argumentation*, 29(2), 115–132. doi: 10.1007/s10503-015-9348-4
- Kress, G. (2010). *Multimodality: a social semiotic approach to contemporary communication*. London: Routledge.
- Lake, R. A., & Pickering, B. A. (1998). Argumentation, the visual, and the possibility of refutation: an exploration. *Argumentation*, 12(1), 79–93. doi: 10.1023/A:1007703425353
- Nichols, B. (2001). *Introduction to documentary*. Bloomington: Indiana University Press.

- Oversteegen, E., & Schilperoord, J. (2014). Can pictures say no or not? Negation and denial in the visual mode. *Journal of Pragmatics*, 67, 89–106. doi: 10.1016/j.pragma.2014.03.009
- Patterson, S. W. (2010). A picture held us captive: The later Wittgenstein on visual argumentation. *Cogency: Journal of Reasoning and Argumentation*, 2(2), 105–134
- Peirce, C. S. (1931–1958). *Collected Papers of Charles Sanders Peirce*. Cambridge: Harvard University Press.
- Popa, E. O. (2016). We have yet to see the 'visual argument'. *Multimodal Communication*, 5(2), 79–92. doi: 10.1515/mc-2016-0016
- Reisigl, M. (2014). Argumentation analysis and the Discourse-Historical Approach. A methodological framework. In C. Hart & P. Cap (Eds.), *Contemporary Critical Discourse Studies* (pp. 67–96). London: Bloomsbury.
- Reisigl, M., & Wodak, R. (2016). The discourse-historical approach (DHA). In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse studies* (3rd ed., pp. 23–61). London: Sage.
- Sedlaczek, A. S. (2012). *Die visuelle Repräsentation des Klimawandels in Dokumentarfilmen. Eine multimodale kritische Diskursanalyse*. (MA Thesis), University of Vienna, Vienna.
- Sedlaczek, A. S. (2018). Multimodal argumentation in factual television. In S. Oswald & D. Maillat (Eds.), *Argumentation and Inference: Proceedings of the 2nd European Conference on Argumentation, Fribourg 2017* (Vol. II, pp. 741–754). London: College Publications.
- Tseronis, A. (2018). Multimodal argumentation: Beyond the verbal/visual divide. *Semiotica*, 220(1), 41–67. doi: 10.1515/sem-2015-0144
- Tseronis, A., & Forceville, C. (2017a). Arguing against corporate claims visually and multimodally: The genre of subvertisements. *Multimodal Communication*, 6(2), 143–157. doi: 10.1515/mc-2017-0008
- Tseronis, A., & Forceville, C. (Eds.). (2017b). *Multimodal argumentation and rhetoric in media genres*. Amsterdam: Benjamins.
- Wildfeuer, J. (Ed.). (2015). *Building bridges for multimodal research: International perspectives on theories and practices of multimodal analysis*. Frankfurt am Main: Peter Lang.
- Žagar, I. Ž. (2016). Against visual argumentation: Multimodality as composite meaning and composite utterances. In D. Mohammed & M. Lewiński (Eds.), *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation, Lisbon 2015* (Vol. I, pp. 829–852). London: College Publications.

Analogueal Argumentation in Philosophical Thought Experiments

STEFAN SLEEUW

University of Groningen, Faculty of Philosophy
s.l.sleeuw@rug.nl

Some thought experiments in philosophy involve analogies with fictitious scenarios. Such scenarios are meant to elicit intuitions that serve to support a judgment regarding a relevantly similar real-world case. In this paper, I specify a criterion for evaluating this type of thought experiment and subsequently apply it to two well-known examples.

KEYWORDS: thought experiment, analogueal argument, warrant, presumption,

1. INTRODUCTION

Many people agree that philosophical thought experiments can be reconstructed as arguments, and evaluated in terms of their argumentative strength.¹ Starting from this assumption, one may observe that not all thought experiments involve the same kind of argumentation, and hence that not all thought experiments are to be evaluated by the same criteria. A significant number of thought experiments contain *reductios*. These are aimed at casting doubt on some theory by bringing out contradictory, counterintuitive or otherwise implausible implications. There are also thought experiments, however, that rely on *analogies*. In this type of thought experiment, a judgment made about some fictitious scenario is brought to bear on a relevantly similar case-type in the actual world. It is on this latter kind, which I will from now on refer to as *analogueal thought experiments*, that I focus in the present article.

A common characteristic of analogueal thought experiments is that they involve appeals to intuition. Given that intuitions are subjective and may not be universally shared, one could doubt whether it is possible to make reasoned judgments about the argumentative strength of analogueal thought experiments. I contend that this is indeed possible: we

¹ Note that this is weaker than the claim, defended by Norton (2004), that thought experiments simply *are* arguments.

can develop a criterion that such thought experiments should minimally satisfy in order to be *prima facie* acceptable.

In **section 2**, I use a well-known example to illustrate what I have in mind when I speak of analogical thought experiments. Then, in **section 3**, I discuss two distinct frameworks for reconstructing and evaluating analogical arguments developed by Douglas Walton and Paul Bartha, respectively. Drawing on their insights as well as on work by Stephen Toulmin, I propose in **section 4** that we interpret analogical arguments as hinging on *warrants* that may come with different kinds of *qualifiers* and *backings*. I argue that for analogical thought experiments to be *prima facie* acceptable, the warrants need to be qualified and backed in such a way that they can give rise to so-called *presumptions*. This point will be further elucidated by means of two case studies in **section 5**.

2. A STARTING EXAMPLE

In her seminal paper ‘A Defense of Abortion’ (1971, p. 48-49), Judith Thomson presented an imaginary scenario that has become one of the most widely discussed thought experiments in moral philosophy:

You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. [If he is unplugged from you now, he will die; but] in nine months he will have recovered from his ailment, and can safely be unplugged from you. Is it morally incumbent on you to accede to this situation?

Thomson presents this passage, which I will refer to as *Violinist*, with the intention of eliciting a definite, intuitive moral judgment, which she (1971, p. 49) herself anticipates as follows: “No doubt it would be very nice of you if you did, but do you have to accede to [the situation]? (...) I imagine you would regard this as outrageous.” In other words, Thomson thinks it is morally permissible to unplug yourself from the ailing violinist, even if this inevitably results in his death. On the basis of this result, Thomson argues that we have a strong reason for endorsing the permissibility of abortion in cases of involuntary pregnancy.

The argumentation underlying this conclusion remains largely implicit in Thomson's article. On most accounts, however, it is reconstructed along the following lines:

- P1: Violinist is relevantly similar to cases of involuntary pregnancy (e.g. pregnancy due to rape)
- P2: In Violinist, you are morally permitted to disconnect from the ailing individual.
- C: Therefore, abortion is permitted in cases of involuntary pregnancy.

This is a clear instance of analogical argumentation. In Section 5 I will discuss two other examples of analogical thought experiments. Throughout sections 3 and 4, however, I will take Violinist as my go-to sample case, meant to exemplify common features of analogical thought experiments in general.

3. ANALOGICAL ARGUMENTATION

Assuming Thomson's thought experiment is indeed an analogical one, what is the general structure of analogical arguments? And how are they to be evaluated? These questions need to be considered before we can develop a specific evaluation criterion for analogical thought experiments. In this section, I will first look at Douglas Walton's important contribution to the study of analogical reasoning in argumentation theory. Then, I will elaborate on a different proposal for representing and assessing analogical arguments that has recently been defended by Paul Bartha.

3.1 Walton's argument scheme approach

Over the course of several books and articles, Douglas Walton has canonized a wide variety of argument schemes including not only standard forms of inductive and deductive reasoning, but also inference types that he labels alternately as 'defeasible', 'presumptive' or 'plausible'. Generally speaking, an argument belonging to this third category "is presented as providing only a defeasible support for its conclusion, subject to critical questioning in a context of dialogue" (Walton, Reed & Macagno, 2008, p. 3). In order to assess the strength of particular instances of such defeasible argumentation, the various schemes for defeasible argument types come with lists of standardized critical questions. According to Godden & Walton (2007, p. 12) these critical questions serve to evaluate the cogency of an argument based on three aspects:

- i. whether its premises are rationally acceptable,
- ii. whether its premises are relevant to the conclusion,
- iii. whether its premises provide sufficient reason to accept the conclusion.

The strength of defeasible arguments is taken to depend on the extent to which someone who puts forward the argument is deemed able to adequately answer critical questions that are appropriate in light of i-iii.

According to Walton, analogical argumentation falls under the heading of defeasible argumentation.² His simplest version of the argument scheme for analogical argumentation looks as follows (Walton, Reed & Macagno, 2008, p. 56):

Major premise: Generally, case C1 is similar to case C2.

Minor Premise: Proposition A is true (false) in case C1.

Conclusion: Proposition A is true (false) in case C2.

This being a defeasible kind of inference, it is associated with a number of generalized critical questions (idem, p. 62):

CQ1: Is A true (false) in C1?

CQ2: Are C1 and C2 similar, in the respects cited?

CQ3: Are there important differences (dissimilarities) between C1 and C2?

CQ4: Is there some other case C3 that is also similar to C1 except that A is false (true) in C3?

In this list, CQ1 and CQ2 pertain to the rational acceptability of the minor and major premise, respectively. CQ3, on the other hand, is meant to test the relevance of the premises: depending on whether there are important differences between C1 and C2, the similarities between the two cases may or may not be rendered insignificant. CQ4, finally, concerns the sufficiency of the premises. That is, depending on the availability of counterexamples to the analogical argument, otherwise relevant similarities between C1 and C2 may or may not turn out to be sufficient for drawing the conclusion.

3.2 Bartha's articulation model

When it comes to accounts of analogical argumentation, philosophy of science has generated useful insights. I am thinking especially of Mary Hesse's analysis of analogical reasoning (Hesse 1966), which was refined and extended by Paul Bartha in *By Parallel Reasoning* (2010). According to Bartha, whose so-called *articulation model* I will concentrate on in the remainder of this section, the structure of analogical arguments is best conceived of along the lines of Figure 1:

² While they do not all use his terminology, several authors have sided with Walton in this regard. See, for instance, Govier (1989) and Guarini (2004).

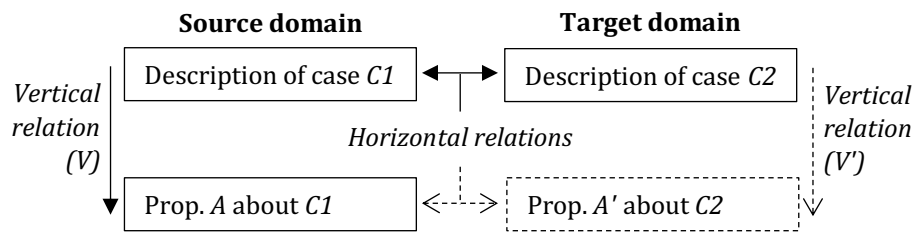


Figure 1 – A generalized and modified version of Bartha's model

According to this model, analogical arguments first of all consist of a source and a target domain, which are taken to be similar in certain respects. These known similarities comprise one part of the so-called *horizontal relations*, signified in Figure 1 by the solid double-headed arrow. In addition, Bartha identifies a *vertical relation (V)*, which can be described as the inference that supports proposition *A* in the source domain. On the basis of the known similarities, this vertical relation is projected onto the target domain (*V'*), where it serves to establish a new proposition *A'* and with that a further similarity between the two domains – represented by the dashed, double-headed arrow.

There are clear correspondences between Bartha's approach and Walton's scheme for arguments by analogy. First of all, what I have called the 'known similarities' match Walton's major premise. Secondly, we can see that Walton's minor premise is found in the bottom left frame of the diagram. Thirdly, the conclusion of Walton's scheme corresponds to proposition *A'* in the diagram's target domain. However, there is also a fundamental way in which the two accounts differ: the notion of a vertical relation does not have an immediate counterpart in Walton's scheme. Bartha (2010, p. 99), who has provided a detailed account of this notion, claims that as a minimal condition for acceptability, an analogical argument must "not simply put forward an unstructured list [of similarities], (...), but rather [present] the relevant factors in some definite relationship to each other". This definite relationship, or *prior association* as Bartha calls it, "may be either deductive or inductive" (Bartha, 2010, p. 97).

To see what role prior associations play in analogical arguments, let us briefly look at examples of the two main kinds recognized by Bartha. First, here is an example from mathematics that involves a deductive prior association:

Suppose we have proved that the three medians of any triangle have a common intersection point. By analogy, we conjecture that the medians of any tetrahedron – the lines joining each vertex with the center of the opposite face – have a common intersection. (Bartha, 2010, p. 95)

The source domain of this analogy consists of the claim that the three medians of any triangle have a common intersection point. Given that this claim is assumed to have been proven and thus established with certainty, the prior association in this analogical argument is deductive. The category of inductive prior associations, secondly, is exemplified by the practice of drug testing on non-human animals, where certain physical effects of a drug observed in non-human test animals are hypothesized to occur in humans as well. Here, the inference in the source domain takes the form of an empirical generalization: sufficiently frequent observation of some physical effect after administering a drug to individuals of a certain species makes it probable that, in general, administering the drug to individuals of that species leads to the same effect.

4. ANALOGICAL THOUGHT EXPERIMENTS

As it turns out, neither Walton's argument scheme theory, nor Bartha's articulation model provides us with a fully adequate instrument for evaluating thought experiments such as *Violinist*. While *Violinist* can be reconstructed according to Walton's scheme, the critical questions are insufficient for making a reasoned assessment of the intuitive judgment. Bartha's account, on the other hand, might not even countenance *Violinist* as a proper analogical argument. After all, the conclusion reached in *Violinist*'s source domain does not rest on an inference that is straightforwardly deductive or inductive. Hence, *Violinist* is not readily interpreted as containing a prior association of the kind Bartha allows for. However, we are not required to opt for one approach to the exclusion of the other. I propose that by syncretizing some of Walton and Bartha's insights, we can think of analogical thought experiments as relying on so-called *warrants* (Toulmin 1958/2003). Based on the further idea that such warrants should act as presumptions in an argumentative dialogue, we can formulate a minimal criterion that analogical thought experiments should satisfy in order to be *prima facie* acceptable.

4.1 Bartha meets Toulmin: prior associations as warrants

In Thomson's original discussion of *Violinist*, the idea that unplugging from the ailing violinist is permissible is presented as more or less self-evident.³ While such appeals to self-evidence are not naturally interpreted as involving implicit inductive or deductive inference-principles, they may in fact be reconstructed as instances of the kind of defeasible reasoning recognized by Walton. Looked at in this way, we can

³ Recall Thomson's remark that "I imagine you would regard this as outrageous".

read Violinist and other analogical thought experiments as relying on prior associations that have the form of *defeasible inference-licensing principles*. In order to make this more tangible, we can liken the idea of a defeasible inference-licensing principle to Stephen Toulmin's notion of a warrant that supports the step from an argument's data to its claim. Warrants, according, to Toulmin (1958/2003, p. 91), are "general, hypothetical statements, which can act as bridges, and authorise the sort of step to which our particular argument commits us". More precisely, the function of a warrant is "to register explicitly the legitimacy of the step involved and to refer it back to the larger class of steps whose legitimacy is being presupposed" (idem, p. 92). Figure 2 sketches how data (D), warrant (W) and claim (C) hang together:

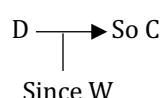


Figure 2 – A schematic rendering of Toulmin's analysis of arguments (Toulmin, 1958/2003, p. 92)

Let us, by way of illustration, apply Toulmin's notions to Thomson's argument. For that thought experiment, the data, warrant and claim can be rendered as follows:

D: the Violinist scenario

W: *if someone comes to vitally depend on your body in a way that uninvitedly and severely restricts your bodily autonomy for a prolonged period of time, it is permissible to withdraw yourself from that predicament*

C: you are permitted to unplug yourself from the violinist

Stated in Bartha's terms, the above warrant functions as the vertical relation that is projected onto the argument's target domain: cases of involuntary pregnancy. It is formulated in a deliberately abstract way, so as to account for the fact that Toulmin thinks of warrants as 'generalized conditionals'.⁴ That is, warrants are not to be expressed in the form 'If [data], then [claim]', but rather in the form 'If [data-like information], then [claim-like conclusion]'. Because of this, implicit warrants can be reconstructed in more than one way. For current purposes, however, the point is simply to show how an analogical thought experiment such as Violinist can be read as involving a Toulminian warrant that is functionally similar to a Barthian prior association.

⁴ This point is also emphasized by Hitchcock (2003).

4.2 Thought-experimental warrants as presumptions

Toulmin thinks of warrants as being *qualified* in some way. In addition, he argues that warrants must come with *backings*. If we assume that, ideally, the warrants in analogical thought experiments are qualified and backed in such a way that they can function as so-called *presumptions*, it becomes possible to formulate an evaluation criterion that is specifically tailored to analogical thought experiments we encounter in philosophy.

For Toulmin, a qualifier essentially refines a warrant (Toulmin 1958/2003, p. 93):

Warrants are of different kinds, and may confer different degrees of force on the conclusions they justify. (...) It may not be sufficient, therefore, simply to specify our data, warrant and claim: we may need to add some explicit reference to the degree of force which our data confer on our claim in virtue of our warrant. In a word, we may have to put in a *qualifier*.

Toulmin recognizes a broad variety of qualifiers. Some are associated with inductive and deductive inferences, such as ‘probably’, and ‘necessarily’. Others belong to defeasible argumentation, such as ‘plausibly’ and ‘arguably’. Obviously, a qualifier like ‘necessarily’ is meant to confer more force on the argument’s conclusion than one like ‘arguably’. But in order for it to justifiably confer this force, the warrant must be supported in some way. This is where the notion of a *backing* comes in. A backing is a statement intended to ground the general acceptability of the warrant. It typically does so by relating the warrant to some principle or source of information that is taken to authoritative within the field or discipline that the argument operates in (Toulmin, 1958/2003, p. 96).

One kind of qualifier recognized by Toulmin is ‘presumably’. If put in front of a warrant, the warrant’s associated claim is turned into a so-called *presumption*. According to Walton (1993, p. 138), a presumption is a kind of speech act that is “halfway between assertion and (mere) assumption”. Walton’s characterization presupposes the framework of a rule-governed argumentative dialogue between a proponent and an opponent (or respondent). In this dialogical setting, there is a *proponent* who aims to persuade the other party of a particular thesis, while the *opponent* seeks to maintain his critical stance towards that thesis. Whoever puts forward an assertion in such a dialogue incurs a *burden of proof*: an obligation to provide reasons for the assertion if challenged to do so by the other party. By contrast, there is no burden of proof associated with presenting something as a mere assumption. Participants can freely put forward and retract assumptions without having to defend these dialogical moves. Presumptions, like assertions, do carry a burden

of proof. Unlike assertions, however, this burden does not fall onto the participant who puts forward the presumption, but rather on the one who intends to refute it. As Walton (1993, p. 138) puts it: “when a presumption is brought forward by a proponent, the burden is on the respondent to refute it, or otherwise it goes into place as a commitment.” The function of presumptions is often practical: in the absence of conclusive evidence, taking recourse to a presumption may be the best way to allow a dialogue to move forward.

Of course, not just any statement can function as a presumption. If a statement is presented by a proponent as a presumption but fails to be a reasonable one, the proponent can be accused of having begged the question. As with other warrants, then, a warrant qualified by ‘presumably’, must be appropriately backed in order for that qualifier to justifiably confer its force upon the claim. As Godden & Walton (2007, p. 337) point out, what kind of normative foundation is appropriate, may vary per context. In certain disciplines, such as law, there are certain presumptions that are firmly grounded in institution-specific rules.⁵ In philosophy, however, the dialectical room for making presumptions is often limited. As I see it, there are two dialectical scenarios in which a statement put forward as a presumption may reasonably function as such:

- a) The statement appeals to claims or theories that the opponent can be expected to accept given his or her known antecedent philosophical commitments.
- b) The statement appeals to claims or theories whose undefended rejection by the opponent would count as unduly revisionary given some sufficiently strong consensus within the relevant context of philosophical discussion.

Note that these are scenarios in which statements may, but need not always function as reasonable presumptions. That is, both (a) and (b) are defeasible as criteria for making reasonable presumptions.

Let us once more clarify the foregoing by means of Violinist. As mentioned earlier, Thomson does not defend the intermediate conclusion of her thought experiment by providing a substantive reason. This indicates that she thinks of this conclusion as a claim that her intended audience can be expected to subscribe to. Because of this, I take it that the qualifier associated with the warrant-reconstruction proposed earlier is best formulated as ‘presumably’:

⁵ Think for instance of the innocence presumption in criminal law.

presumably, *if someone comes to vitally depend on your body in a way that uninvitedly and severely restricts your bodily autonomy for a prolonged period of time, it is permissible to withdraw yourself from that predicament*

Whether this particular warrant is indeed adequately backed, such that its associated conclusion can function as a reasonable presumption in the sense of satisfying (a) and/or (b), is not an issue I will tackle here. What I have wanted to show so far is simply the way in which analogical thought experiments can be reconstructed in a way that opens a new avenue for evaluating them. Given the defeasibility of (a) and (b), however, conclusive verdicts regarding the reasonableness of presumptions may not be feasible. I do think, however, that my analysis allows for comparative assessments of different analogical thought experiments. I will illustrate this in the next section.

5. (UN)REASONABLE PRESUMPTIONS IN ANALOGICAL THOUGHT EXPERIMENTS: TWO CASE STUDIES

The insights from the previous sections can be clarified by contrasting two famous analogical thought experiments in philosophy. This enables us to see how my framework may yield comparative judgments regarding the reasonableness of the presumptions that are implicit in analogical thought experiments.

5.1 First case study: Peter Singer's Drowning Child

In his seminal paper 'Famine, Affluence, and Morality' (1972), Peter Singer presented a thought experiment that draws on an analogy. The source domain of this analogical argument is found in the following passage, which we may refer to as *Drowning Child*:

If I am walking past a pond and see a child drowning in it, I ought to wade in and pull the child out. This will mean getting my clothes muddy, but this is insignificant, while the death of the child would presumably be a very bad thing. (Singer, 1972, p. 231)

Singer thinks it is clear that I should attempt to save the child despite the repercussions this might have for my clothes. Drowning Child, along with this moral imperative, can be inserted into Bartha's model as shown in figure 3:

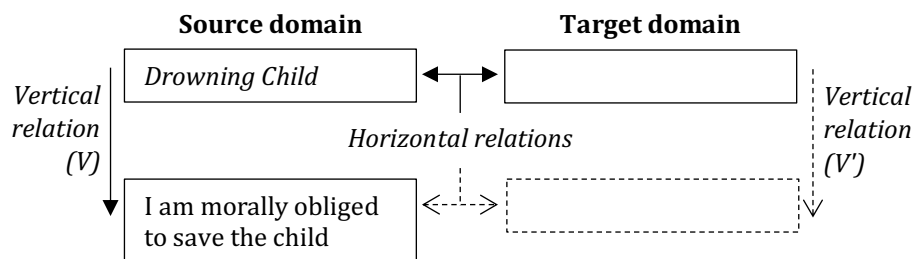


Figure 3 – A Barthian reconstruction of Drowning Child's source domain

The target domain is formed by a hypothetical humanitarian crisis that can be considerably mitigated through a relief fund. According to Singer, I am – by analogy with Drowning Child – morally obligated to donate to such a relief fund, provided that doing so is not likely to harm me in any significant way. By completing the above diagram, we arrive at figure 4:

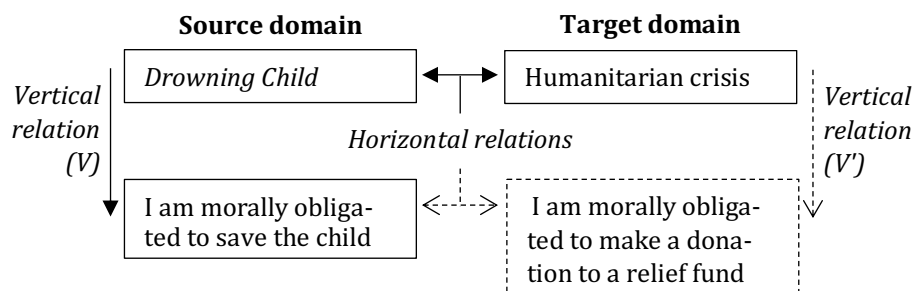


Figure 4 – The target domain added

Here the vertical relation or warrant apparently comes in the form of a general moral principle:

[I]f it is in our power to prevent something very bad from happening, without thereby sacrificing anything morally significant, we ought, morally, to do it. (Singer 1972, p. 231)

Singer deems this principle to be so uncontroversial as to require no further substantive argument. As with Thomson's *Violinist*, then, we can regard this warrant as being qualified by 'presumably'. The remaining question now is, whether the warrant is adequately backed in the sense of yielding a reasonable presumption. While we need not arrive at a definitive verdict here, there is at least strong textual evidence that Singer has taken criterion (a) into account, by giving some consideration to the possible positions of his interlocutors. First of all, Singer (1972, p. 231) points out that "[the principle] requires us only to prevent what is bad, and not to promote what is good"; thereby arguably forestalling a

potential demandingness objection. Secondly, he has attempted to formulate the principle in such a way that it imposes the least requirements on one's ethical commitments. This is evidenced by Singer reformulating the initial clause "without thereby sacrificing anything of comparable moral importance" into "without thereby sacrificing anything morally significant" (Singer, 1972, p. 231). Given these two pieces of textual evidence, there is some reason to believe that Singer has paid heed to criterion (a).

5.2 Second case study: John Searle's Chinese Room

John Searle's *Chinese Room* is arguably one of the most famous philosophical thought experiments:

I am sitting alone in a room, following instructions in English for responding to Chinese characters slipped under the door. My native language is English, and my knowledge of Chinese is negligible. Outside observers have the impression that they are communicating with a competent speaker of Chinese.⁶

According to Searle, I do not understand Chinese in this scenario, despite my apparent communicating abilities. We can thus fill in the left hand column of our diagram, seen in figure 5:

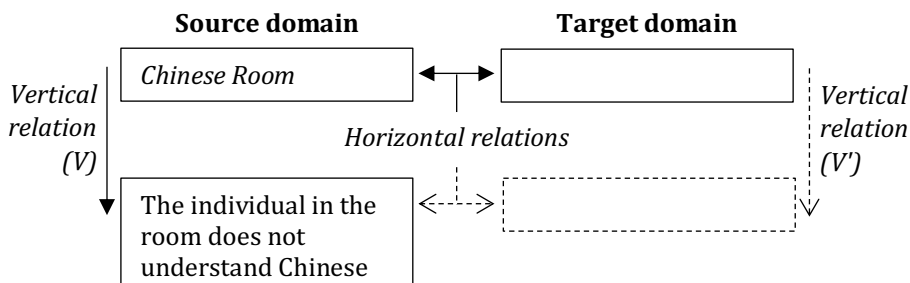


Figure 5 – A Barthian reconstruction of Chinese Room's source domain

Searle deems this source domain to bear relevant similarities to the following, real-world scenario:

A sufficiently advanced digital computer running a program of instructions for responding to Chinese characters, may produce correct and appropriate answers.⁷

⁶ This is an abridged version of the original scenario (Searle, 1980, p. 417-418).

⁷ Again, a summarized version.

By analogy, Searle argues, such a computer does not understand Chinese either. The resulting diagram thus looks as follows:

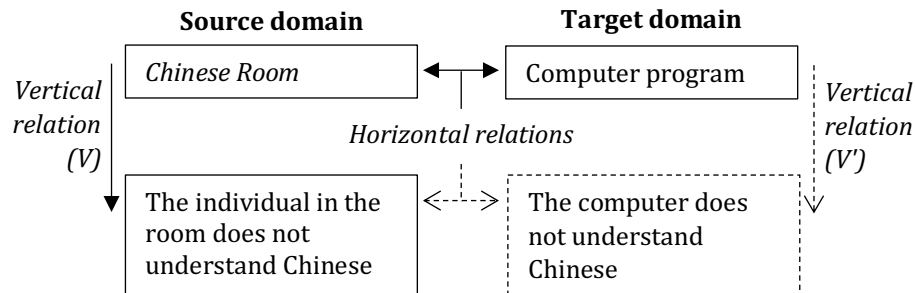


Figure 6 – The target domain added

Unlike Singer, Searle does not explicitly formulate a warrant, but merely appeals to our intuitions: “it seems to me quite obvious in the example that I do not understand a word of the Chinese stories” (Searle, 1981, p. 418). As becomes apparent from various other passages in the article, Searle strongly believes that ‘mere’ manipulation of symbols is insufficient for understanding; we also need to know what the symbols mean (Searle, 1980, p. 418). Taking this into account, I propose the following principle is implicitly operating as the warrant in Searle’s argument:

presumably, *if someone or something generates appropriate messages merely through manipulating uninterpreted symbols of a language, he, she or it does not thereby understand that language*

While this principle may seem unproblematic to many, it is questionable whether it can really yield a presumption that is reasonable. Here is why. By Searle’s own admission, his intended opponents are defenders of what he calls *strong AI*, according to which “the appropriately programmed computer really *is* a mind, in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states” (Searle, 1981, p. 417). Clearly, the warrant just formulated is not acceptable for adherents of this view: it begs the question against them. Hence, it cannot serve to raise a reasonable presumption in the sense of criterion (a). Moreover, since Strong AI is arguably not a fringe position that goes against some overwhelming and long-standing consensus, Searle’s thought experiment relies on a presumption that also fails to satisfy (b). This means that in terms of *prima facie* acceptability, Searle’s Chinese Room Experiment has a dubitable status, at least in comparison to Singer’s Drowning Child.

6. CONCLUSION

Analogical thought experiments in philosophy often hinge on appeals to intuition. Because of this, it may be difficult to form reasoned judgments concerning their persuasiveness. By drawing on Douglas Walton and Paul Bartha's accounts of analogical argumentation, I have argued that otherwise intractable appeals to intuition can be made more explicit if we reconstruct them as Toulminian warrants. The main finding of this paper is that in order for an analogical thought experiment to be at least *prima facie* acceptable, its warrant needs to function as a reasonable presumption in an argumentative dialogue. This criterion enables us to compare the strength of analogical thought experiments in a more objective manner.

REFERENCES

- Bartha, P. (2010). *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*. Oxford: Oxford University Press.
- Brun, G. (2018). Thought Experiments in Ethics. In M.T. Stuart, Y. Fehige and J.R. Brown (eds.), *The Routledge Companion to Thought Experiments* (pp. 195-210). New York: Routledge.
- Guarini, M. (2004). A Defence of Non-deductive Reconstructions of Analogical Arguments. *Informal Logic*, 24(2), pp. 153-168.
- Godden, D.M. & Walton, D. (2007). A Theory of Presumption for Everyday Argumentation. *Pragmatics & Cognition*, 15(2), pp. 313-346.
- Govier, T. (1989). Analogies and Missing Premises. *Informal Logic*, 11(3), pp. 141-152.
- Hesse, M.B. (1966). *Models and Analogies in Science*. Notre Dame: University of Notre Dame Press.
- Hitchcock, D. (2003). Toulmin's Warrants. In Van Eemeren et al. (eds.), *Anyone Who Has a View: Theoretical Contributions to the Study of Argumentation* (pp. 69-82). Dordrecht: Kluwer.
- Norton, J.D. (2004). On Thought Experiments: Is there More to the Argument? *Philosophy of Science*, 71, pp. 1139-1151.
- Searle, J.R. (1980). Minds, Brains, and Programs. *The Behavioral and Brain Sciences*, 3, pp. 417-457.
- Singer, P. (1972). Famine, Affluence, and Morality. *Philosophy and Public Affairs*, 1(3), pp. 229-243.
- Thomson, J.J. (1971). A Defense of Abortion. *Philosophy and Public Affairs*, 1(1), pp. 47-66.
- Toulmin, S.E. (2003). *The Uses of Argument*. Cambridge: Cambridge University Press. (Original work published 1958).
- Walton, D. (1993). The Speech Act of Presumption. *Pragmatics & Cognition*, 1(1), pp. 125-148.
- Walton, D., Reed, C. & Macagno, F. (2008). *Argumentation Schemes*. Cambridge: Cambridge University Press.

Detecting and handling disagreement in multi-party health coaching

MARK SNAITH

Centre for Argument Technology, University of Dundee, UK
m.snaith@dundee.ac.uk

ALISON PEASE

Centre for Argument Technology, University of Dundee, UK
a.pease@dundee.ac.uk

DOMINIC DE FRANCO

Centre for Argument Technology, University of Dundee, UK
d.f.defranco@dundee.ac.uk

We describe an empirical study of disagreement in multi-party healthcare coaching. We analyse a patient interview dataset for dissent, including both highly explicit examples using well-established conflict keywords and actions, and more subtle dissent in terms of language and gestures. We use our analysis to address questions such as: “What types of conflict occur in this context?”, and “Can we identify different types of disagreement and corresponding resolution strategies?”

KEYWORDS: conflict, dialogue, disagreement, multi-party healthcare coaching

1. INTRODUCTION

Disagreement is common in health coaching, especially in contexts where the aim is to bring about behaviour change. For instance, a patient may disagree with a proposed course of action because they are reluctant to change their behaviour. Furthermore, if the health coaching is also multi-party, where more than one practitioner is in consultation with the patient, disagreement can then occur not just between the patient and the practitioners, but also between the practitioners themselves. In some cases, this is due to a lack of knowledge by one coach of the domain of another rather than outright contradiction of the other coach’s statements (e.g. a GP questioning a proposal made by a

nutritionist), but in others it may be a genuine disagreement (e.g. a nutritionist and physical activity coach clashing over the best way to lose weight).

Examining real patient consultations presents a significant challenge from both an ethical and a practical point of view. We therefore constructed a data set from simulated sessions involving real healthcare professionals in consultation with a patient played by an actor, playing to a realistic persona.

In this paper we examine different ways in which disagreement is expressed in multi-party health coaching. Using data collected from sample coaching sessions, we highlight specific examples of both verbal and non-verbal disagreement from the patient to the medical practitioners; the medical practitioners to the patient; and between the practitioners themselves. We find that disagreement is communicated in a variety of different ways, from the highly explicit, using well-established conflict keywords and actions, to ways that are subtler in terms of language and gestures. We start to answer our two research questions: *"In what contexts does disagreement occur and how is it expressed in those contexts?"*, and *"Can we identify resolution strategies for different types of disagreement?"*.

2. METHODOLOGY

This paper forms part of a larger study into conflict in healthcare. We used an approach based on grounded theory (Strauss and Juliet, 1994) to conduct a fine-grained study, in order to answer our research questions. This is a data-driven method to systematically build integrated sets of concepts in a topic where little is known. Researchers keep an open mind in order to build a theory which is purely grounded in data rather than influenced by prior work. As is the standard in grounded theory, once we had gathered our data, we followed four stages to analyse it:

1. **Open coding:** Use the raw data to suggest code definitions (anchors that help to identify key points in the data).
2. **Axial coding:** Development of concepts by combining codes into collections of similar content.
3. **Selective coding:** Grouping the concepts into categories – put the data back together by making connections across codes, categories, and concepts.
4. **Theory building:** Compare the central phenomenon across several dimensions, and formulate the major themes which have emerged.

Here codes, concepts and categories are different levels of abstraction and are the building blocks for a grounded theory. Grounded theory is a highly iterative process: in this paper we describe preliminary findings on a small subset of the data (examples 1-3 below). Further work will include iterating through the stages multiple times on the entire dataset. Features of our approach include:

a focus on empirical data from which our work derives

a focus on multi-disciplinary healthcare teams

an open exploration of conflict on healthcare, demonstrated with illustrative examples.

3. CONFLICT IN HEALTHCARE SETTINGS

Research on conflict is multidisciplinary, including argumentation, philosophy, rhetoric, anthropology, sociology, psychology, and linguistics.

Theories of conflict and disagreement in argumentation focus on the structural properties of arguments or disputes. For instance, Pollock characterises two types of disagreement as rebutting defeaters that deny a conclusion P, and undercutting defeaters that attack the reason for believing P, but not P itself (Pollock, 1995). He argues that these two types can describe the full complexity of defeasible reasoning.

Scholars in related fields such as sociolinguistics and pragmatics have investigated situated, cultural, and social meanings of conflict. Kakavá reviews contextual elements of disagreement, such as how conflict is initiated and how it develops, how it is organised sequentially, and its status and use in social interaction (Kakavá, 2001). In particular, she surveys work on (1) the structural properties of conflict; (2) the communicative strategies of conducting conflict; (3) conflict negotiation and resolution; and (4) the meanings of conflict.

There is much work on both intra-team and patient-team conflict in healthcare settings. Here, sources and costs of conflict are often considered to be established, for instance costs including burnout, higher absenteeism, higher turnover (Almost *et al.*, 2016), hostility, emotionality and distraction from actual task accomplishment (Greer *et al.*, 2012). Given the costs, the current focus of much work is now on researching conflict resolutions and interventions in order to minimize it. While conflict is widely seen as a negative force, impeding healthy collaborations, some authors point out positive roles it can play, such as enhancing team members' understanding and performance of a given task (*ibid.*).

4. THE PATIENT CONSULTATION CORPUS

Understanding how healthcare professionals and patients interact is challenging, both ethically and practically. Observing real consultations risks changing the dynamics and outcomes of those consultations through the presence of a recording device – for instance, a patient may be less forthcoming about a medical problem if they know they are being recorded, even if they have been given assurances of privacy. Furthermore, it is unusual to find two or more healthcare professionals in a single consultation. To address these problems, we employed the use of healthcare simulation to develop the **patient consultation corpus**, which we will now briefly describe in the remainder of this section.

Simulation within medical practice is a common process in used in medical training, underpinned by a number of educational theories (Gaba, 2004). In such an approach, the patient is portrayed by an actor playing to a specified persona and associated medical history, and medical professionals do as they would if the actor were a real patient. Similar role-playing techniques have been successfully used as a data collection tool in other sensitive contexts such as dispute mediation (Janier & Reed, 2016) in which attempting to record real consultations raises similar ethical and practical questions to the medical domain.

Key to successful healthcare simulation is careful design of the patient personas. It is important to ensure that the actors portray patients that are realistic, both in terms of their overall personality and the medical conditions they are experiencing. Using an iterative process that followed several cycles (including analysis by a medical practitioner) we developed four personas, which are summarised in Table 1. Note that while a gender is specified for the persona, this was not fixed: through only tweaking minor details, each persona was adaptable to be played by an actor of any gender.

No	Gender	Age	Personality
1	Male	57	Know-it-all
2	Female	63	Anxious
3	Female	50	Unengaged
4	Male	67	Benchmark

Table 1: Patient personas

The benchmark patient was regarded as being an engaged and consciences patient so as to provide a baseline comparator for the other personality types.

The Patient Consultation Corpus consists of nine sessions in total, using three different actors and five different healthcare

professionals in varying combinations. The different healthcare professionals are: a general practitioner (GP); a diabetes specialist; a podiatrist; a dietician; and a motivational interviewer (MI). Motivational interviewing is a patient-centred counselling approach, that is designed to help clients explore and resolve ambivalence.

5. EXAMPLES OF DISAGREEMENT

Here, we provide three examples of disagreement identified through an initial observation of the patient interview corpus while it was being constructed. In some examples, non-verbal cues are important to the style and type of disagreement; where relevant, we highlight these in square brackets (e.g. *[voice pitch raises]*).

5.1 Intra-team disagreement

The first example of disagreement we identify is between two healthcare professionals:

Dietician: *Cravings can be pretty tricky to manage, but the good thing is they do pass. And, actually, if you're having that at the same time every day, what you're developing there is a habit, of then every time you go and have a coffee, you're like mm, I could really do with my chocolate bar. And don't get me wrong, habits are pretty easy to develop, but actually to break them is a really, really tricky thing. But, within time, that craving will pass, which is a bit reassuring.*

General Practitioner: *So, are you suggesting, and I don't want to put words in your mouth, but the chocolate bar, do you just cut them out? Because that seems a bit harsh! [raises eyebrows; slight laugh]*

Dietician: *It does seem a little bit harsh, I suppose, and that's where we could have a discussion about actually what would be a realistic target for you.*

Nature of the disagreement: In this example, the general practitioner (GP) disagrees with the dietician as to whether or not the patient should stop eating chocolate entirely.

How is the disagreement expressed? After explaining to the patient about cravings and how to break them, the general practitioner (GP) interjects, asking if that means to cut out chocolate completely because

it “...seems a bit harsh”. This was said in a light-hearted way with an element of laughter and raised eyebrows, emphasising that the GP is not outright contradicting the dietician, but instead is genuinely questioning the (believed) suggestion that the patient should completely cut chocolate out of their diet. Use of the word “seems” also demonstrates that the disagreement arises from intuition rather than knowledge (i.e. he did not say “...that **is** harsh”).

Resolution strategy: The resolution strategy for this disagreement was seeded in the expression. The element of laughter from the GP along with the use of the word “seems” emphasised that the disagreement was casual, almost friendly. Furthermore, it was framed as being from intuition rather than knowledge, leaving an easy route for the dietician to respond.

Was the resolution successful? The dietician’s response – “*It does seem a little bit harsh, I suppose*” – shows that the disagreement was immediately resolved. The dietician took on board what the GP said and used it as a way of seeding a further discussion about exactly what the patient should do to achieve a reduction in their consumption of unhealthy snacks.

5.2 Embedded disagreement

Our second example of disagreement is between a patient with a “know-it-all” persona, and two healthcare professionals – a dietician, and a motivational interviewer. The patient is determined not to eat carbohydrates (sometimes referred to as “carbs”), even if this helps control their blood-sugar levels. They also do not believe that there is a link between carbohydrates and sugar thanks to their own, online research.

Dietician: *No. So, actually about a third of our diet, we recommend coming from starchy carbohydrate, so that’s things like your bread and your pasta and your rice. You know, the types of carbohydrates that you’ve just actually mentioned.*

Patient: *Well, I think we’ll just have to agree to disagree on that one, because there’s absolutely no way that I’m going to start taking breads and pastas and things into my diet, no.*

...

Motivational Interviewer (MI): Okay, so you've kind of researched this, felt that this works for you. I guess Barbara's question was projecting forward, around about how might it be to continue on this long term? And what's your thoughts about that, and how you might see things, if you are experiencing that diet long term?

Patient: I don't know how else you could possibly change it, apart from, you've said about carbs, which I'm obviously not willing to do, so I mean, is there another way around it, if you think that eventually I'm not going to be able to follow it, which I think is very critical of you, but?

...

MI: Okay, so you don't see any relation between your low carb diet and your blood sugars dipping?

Patient: No, no, because Google tells me that diabetes is due to sugar intake, and I'm not taking sugar, so...

MI: Okay. So, I get that you have researched. Can I just ask, in your work, if you research a piece for your work, what do you search through?

Patient: Well, I use the internet, but I also use libraries, there's different firms who have got past cases that you can look at, and you can look at manuals that you've already, I work for a toy manufacturing company, so there's different manuals that you can look at, to just get guidance from. There's lots of different things you can compare them.

...

Motivational Interviewer (MI): How does that compare to your research about something as complex as diabetes, with a general search engine?

Patient: Well, I mean, Google, you put in a question, you get information from all over the world, that's like the easiest way to get all the information in one place and there's lots of different sources you can look at, to make sure the facts are consistent.

MI: So I could do your job just by putting a search term to Google?

Patient: *Oh no, you need years of experience, you need to train properly as a lawyer, learn all that, you need to do the work-shadowing and the placements, that's a completely different thing.*

MI: *So, and I guess what I'm wondering is, as you draw that comparison in your mind, about the research you have done and the methods you've done to research diabetes to enhance your understanding of, what is effecting your wellbeing, it's not something you would do for your work base, because you wouldn't trust to that level of search and research.*

[Period of silence]

Patient: *Well, I haven't thought about it like that, it's something I'd have to reflect on.*

...

MI: *And I just wonder whether it might be useful to have a thought about that, and think whether that might be a conversation worth having, where we can have that open discussion, and you can understand where Barbara, as a dietician, is coming from.*

Patient: *Well, I'm certainly open to listening.*

...

Dietician: *So, the glycaemic index is a measurement really, it goes from zero to 100. The highest food you'll get for a glycaemic index is pure sugar. Because that raises your blood sugar the quickest. Yeah? So, anything lower in the glycaemic index, will raise your blood sugar much slower, it'll keep you feeling fuller for longer, and those are the types of carbohydrates that we really should be focusing more on, and as a population we could all do with including more low GI foods in our diet.*

Patient: *Aha.*

Dietician: *So, types of low GI foods are things like wholemeal granary breads, wholemeal pasta, wholemeal rice, beans and pulses as well.*

Patient: *Now, that's something I would try. I'm really not keen on the bread and the pasta and the rice, but beans and pulses, I could incorporate that into a vegetable based meal.*

Nature of the disagreement: This example contains two instances of disagreement: an overall disagreement about whether or not the patient should eat more carbohydrates, and a second, embedded disagreement about taking advice.

The overall disagreement came about because the dietician recommended to the patient that they eat carbohydrates as part of a balanced diet; however, the patient believes that all carbohydrates leave them feeling bloated and as such they are not prepared to take on board the advice.

While attempting to resolve this overall disagreement, the motivational interviewer discovered that the patient does their own research using Google, and is prepared to believe that advice over what they are told by healthcare professionals.

How is the disagreement expressed? This is an example of very explicit disagreement: on several occasions the patient outright refuses to accept they should eat more carbohydrates. The embedded disagreement was also highly explicit, with the statement that “*Google tells me that diabetes is due to sugar intake*” when asked if they see the connection between low blood sugar and not eating carbohydrates. Furthermore, they used the expression “agree to disagree” in an attempt to stymie further discussion around the subject because their view was so deeply entrenched.

Resolution strategy: With the patient unmoving on the subject of eating carbohydrates, the motivational interviewer first tries to link their “low-carb” diet to their blood sugars dropping; this however deepens the disagreement because it reveals that the patient has done their own research via Google, and is believing that over what they are being told.

The motivational interviewer decides to therefore adopt a different approach: he steers away from the general conflict to something, seemingly, unrelated – asking the patient how they go about their job. He finishes this ostensible diversion by putting the patient’s view back at them – that is someone could do their job just by using Google (in the same way the patient believes they can do the medical professionals’ jobs).

By disengaging from the disagreement, the motivational interviewer encouraged the patient into reflecting on their stance. Using the patient’s own job as a comparison to demonstrate by they should

listen allowed the motivational interviewer led them to reflect on their position. This in turn allowed the dietician to rejoin the discussion, providing new information that allowed the overall disagreement to be resolved.

5.3 Addressing frustration head-on

Our final example of disagreement stems from a patient feeling frustrated at the amount of advice she has been receiving. Having previously been advised to follow a certain diet, she has now been told she needs to take more care of her feet. This led to the following exchange:

Patient: *Well, I mean, I'm trying to follow a diet. You know, the doctor's given me some sheets and things so I'm trying to do that but...*

Podiatrist: *I know, it's difficult.*

Patient: *...it's hard and it's...when you're out on the road all day, I mean, I am taking my lunch and things but, you know, you stop for petrol, you pop into the garage, it's easy to think, oh, I'll just pick up a quick snack...*

Podiatrist: *Yes. Yes.*

Patient: *...and what you think is healthy, you know, it's got these, sort of, oh, I can have something with nuts in it and then you look at the calories or the whatever, the carbohydrate I'm supposed to look at, aren't I...*

Podiatrist: *Yes. Yes.*

Patient: *...and if...you know, and you're like, oh, great, it's really high. Well, I've got it now, I'm going to eat it anyway.*

General Practitioner (GP): *Can I...just, Linda, can I ask, you don't look very happy with what [the podiatrist] is saying. You look a bit pee'd off, actually. Is that fair to say?*

Patient: *Well, yes, I just feel like there's...I feel like you've taken everything away. You know, like I'm dieting. I'm supposed to be not drinking. I'm supposed to be cutting out smoking and here*

you are now telling me that I'm going to have something else to do.

GP: *Yes.*

Patient: *It's like, oh my god, this just overtakes your life.*

GP: *Yes.*

Nature of the disagreement: This disagreement stemmed from the patient's frustration at being given advice regarding how to take care of their feet. Having previously been given a lot of previous medical advice, this prompted the patient to express significant frustration at everything they have been told to do.

How is the disagreement expressed? While the disagreement itself is not made explicit, the patient's vocal expression of their frustration clearly demonstrates a reluctance to follow the advice being given.

Resolution strategy: The overall strategy adopted here was to first give the patient space to express their frustration, before then allowing them to vocalise the precise root cause of their issue.

Initially, the podiatrist gave the patient space: interjecting only occasionally to say "yes, yes". This however could not continue indefinitely, so the GP intervened and explicitly asked the patient if they were "p'eed off". This provided an opening for the patient to vocalise their general problem, that is the feeling of being overwhelmed and that adhering to medical advice "overtakes [their] life".

Was the resolution successful? Partly; while the GP's intervention prompted the patient to express their overall frustration more clearly and succinctly, it did not address the overall issue of requiring the patient to take more care of their feet. Nevertheless, it was still important to identify the patient's general complaint of being overwhelmed.

6. PRELIMINARY FINDINGS

In order to build a substantive theory of conflict in healthcare settings, we are following the stages of grounded theory as described above. We describe preliminary findings on the illustrative examples described above, which were selected as being interesting examples of conflict in healthcare.

Open coding: In the first stage, we identified and coded code definitions of interest via open coding. In example 1, for instance, we highlighted the use of humour, qualifiers (“seems”, “a bit”) and gestures (raised eyebrows, smiling) in expressing the conflict, which all helped to keep the setting friendly and non-confrontational. Example 2 showed examples of conflict both as goal (how best to manage low blood sugar) and process (what is the best way to learn about diabetes management). Example 3 showed the resolution strategy of gentle listening and agreeing, with both practitioners using the strategy to sooth the patient's frustration.

Axial coding: In the second stage, we identified interrelationships between our codes and formed concepts by combining codes, to describe repeated patterns of interactions and conflict solving strategies in the conversation. Here, for instance, Example 1 showed the method of retreat as a resolution strategy (an open code) -- seen here via the concept of unequal power dynamics (senior male and junior female). We also see the concept of language, currently with the single qualifiers code. We also see how conflict can be used to identify a lack of understanding or knowledge, and if handled in a productive way, can play a role as trigger for negotiation and deliberation.

Selective coding/Theory building: In the third and fourth stages, we grouped concepts into categories, making connections across codes, categories, and concepts. The following main categories emerged: *conflict type, conflict expression, resolution and value*.

We show a visual representation of how these categories relate to each other and to different kinds of conflict in Figure 1.

The three examples of disagreement we have provided were examined primarily based on initial observation while the sessions were being recorded. While this gives us an overall view of the manifestation and eventual resolution of disagreement in a healthcare context, we do not assume that it provides a full picture.

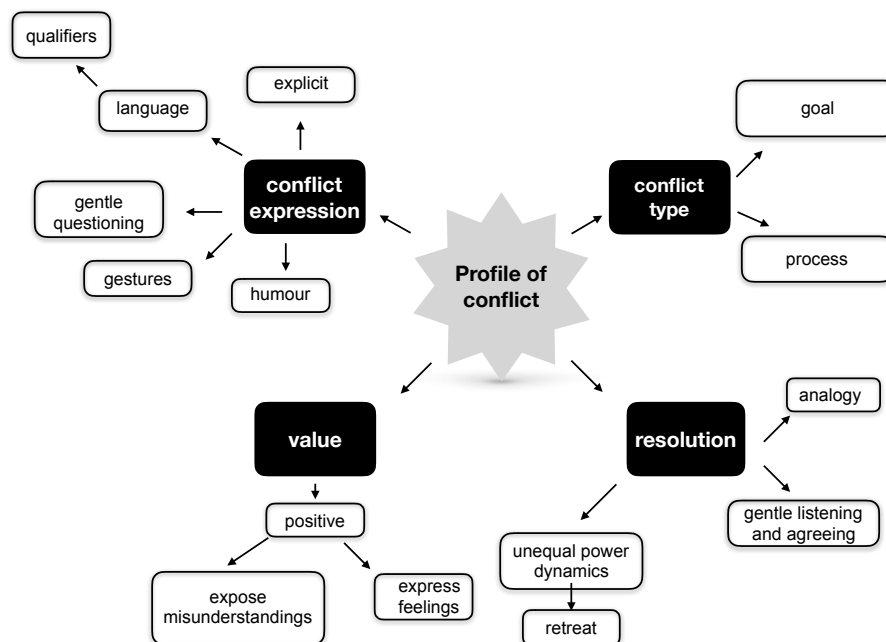


Figure 1– Preliminary analysis of our three illustrative examples

While this work is still in preliminary stages, the analysis above shows how we might begin to answer our research questions: *“In what contexts does disagreement occur and how is it expressed in those contexts?”*, and *“Can we identify resolution strategies for different types of disagreement?”*. The contexts so far in our data show that it arises both as a disagreement over a particular goal, or how that can best be satisfied, as well as questions over what the best way to do something. It is expressed explicitly as well as in more subtle ways, with qualified language. Gestures and humour are also important here. In our examples the resolution strategies all avoided direct confrontation, and involved use of analogies (which guides the conversation away from the conflict) and retreating.

Our pilot study also suggests further research questions, such as *“Is there a mirroring of disagreement type (explicit vs. subtle) and the strategy for effective resolution?”*; *“What positive outcomes arise from expressing disagreement?”*, *“Are there any negative outcomes from expressing disagreement?”*, and *“How is silence used to express or resolve conflict?”*. The effect of power dynamics also merits further investigation.

7. CONCLUSION

Understanding disagreement in a medical domain is important if such disagreements are to be successfully and effectively resolved. In this paper, we presented three examples of such disagreement taken from a corpus of mock patient consultations.

Our first example showed a subtle disagreement between two medical professionals: a general practitioner (GP) questioned a dietician over cutting chocolate out of a diet. This disagreement was borne out of intuition rather than knowledge, demonstrated by the tone of surprise and element of humour the GP injected into his question. The dietician subsequently agreed that cutting chocolate out completely “seems...harsh”. Resolving this disagreement was seeded by the disagreement itself: by expressing it in a casual way, the GP left it open for the dietician to reaffirm her view without any consequences.

The second example consists of an overall disagreement, about whether or not the patient should eat carbohydrates, and a second, embedded disagreement about taking on board advice. This disagreement was largely borne out of the patient’s “know-it-all” personality and as such took significant time and effort to resolve. The resolution of the embedded disagreement was led by the motivational interviewer, who used the patient’s job to encourage them to reflect on their viewpoint. This helped lead into the dietician resolving the overall conflict by providing the patient with new knowledge that allowed them to better understand types of carbohydrate that are good for them.

We have presented a pilot study into disagreement in healthcare: reporting our work on collecting data of both intra-team and patient-team disagreements, and our initial coding for thematic analysis to develop main categories and concepts. We now plan to conduct a full coding in which all data is classified according to these categories and concepts, and then develop and evaluate our theory.

ACKNOWLEDGEMENTS: This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant number #769553. This result only reflects the authors’ view and the EU is not responsible for any use that may be made of the information it contains. The authors are also grateful to the anonymous reviewers and the audience in our talk for their valuable feedback.

REFERENCES

- Almost, J., Wolff, A. C., Stewart-Pyne, A., McCormick, L. G., Strachan, D., & D’Souza, C. (2016). Managing and mitigating conflict in healthcare teams: an integrative review. *Journal of Advanced Nursing*, 72(7), 1490-1505.

- Gaba, D. M. (2004). The future vision of simulation in health care. *BMJ Quality & Safety in Health Care*, 13(Suppl 1), i2-i10.
- Greer, L. L., Saygi, O., Aldering, H., & de Dreu, C. K. (2012). Conflict in medical teams: opportunity or danger? *Medical Education*, 46(10).
- Janier, M., & Reed, C. (2016). Corpus resources for dispute mediation discourse. *Proceedings of the 10th Language Resources and Evaluation Conference (LREC 2016)*, (pp. 1014-1021). Portoroz.
- Kakavá, C. (2001). Discourse and Conflic. In D. Schiffrin, D. Tannen, & H. E. Hamilton, *The Handbook of Discourse Analysis*. Blackwell Publishers Ltd.
- Pollock, J. (1995). *Cognitive Carpentry*. Cambridge, MA: The MIT Press.
- Strauss, A., & Juliet, C. (1994). Grounded theory methodology: An overview. In N. Denzin, & Y. Lincoln, *Handbook of Qualitative Research (1st ed)*. Thousand Oaks, CA: Sage Publications.

Explaining away and the propensity interpretation of probability: the case of unequal priors

ALICE LIEFGREEN

*Department of Experimental Psychology,
University College London
alice.liefgreen.15@ucl.ac.uk*

MARKO TEŠIĆ

*Department of Psychological Sciences,
Birkbeck University
mtesic02@mail.bbk.ac.uk*

Explaining away is a pattern of inference that occurs in situations where independent causes compete to account for an effect. Empirical studies have found that people ‘insufficiently’ explain away. In this paper we explore whether this insufficiency could be partly due to people’s different interpretations of probabilities. In particular, we tested people on reasoning tasks involving unequal priors of causes and we provide evidence indicating that some people may interpret probabilities as propensities, which would then drive the insufficiency effect of explaining away.

KEYWORDS: Causal Bayesian networks; Causal inference; Diagnostic reasoning; Evidential reasoning; Explaining away; Probability interpretation; Propensity

1. INTRODUCTION

Judgments and inferences that are reliant on beliefs about how events or items of information are causally related to each other are extremely ubiquitous in people's daily and professional lives. The vast majority of these causal judgments occur *under uncertainty*. Consider for example a scenario in which a social worker is trying to ascertain whether action should be taken to remove a child displaying bruises from the custody of his parents under the suspicion that he is being physically abused. The social worker, however, knows from her experience that the bruises could also be the product of a blood disorder termed ‘haemophilia’. Upon

observing the bruises, she should then *increase* the probability of each potential cause as bruises are indicative of each one. After a medical test, the social worker learns that the child definitely suffers from haemophilia. Given this new piece of information, she should now *decrease* the probability of the child being physically abused, since haemophilia is sufficient to explain the bruises. If, however, the medical test had revealed that the child definitely *did not* suffer from haemophilia, then the probability of him being physically abused would *further increase* as a result. This scenario illustrates a pervasive pattern of reasoning known as ‘explaining away’. In more general terms, explaining away occurs in situations in which multiple independent causes (e.g. physical abuse and haemophilia) compete to explain a common effect (e.g. bruises). After observing the occurrence of the effect, the probability of the two causes increases (step 1). Subsequently, after learning of the occurrence of one the probability of the alternative cause(s) decreases (step 2a). If, conversely, we learned that a cause *did not* happen, the probability of the other cause(s) further increases (step 2b).

1.1. Explaining away: Normative account

Over the past few decades, patterns of inference in causal reasoning such as explaining away have been modelled in the cognitive sciences utilising graphical models called ‘Causal Bayesian Networks’ (CBNs). These can be used to represent probabilistic knowledge in a graphical manner (for overview see Pearl, 2009; Neapolitan, 2003).

The computational machinery of CBNs, grounded in probability theory, allows one to perform exact quantitative computations of the probability of any random variable(s) in the network being present/absent given the presence/absence of any other variables.

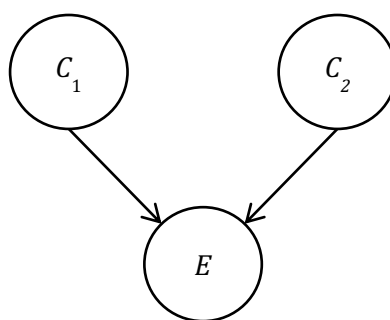


Figure 1: A CBN model of explaining away

Consider the graph in Figure 1, consisting of three nodes representing three random variables: two causes, C_1 and C_2 , and one common effect, E ¹. Situations involving explaining away can be modelled utilising common-effect CBNs such as the one in Figure 1 (see Pearl, 1998; 2009). For example, we could model the aforementioned example by representing physical abuse as C_1 , haemophilia as C_2 and finally the bruises on the body as E . The two causes are (unconditionally) independent when we do not know whether the child has bruises on his body or not, which follows our intuitions that physical abuse and haemophilia cannot probabilistically influence each other, *before* learning anything about the bruises. However, this is dependent on the network parameterization.

In order for common-effect CBNs to lead to the pattern of explaining away described earlier, they need to be parameterized such that the following inequality holds (see Wellman, 1993):

$$\frac{P(E|\neg C_i, \neg C_j)}{P(E|\neg C_i, C_j)} < \frac{P(E|C_i, \neg C_j)}{P(E|C_i, C_j)} \quad (1)$$

for $i, j \in \{1, 2\}$. From Inequality (1) it follows (see Griffiths, 2001; Morris & Larrick, 1995):

$$P(C_i|E, C_j) < P(C_i|E) < P(C_i|E, \neg C_j) \quad (2)$$

The inequalities in (2) comply with the general intuition of explaining away mentioned above and serve as a definition of explaining away in the empirical research outlined in the present paper (see also Rehder & Waldmann, 2017; Rottman & Hastie, 2016).

1.2. Explaining away: Empirical account

Despite its ubiquity in human reasoning (Kelley, 1973; Pearl, 1988; Rottman & Hastie, 2016), empirical research on explaining away in the psychological sciences adopting the constrained definition outlined by the inequalities in (2) is somewhat limited and has insofar yielded mixed findings (for an overview see Rottman & Hastie, 2014).

Overall however, it appears that human explaining away inference, even in simple three-node common-effect causal structures (as

¹ Throughout the paper all random variables are binary: a random variable X (denoted by italicized letters) can take exactly one of the two values X or $\neg X$ (denoted by non-italicized letters), where X indicates that X is present and $\neg X$ indicates that X is absent.

in Figure 1), is fallible, thus emphasizing the significance of further investigating this evasive phenomenon.

Most of the studies exploring explaining away have reported that people explain away insufficiently or not at all (Davis & Rehder, 2017; Fernbach & Rehder, 2013; Liefgreen, Tesic & Lagnado, 2018; Morris & Larrick, 1995; Rehder & Waldmann, 2017; Rottman & Hastie, 2016; Sussman & Oppenheimer, 2011) or in some cases even display behaviour directly opposite to that of explaining away: $P(C_i|E, C_j) > P(C_i|E, \neg C_j)$ (Ferbach & Rehder, 2014; Rehder, 2014a) or $P(C_i|E, C_j) > P(C_i|E)$ (Rottman & Hastie, 2016, Experiment 1a).

1.3. Limitations of previous studies

Although the empirical studies on explaining away insofar speak to the robustness of people's deviation from the normative model, it is worth mentioning some limitations that are commonly found in these studies.

Firstly, the majority of studies neither conveyed to nor elicited from participants the prior probabilities of causes (see Rottman & Hastie, 2014), rendering any comparison to a normative model problematic. In most cases, priors indirectly dictate the amount of explaining away found in the normative model (see Morris & Larrick, 1995): lower priors imply a larger amount of explaining away than higher priors. As really high prior probabilities lead to minimal amounts of explaining away in the normative model, even if participants adopted the priors given to them and engaged in the correct pattern of inference, explaining away would most probably remain undetected. In the present work we : (i) provided participants with explicit priors and subsequently re-eliciting these to ensure they have been accepted, (ii) utilised low priors to maximise the amount of explaining away in the normative model and facilitate its detection and (iii) assigned different (low) priors to the two causes in the model to vary the amount of explaining away.

Secondly, the majority of studies exploring explaining away in common-effect structures report a violation of the Markov condition of independence, i.e. $P(C_i|C_j) \neq P(C_i|\neg C_j)$ (Rehder, 2014a, 2014b; Rehder & Burnett, 2005). In these cases, participants are misconstruing the two causes to be initially dependent, typically by assuming they are positively correlated. This is problematic since the higher the degree of positive correlation, the lower the normative amount of explaining away, with very high degrees of positive correlation potentially leading to a pattern opposite to explaining away (see Morris & Larrick, 1995). In order to guard ourselves against potential violations of the independence assumption, in the present work we: (i) explicitly stated that the two causes are independent, (ii) utilised cover stories that intuitively minimized participants' inclination to view the two causes as

unconditionally dependent, and (iii) utilised qualitative relational questions to investigate people's understanding of independence.

Finally, despite explaining away being a relational concept, the majority of empirical studies on explaining away elicit participants' belief estimates in isolation and do not investigate whether participants understand the relational nature of this pattern of reasoning. To rectify this issue, in the present work we complement quantitative questions asking for numerical probability estimates of, for example, $P(C_i | E, C_j)$, with qualitative relational questions asking them to consider whether $P(C_i | E, C_j)$ is less than, greater than, or equal to $P(C_i | E)$.

2. MOTIVATIONS FOR PRESENT WORK

In our previous study (Liefgreen et al., 2018), we tried to address some of the above-mentioned methodological issues often found in empirical studies on explaining away. Despite concluding that participants accepted priors of causes and did not violate the assumption of independence, we still observed insufficient explaining away. Moreover, a large cluster of participants did not update the probabilities of causes from their priors, given the presence of the effect or even given the presence of the effect and the other cause. This together with participants' explanations of the way they updated the probabilities led us to hypothesize that participants in this cluster may be interpreting probabilities differently, or more specifically, as propensities.

2.1. *Probability interpretations*

A large number of studies exploring human reasoning under uncertainty implicitly or explicitly assume the subjective probability interpretation where probabilities are identified as degrees of belief of a particular person about a certain event occurring. However, in philosophy of statistics one finds a whole spectrum of probability interpretations, one of which is the propensity interpretation (Popper, 1959; Giere, 1973). According to this interpretation probabilities are propensities (or tendencies and dispositions) of a particular physical system to produce an outcome (Hajek, 2012). For example, the statement that the probability of a coin to land Heads equals $\frac{1}{2}$ is equivalent to the statement that there is a coin tossing set-up and that on a particular trial the strength of the propensity for this coin to land Heads is $\frac{1}{2}$. This propensity is objective, it is part of the physical world, and it does not depend on our subjective beliefs about the coin landing Heads.

How does this relate to explaining away? Imagine a situation where there are two coins tossed at the same time, each with a coin bias

of $\frac{1}{5}$ for Heads. In this set-up there is also a light bulb that will turn on if at least one coin lands Heads. Here, it is perfectly natural to ask about the propensity for the light bulb to turn on if Coin 1 landed Heads, i.e. $P(E|C_1)$. However, the propensity of Coin 1 to have landed Heads given that the light bulb turned on is simply the original propensity for Coin 1 to land Heads: whether or not the light bulb turns on does not affect the propensity/the coin bias of Coin 1 to land heads, i.e. $P(C_1|E) = P(C_1) = \frac{1}{5}$.² In the same vein, according to the propensity interpretation, observing the effect (or another cause) would not change the propensity of the cause in question to happen. This implies that people who interpret probabilities as propensities in explaining away situations will violate the normative account by not updating their estimates given the presence of the effect, or the presence of the alternative cause (i.e. they would be repeating the priors).

This behaviour could therefore be partly driving the insufficiency observed in empirical studies of explaining away as repeating the priors would drive the average sample estimate away from the normative one. This seems increasingly plausible in light of the psychology literature suggesting that people may be able to distinguish between different variants of uncertainty, one of which is propensity (see Fox & Ülkümen, 2011; Kahneman & Tversky, 1982), and studies suggesting that people are sensitive to different probability interpretations (Ülkümen, Fox, & Malle, 2016) and may in fact be thinking of probabilities as propensities (Keren & Teigen, 2001).

3. EXPERIMENT OVERVIEW

The main aim of the present experiment was to empirically test whether propensity interpretations of probability partly drive the observed deviation of people's explaining away inferences from the normative ones. We adopted a novel experimental design that addressed the methodological confounds employed by previous studies and manipulated the properties of cover stories within which we embedded our CBN. In our experiment, all participants were required to reason with the same three-node common-effect structure depicted in Figure 1, parameterized such that causes had unequal low priors ($P(C_1) = 0.2$ and $P(C_2) = 0.1$) to increase the normative amount of explaining away in the model. Moreover, we utilised a deterministic setup wherein the

² This intuition has been (formally) outlined in Humphreys (1985), who employs it to argue that propensities are inconsistent with probabilities. This inconsistency is commonly known as 'Humphreys' paradox' in the literature.

presence of one cause entailed the presence of the effect ($P(E|C_1, C_2) = P(E|C_i, \neg C_j) = 1$, and the absence of both causes entailed the absence of the effect ($P(E|\neg C_1, \neg C_2) = 0$).

To test whether the propensity interpretation affected people's judgements on inferences relating to independence of causes, diagnostic reasoning, and explaining away, we manipulated properties of cover stories that, to a larger or a lesser extent, accentuated the propensity interpretation.

3.1. Manipulating cover stories

We embedded our common-effect structure within three different cover stories: one involving coin-tossing, one involving balls and containers, and one involving a dinner party.

In the coin-tossing cover story, the two causes (C_1 and C_2) were represented by two coins (binary variables; either Heads or Tails) tossed with the probability p_i for Heads by two coin-tossing mechanisms in separate rooms. If at least one coin landed Heads, a light bulb (common effect) stored in a different unit would switch on. From the propensity interpretation point of view, p_i is the propensity for a coin to land Heads given a coin-tossing set-up and that propensity does not change whether or not the light bulb (i.e. the effect) is on or off. As the questionnaire prompted participants to answer diagnostic reasoning and explaining away questions pertaining to the coins (see Section 4 below), we argue that the propensity interpretation would be strongly pronounced in this scenario.

In the balls and containers cover story the two causes were represented by two balls (binary variables; either copper or rubber) randomly selected from independent containers and placed on two gaps in an electric circuit. If at least one of the two balls was copper, a light bulb in the circuit (common effect) would turn on. Here, we follow Giere (1973) in arguing that the propensity is still present in this set-up, but it is at the level of a random sampling mechanism, not at the level of balls. As we prompted participants to questions pertaining to the balls and not to the random sampling mechanism, we argue that the propensity interpretation is less pronounced in this cover story compared to the coin-tossing one.

Finally, in the dinner party cover story the two causes were represented by two individuals, Michael and Tom, and the common effect was represented by a third individual, Helen, who would drink wine only if at least one of the two aforementioned people brought wine to a dinner party ('Helen' was a binary variable; either 'drinking wine' or 'not drinking wine'). In this set-up, the probability p_i of whether a person brings wine to the party was determined purely by host's subjective

estimates, implying that in this scenario the propensity interpretation is the least pronounced (if at all present).

Given the above rationale, we predicted that the proportion of participants whose reasoning aligns with the propensity interpretation, i.e. who would respond $P(C_i|E) = P(C_i|E, C_j)$ would be the highest when reasoning with the coin-tossing cover story, smallest when reasoning with the dinner party cover story, and fall in between these when reasoning with the ball containers cover story.

4. METHODS

4.1. Participants and design

A total of 271 participants ($N_{\text{MALE}} = 111$, 4 identified their gender as other; $M_{\text{AGE}} = 32.2$ years, $SD = 10.2$) were recruited from Prolific Academic (www.prolific.ac). All participants were native English speakers who gave informed consent and were paid £1 for partaking in the present study, which took on average 10.8 minutes ($SD = 5.4$) to complete. Eight participants were excluded as they did not pass the attention check, leaving a total of 263 participants in the analyses.

A between-subjects design was employed, and participants were randomly allocated to one of the three groups which differed in the cover story they were required to reason with. Group 1 ($n=87$) was presented with the coin-tossing cover story; Group 2 ($n=87$) with the ball containers cover story and Group 3 ($n = 89$) with the dinner party cover story.

4.2. Materials

Each of the three groups was asked to complete the same inference questionnaire ($N_{\text{QUESTIONS}} = 12$) comprising of the questions outlined in Table 1. Although all participants completed this inference questionnaire, as mentioned in Section 4.1, participants in each group were required to reason with different cover stories, within which we embedded the common-effect structure. For cover story details see Section 3.1 and for full details on the inference questionnaire visit Open Science Framework, <https://osf.io/zm6ec/>.

Table 1: Inference types and questions found in questionnaire.

Question number	Inference Type	Key Inferences	Question Type
1	Priors	$P(C_1)$	Quantitative
2		$P(C_2)$	Quantitative
3	Independence	$P(C_2 C_1)$	Qualitative

4		$P(C_1 \neg C_2)$	Qualitative
5,6	Diagnostic	$P(C_1 E)$	Qual. + Quant.
7,8	Reasoning	$P(C_2 E)$	Qual. + Quant.
9, 10	Explaining Away	$P(C_1 E, C_2)$	Qual. + Quant.
11,12	Logic ³	$P(C_1 E, \neg C_2)$	Qual. + Quant.

4.3. Procedure

Participants in each of the three groups were initially presented with the pertinent cover story and were given explicit information on the common-effect model embedded within the cover story including the prior probability of each cause, and the causal relationships within the model. This was done in both textual form and through a graphical representation. Participants were provided with a textual account by which each cause could independently bring about the common effect. Subsequently, participants were presented with the inference questionnaire (for questions and associated inferences see Table 1).

Questions marked as quantitative in Table 1 required participants to provide numerical estimates on a slider (scale of 0-100 %). Questions marked as qualitative required participants to select one of three options: the probability increases, decreases, or stays the same when asked about e.g. $P(C_2|C_1)$ given no knowledge about the state of E. To investigate participants' diagnostic and explaining away reasoning we employed both qualitative and quantitative question formats. This enabled us to capture the relational nature of explaining away, and retain focus on the direction and magnitude of change of beliefs given certain evidence. Additionally, in order to better understand participants' reasoning, some questions prompted participants to provide written explanations for their answers.

5. RESULTS

Participants' answers to all quantitative questions in the inference questionnaire are graphically represented in Figure 2. The results section will be sub-divided by analyses carried out for each inference type.

³ We have labelled questions 11 and 12 as 'logic' questions, since our set-up is deterministic and learning that one cause did not happen, whilst knowing that the effect happened, entails (by logic) that the other cause must have happened, i.e. $P(C_1|E, \neg C_2) = 1$.

5.1. Prior probabilities and independence of causes

Within each group we obtained the percentage of people who correctly answered⁴ both questions on prior probabilities of causes (Q1 and Q2 in Table 1). Within Group 1 this was 88.5% of participants, within Group 2, 77% and within Group 3, 72%. Additionally, we obtained the percentage of people who correctly answered *both* questions regarding the independence of causes (Q3 and Q4 in Table 1). Within Group 1, this was 88.5%; within Group 2, 87.4%; and within Group 3, 91%. These high percentages illustrate that overall participants accepted the priors of causes that the experimenters explicitly stated and correctly regarded the causes as initially independent in all groups.

5.2. Logic

Independent analyses were conducted on qualitative and quantitative 'logic' questions (Q11 and Q12 in Table 1).

5.2.1. Qualitative

The percentage of participants who correctly answered the qualitative logic question was 62.1% in Group 1, 81.6% in Group 2 and 74.3% in Group 3. A Chi-Square test of independence illustrated these proportions significantly differed, $\chi^2(2) = 8.5, p = 0.001$. Bonferroni corrected post-hoc pairwise comparisons illustrated the only significant difference to be between the proportions of Group 2 and Group 3, $p = 0.004$.

5.2.2. Quantitative

The percentage of participants who correctly answered the quantitative logic question was 67.8% in Group 1, 81.6% in Group 2 and 70.8% in Group 3. A Chi-Square test of independence illustrated no significant difference in these proportions, $\chi^2(2) = 4.7, p = 0.09$. These results suggest that participants correctly understood the deterministic set-up of our experiment, in contrast to, for instance, Rottman and Hastie (2016).

⁴ Answers to quantitative questions were coded as correct (1) if they were $\pm 2\%$ of normative answers, otherwise they were coded as incorrect (0).

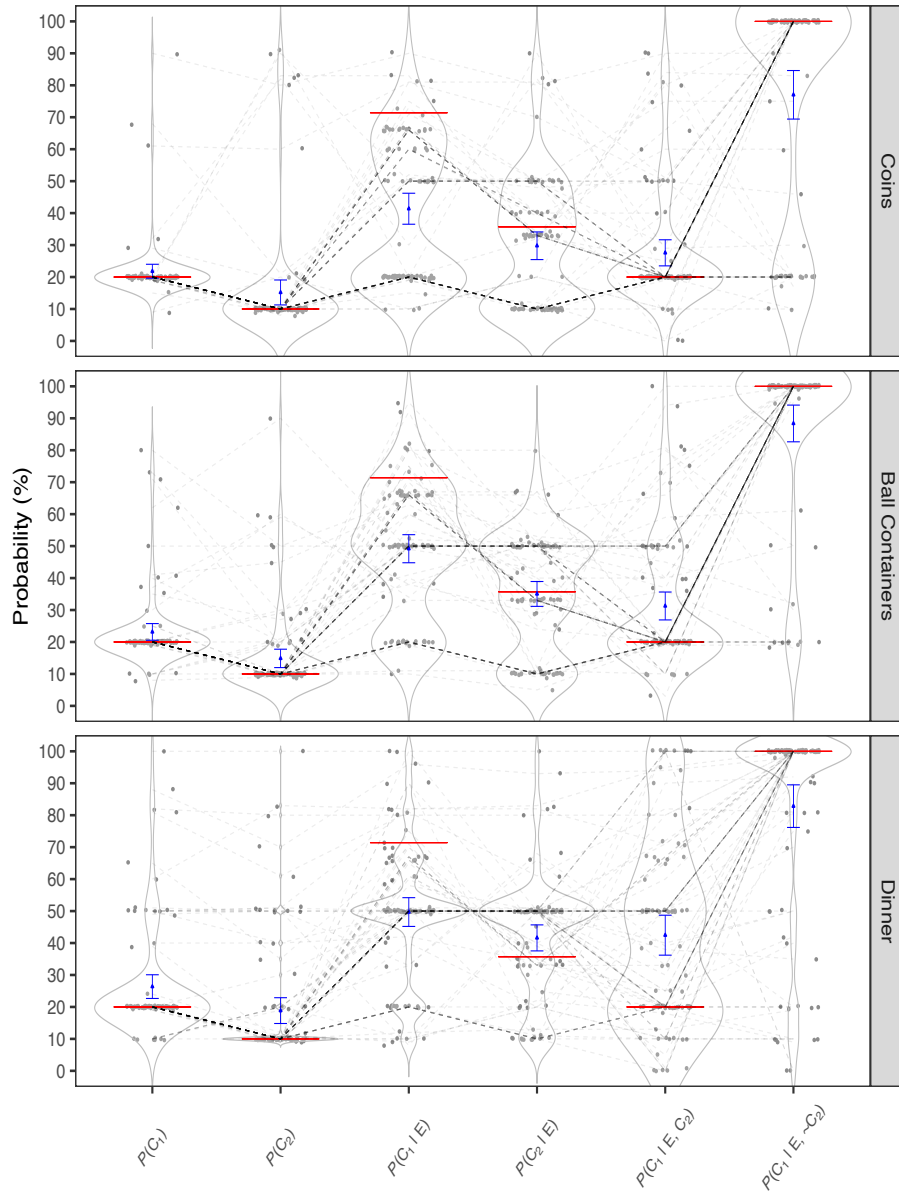


Figure 2: Participants' responses to quantitative questions. Red lines are normative answers. Blue dots are empirical averages with 95% confidence intervals as error bars. Dotted lines depict how participants changed their probability estimates from one questions to another, with darker lines indicating more participants changing the probabilities in the same way

5.3. Explaining away: Relational concept

Given the relational nature of explaining away, to better investigate participants' updating behaviour across this pattern of inference, we conducted aggregate analyses on questions pertaining to diagnostic reasoning, explaining away and logic (Q5-12 in Table 1). Independent analyses were conducted on qualitative and quantitative relational explaining away questions.

5.3.1. Qualitative

Given that we required participants to make two qualitative diagnostic reasoning inferences, i.e. $P(C_1|E)$ and $P(C_2|E)$, if a participant answered both questions regarding qualitative diagnostic reasoning correctly, we coded the response as 1; otherwise 0.

The proportion of participants who correctly answered all qualitative questions pertaining to relational explaining away was 28.7% in Group 1, 33.3% in Group 2 and 21.3% in Group 3. A Chi-Square test of independence illustrated no significant difference between these proportions, $\chi^2(2) = 3.2, p = 0.2$ suggesting that participants across all conditions performed relatively poorly on qualitative relational explaining away.

5.3.2. Quantitative

In regard to quantitative relational explaining away, we analysed questions relating to the updating of C_1 , namely $P(C_1|E)$, $P(C_1|E, C_2)$ and $P(C_1|E, \neg C_2)$.

A repeated-measures ANOVA with a Greenhouse Geisser correction was carried out on the average probability estimates on the relational explaining away questions, within each of the groups (see Figure 3). Results illustrated a significant difference between these estimates within Group 1; $F(1.59, 122.6) = 95.6, p < 0.0001$, within Group 2; $F(1.8, 140) = 167.5, p < 0.0001$ and within Group 3; $F(1.6, 126) = 57, p < 0.0001$. Post-hoc paired t-tests allowed us to obtain the 95% confidence intervals (CI) of the difference in participants' average probability estimates between pairs of inferences of interest (see Table 2 below).

Since participants in all groups under-adjusted their belief estimates (i.e. the normative difference was not included in any of the 95% CI of the empirical differences) we were able to conclude that there was insufficient explaining away in all groups.

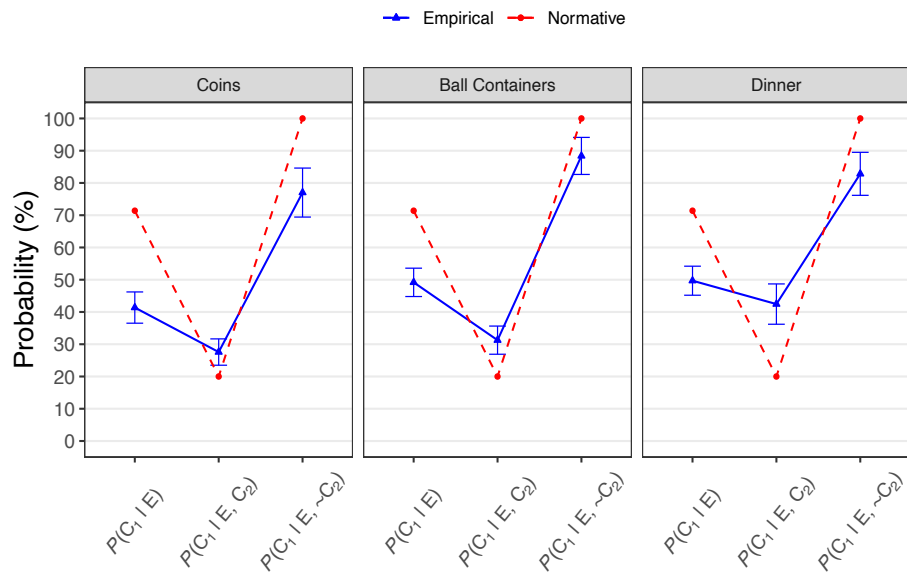


Figure 3: Participants' quantitative relational explaining away responses. Error bars are 95% confidence intervals.

Table 2: Within group explaining away.

Inferences	Normative difference	Empirical difference	95% CI of empirical difference
<i>Group 1</i>			
A-B	36	13.8	[8.5, 19]
C-B	80	49.4	[40.7, 58.2]
<i>Group 2</i>			
A-B	36	17.9	[12.6, 23.3]
C-B	80	57.1	[49.8, 64.4]
<i>Group 3</i>			
A-B	36	7.2	[1.5, 13]
C-B	80	40.4	[31.9, 48.9]

Note: A: = $P(C_1|E)$, B: = $P(C_1|E, C_2)$, C := $P(C_1|E, \neg C_2)$.

5.4. 'Stay the same'

To test the hypothesis that participants in certain groups would be more prone to interpret probabilities as propensities, we obtained the proportions of participants in each group who did not update from their priors in diagnostic reasoning and explaining away questions (Q5-10 in Table 1). Independent analyses were conducted on qualitative and quantitative inferences in this section.

5.4.1. Qualitative

We computed the proportion of participants who selected the 'stay the same' option to both diagnostic reasoning and to explaining away qualitative questions, i.e. after being asked about $P(C_1|E)$, $P(C_2|E)$ and $P(C_1|E, C_2)$ (see Figure 4). In Group 1, this was 44.8%, in Group 2, 33.3% and in Group 3, 23.6%. A Chi-Square test of independence illustrated that these proportions significantly differed, $\chi^2(2) = 8.87, p = 0.013$. Bonferroni corrected ($\alpha = 0.017$) post-hoc pairwise comparisons showed the only significant difference to be between the proportions of Group 1 and Group 3, $p = 0.005$.

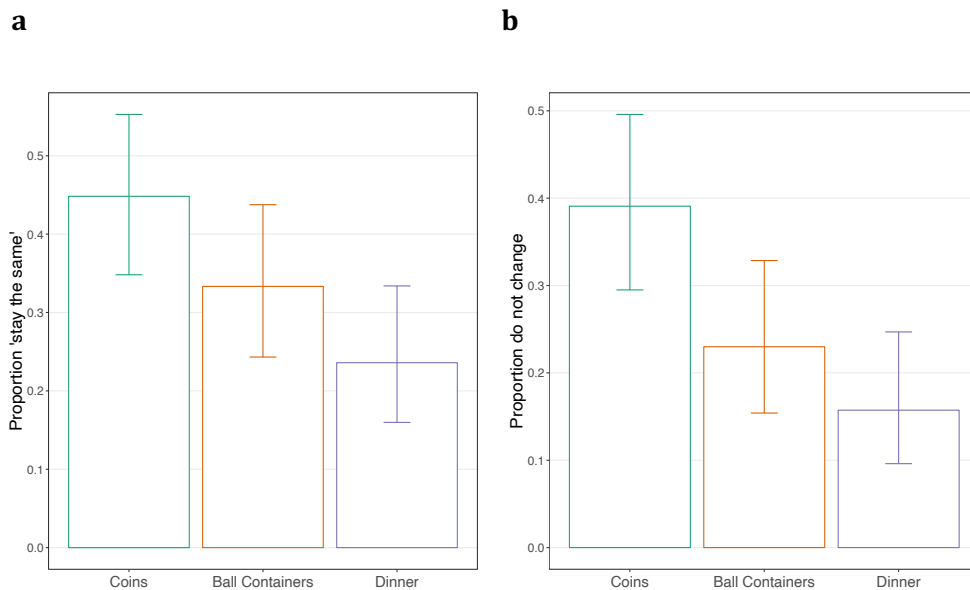


Figure 4. (a). The proportions of participants who chose 'stay the same' option on qualitative Q5, Q7, and Q9 and (b) who did not change their estimates in Q6, Q8, and Q10 compared to their own stated priors. Error bars are 95% confidence intervals.

5.4.2. Quantitative

We obtained the proportion of participants who did not update from their own stated priors on both diagnostic reasoning and explaining away quantitative questions (see Figure 5). In Group 1, this was 39.1% of participants, in Group 2, 23%, and in Group 3, 15.7%. A Chi-Square test of independence illustrated these proportions significantly differed, $\chi^2(2) = 13.07, p = 0.0014$ post-hoc pair-wise comparisons ($\alpha = 0.017$) showed the only significant difference to be between the proportions of Group 1 and Group 3, $p = 0.0009$.

Overall, these results are in support of our hypothesis as Group 1 'stayed the same' significantly more, in both qualitative and quantitative inferences, compared to Group 3, with Group 2 falling in between.

5.5. Explaining away: excluding 'stay the same'

In this section we explored whether participants interpreting probabilities as propensities accounted for the observed insufficiency in explaining away. For qualitative inferences, we ran the same analysis on explaining away as in Section 5.4, but removing the set of participants who answered 'stay the same' to all qualitative questions relating to $P(C_1|E)$, $P(C_2|E)$, and $P(C_1|E, C_2)$. For quantitative inferences we removed the set of participants who did not change their quantitative explaining away estimates regarding $P(C_1|E)$, $P(C_2|E)$ and $P(C_1|E, C_2)$ compared to their stated priors.

5.5.1. Qualitative

Within the new subset, the proportion of participants who correctly answered all qualitative relational explaining away questions was 52.1% in Group 1, 50% in Group 2 and 27.9% in Group 3. A Chi-Square test of independence illustrated a significant difference between these proportions, $\chi^2(2) = 9, p = 0.01$. Bonferroni corrected ($\alpha = 0.017$) post-hoc pairwise comparisons showed a significant difference between Group 1 and Group 3, $p = 0.008$ and between Group 2 and Group 3, $p = 0.011$. These percentages are notably higher than those reported in Section 5.4.1, suggesting that participants who interpreted probabilities as propensities were mostly driving the insufficiency in qualitative explaining away.

5.5.2. Quantitative

Similarly, within the new subset of data, a repeated-measures ANOVA with a Greenhouse Geisser correction was carried out on the average

probability estimates on the relational explaining away questions, within each of the groups. Results illustrated a significant difference between these estimates within Group 1; $F(1.58, 82.6) = 70.43, p < 0.0001$, within Group 2; $F(1.8, 116.4) = 141.5, p < 0.0001$ and within Group 3; $F(1.6, 119.4) = 63, p < 0.0001$. Post-hoc pairwise t-tests allowed us to obtain 95% confidence intervals (CI) of the difference in the average empirical probability estimates between pairs of inferences of interest (see Table 3 below). This analysis illustrated that there was no sufficient explaining away in any group, since the normative difference was not included in any of the 95% CI of the empirical differences.

Notably, however, compared to findings reported in Section 5.4.2, the insufficiency was less pronounced as the 95% CI of the empirical differences were now closer to the normative differences. This suggests that participants interpreting probabilities as propensities were significantly contributing to the observed insufficiency of quantitative relational explaining away.

Table 3: Within group explaining away excluding participants who did not change their quantitative estimates from their stated priors.

Inferences	Normative difference	Empirical difference	95% CI of empirical difference
<i>Group 1</i>			
A-B	36	22.7	[14.9, 30.4]
C-B	80	53.9	[42.7, 65.2]
<i>Group 2</i>			
A-B	36	23.2	[16.9, 29.6]
C-B	80	58.7	[50.47, 66.9]
<i>Group 3</i>			
A-B	36	8.6	[1.8, 15.4]
C-B	80	41.5	[31.9, 50.9]

Note: $A := P(C_1|E), B := P(C_1|E, C_2), C := P(C_1|E, \neg C_2)$.

6. DISCUSSION

Over the past few decades, causal Bayesian networks have been successfully employed in many domains of human reasoning. Despite this, empirical work in the psychological literature has repeatedly illustrated that when engaging in explaining away, people violate the normative CBN model in numerous ways. One recurrently reported

violation in empirical studies of explaining away pertains to the violation of the Markov assumption of independence (Mayrhofer & Waldmann, 2015; Rehder & Waldmann, 2017; Rottman & Hastie, 2016). Another pertains to people's under-adjustment of probabilities and insufficient explaining away (Davis & Rehder, 2017; Fernbach & Rehder, 2013; Liefgreen et al., 2018; Morris & Larrick, 1995; Rehder & Waldmann, 2017; Rottman & Hastie, 2016; Sussman & Oppenheimer, 2011). Although this insufficiency has partly been attributed to structural violations of the normative model such as the assumption of independence, we argue it instead to be the product of methodological confounds of previous studies and participants interpreting probabilities as propensities.

In the present study, we utilised a novel methodology to address the issues often found in empirical studies of explaining away. For example, we explicitly stated priors and re-elicited these from participants and we utilised relational qualitative and quantitative questioning. This approach was seemingly successful in making participants understand the parameters and relational properties found within the common-effect structure they were required to reason with. As such, in all three conditions a high proportion of participants correctly answered questions regarding prior probabilities of causes, independence of causes, and the final logic question. This allowed us to conclude that the assumption of independence remained intact in all conditions, participants had accepted the priors given to them, and they understood what circumstances were necessary to bring about the common effect. This is in contrast to the vast majority of findings reported by studies in the extant literature (e.g. Rottman & Hastie, 2016) and allowed us to make meaningful comparisons between people's inferences and those dictated by the normative model.

Despite these encouraging improvements, our main findings echo those of the extant literature, including our previous study (Liefgreen et al., 2018), as participants systematically violated the normative account of explaining away by under-adjusting their belief estimates in all three conditions. Pitfalls in relational explaining away were visible at the level of both diagnostic reasoning and explaining away. As such, participants in all groups performed extremely poorly in qualitative and quantitative diagnostic reasoning questions and slightly better, but still sub-optimally, in questions relating to explaining away. In our study, deviations from the normative model could not be attributed to violations of the independence assumption, but instead seem to arise, at least in part, from some participants interpreting probabilities as propensities. As predicted by our propensity interpretation hypothesis, a larger number of participants engaged in this 'stay the same' behaviour on both qualitative and quantitative relational explaining away inferences in Group 1 reasoning with the coin-tossing cover story than in Group 3,

reasoning with the social dinner party cover story, with Group 2 reasoning with balls and containers cover story falling in between. Further analyses on people's qualitative and quantitative relational explaining away without the cohort of participants who seemed to have interpreted probabilities as propensities showed that participants' insufficiency was now noticeably less pronounced, suggesting that the cluster of participants who interpreted probabilities as propensities was, at least partly, driving the extreme insufficiency observed in the overall sample.

Another large cluster of participants' responses in our data that seemed to have been responsible for the overall insufficient explaining were those that updated their probabilities in diagnostic reasoning using an erroneous strategy whereby $P(C_1|E) + P(C_2|E) = 1$, with $P(C_1|E) = P(C_2|E) = 0.5$ (assigning equal probability to each cause), or $P(C_1|E) = 0.67$ and $P(C_2|E) = 0.33$ (assigning a probability to each cause that reflects the 2:1 priors ratio) (see Figure 2). Further experimental work is needed to test for the pervasiveness of this erroneous strategy in diagnostic reasoning within explaining away.

Overall, our findings advocate for future work that not only investigates whether people's explaining away reasoning differs from normative predictions, but, given the proven robust nature of the insufficiency of explaining away, also explores when and why these deviations occur. Moreover, in order to investigate people's differential interpretations of probability more directly, we suggest future research could manipulate the phrasing of questions (see Ülkümen et al., 2016) so that participants are being asked either about propensities or subjective probabilities when making inferences on independence, diagnostic reasoning and explaining away.

REFERENCES

- Davis, Z., & Rehder, B. (2017). The causal sampler: A sampling approach to causal representation, reasoning, and learning. In *Proceedings of the cognitive science society*.
- Fernbach, P. M., & Rehder, B. (2013). Cognitive shortcuts in causal inference. *Argument & Computation*, 4(1), 64–88.
- Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. In W. Brun, G. Kirkebøen, & H. Montgomery (Eds.), *Essays in judgment and decision making*, 21–35. Oslo, Norway: Universitetsforlaget.
- Giere, R. N. (1973). Objective single-case probabilities and the foundations of statistics. In *Studies in logic and the foundations of mathematics*, 74, 467–483.
- Griffiths, T. (2001). Explaining away and the discounting principle: Generalising a normative theory of attribution. *Unpublished manuscript*.

- Hájek, A. (2012). Interpretations of probability. In *The stanford encyclopedia of philosophy*.
- Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, 11(2), 143–157.
- Kelley, H. H. (1973). The processes of causal attribution. *American psychologist*, 28(2), 107–128.
- Keren, G., & Teigen, K. H. (2001). The probability-outcome correspondence principle: A dispositional view of the interpretation of probability statements. *Memory & cognition*, 29(7), 1010–1021.
- Liefgreen, A., Tešić, M., & Lagnado, D. (2018). Explaining away: significance of priors, diagnostic reasoning, and structural complexity. In *Proceedings of the 40th annual conference of the cognitive science society*. Austin: Texas.
- Mayrhofer, R., & Waldmann, M. R. (2015). Agents and causes: Dispositional intuitions as a guide to causal structure. *Cognitive Science*, 39, 65–95.
- Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review*, 102(2), 331–355.
- Neapolitan, R. E. (2003). *Learning Bayesian networks*. Pearson Prentice Hall UpperSaddle River, NJ.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, CA: Morgan Kaufman.
- Pearl, J. (2009). *Causality*. Cambridge university press.
- Popper, K. R. (1959). The propensity interpretation of probability. *The British journal for the philosophy of science*, 10(37), 25–42.
- Rehder, B. (2014a). Independence and dependence in human causal reasoning. *Cognitive psychology*, 72, 54–107.
- Rehder, B. (2014b). The role of functional form in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3), 670–692.
- Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive psychology*, 50(3), 264–314.
- Rehder, B., & Waldmann, M. R. (2017). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory & cognition*, 45(2), 245–260.
- Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological bulletin*, 140(1), 109–139.
- Rottman, B. M., & Hastie, R. (2016). Do people reason rationally about causally related events? Markov violations, weak inferences, and failures of explaining away. *Cognitive psychology*, 87, 88–134.
- Sussman, A. B., & Oppenheimer, D. M. (2011). A causal model theory of judgment. In *Proceedings of the 33rd annual conference of the cognitive science society*. Austin: Texas.
- Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of experimental psychology: General*, 145(10), 1280–1297.
- Wellman, M. P., & Henrion, M. (1993). Explaining “explaining away”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3), 287–292.

Responsible Agent Deliberation

DOUGLAS WALTON

*Centre for Research in Argumentation, Reasoning and Rhetoric,
University of Windsor
dwalton@uwindsor.ca*

ALICE TONIOLO

*School of Computer Science, University of St Andrews
a.toniolo@st-andrews.ac.uk*

Our question is whether by adhering to the designed rules of today's deliberation dialogue protocols, intelligent rational agents can act fairly, transparently and responsibly. In this paper, we propose some reflections and guidelines on how deliberation dialogue should be held to these principles using norms to define protocols.

KEYWORDS: Deliberation dialogue, multiagent systems, responsibility

1. INTRODUCTION

In recent years, advances in the development of Artificial Intelligence systems have called for a reflection on principles that such systems should adhere to. Fairness, responsibility and transparency in decision making among others are essential in today's design of agent systems. Agents engaging in deliberation to make a decision or to solve a problem act on behalf of their users and the development of such systems should be held to these principles. In this paper, we reflect on the effect that current design choices in deliberation dialogue frameworks have towards fairness, responsibility and transparency of protocols for such a dialogue.

Formal dialogue protocols define how to move forward in a dialogue prescribing how an agent might respond to a particular statement, and when they are allowed to speak. This approach is commonly used in current deliberation systems, such as the McBurney, Hitchcock, and Parsons (MHP) model (2007). Commenting on these conversational policies among agents, Maudet et al (2002) suggest that they may or may not represent deliberations of the kind found in natural conversation. In a more general context, Shi et al (2010) show

that even in a flexible protocol, undefined sequences or unexpected additional sub-sequences of speech acts occur in natural dialogue. Later protocols have suggested that additional elements are required to represent more natural deliberation (Walton et al, 2016). However, by observing instances of human dialogue, we note that the design of dialogue protocols rules on how the agents are required to act might have consequences on the information shared, on the decisions taken, and beyond, to affect the resulting actions.

Hence our question is whether by adhering to the designed rules of today's deliberation dialogue protocols, agents will behave fairly, transparently and responsibly. In this paper, we propose some reflections and guidelines on how deliberation dialogue should be held to these principles using norms to define protocols. Fairness requires that protocols are designed to not discriminate against agents. Our initial observations show that the turn-taking function may cause unfair behavior, and that this function is not typical of deliberation, even though it is typical of persuasion dialogue. In deliberation dialogue, dialogues need to be more transparent so all can see the reasons given supporting or attacking the various alternatives (Yu et al, 2018).

Responsibility of agents has been studied in many contexts, predominantly in social and ethical behavior. Here we reflect on the problem of omission of information and attribution of responsibility due to protocol prescriptions. Castelfranchi (2000) holds that agents will inevitably deceive each other, and one way is by making an agent ignore something crucial for them. We show that the dialogue protocol rules may cause agents to be unable to state crucial information about an action. An agent may then be held responsible later if that action causes serious negative consequences. Responsibility should be considered as one of the principles for protocol design. We conclude our paper with some desirable properties that deliberation dialogue protocols should adhere to, to achieve better fairness, responsibility and transparency in decision making.

For the purposes of this paper, we define an *intelligent autonomous agent* (IAA) as an entity minimally having the following five capabilities, following in broad outline the approach of (Wooldridge, 2009). First, an IAA has the capability for perception and for collecting information. Second, an IAA can foresee some (but not all) of the future consequences of its actions and can change its planned actions accordingly. Third, an IAA can communicate with other agents so they can act together. Fourth, it can be inferred from the speech acts of an IAA that it is committed to a proposition, an action or a goal. Fifth, an IAA has the capability to add or retract commitments from its commitment store.

2. AGENT DELIBERATION

Deliberation can mean a wide variety of things in natural language (pretty much any activity involving some kind of thought can be called deliberation), but in recent computer science it has been given a more precise meaning. McBurney et al. (2007) cite three characteristics that have been widely adopted. First, deliberation is concerned with actions rather than propositions (and so is different from inquiry). Second, there are no initial commitments on either side (and so it is different from persuasion). Third, deliberation is cooperative rather than adversarial. The object is to achieve consensus, rather than conversion (persuasion) or compromise (negotiation).

Below is a simple example of a deliberation dialogue adapted from Kok et al (2011).

Ann: Where should we go for dinner?
Bob: We should go to the Italian restaurant.
Ann: Why?
Bob: It serves very tasty pizza.
Ann: But it is too expensive. We should go to the
Japanese restaurant.
Bob: Why?
Ann: It's close to my place.
Bob: But I have to go home early and the Japanese
restaurant is too far.

So far the deliberation dialogue has reached an impasse. But suppose the dialogue continues when Ann offers some new information which gives rise to a new option.

Ann: I noticed this new Greek restaurant on my way to
work today, it is close to your place, and much cheaper.
Bob: OK.

Ann has offered two arguments supporting this new option. Bob indicates that he is OK with the proposal she has made, and so the dialogue has reached a successful resolution. They can go ahead with this proposal.

What made the dialogue terminate with a good outcome based on the arguments and proposals put forward by both sides? The introduction of the new information that Bob intended to go home early guided the subsequent identification of a new option revising the initial issue, helping the parties to find a suitable agreement.

In order for agents to engage in this dialogue, an agent model requires a representation of plans, actions, commitments and goals. A model of arguments is then required for agents to construct instantiated arguments about plans and actions to put forward in the dialogue. Finally, a dialogue protocol must be defined to identify when one agent is allowed to speak and what arguments can be stated. At each turn, the agent will identify from the protocol the possible speech acts that can be used to respond to a previous speech act. These will include arguments that could be exchanged, identified according to plans, actions, and goals. A selection of the next move is then to be made among the set of potential answers available. Figure 1 shows the layered representation of the agent knowledge as adapted from Prakken and Sartor (2002).

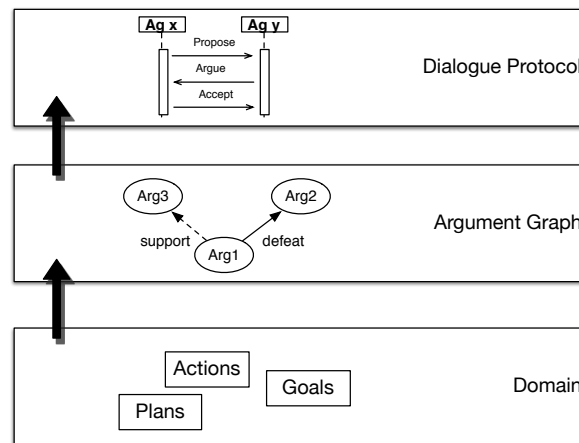


Figure 1. An agent layered dialogue model.

Arguments exchanged during the dialogue can be woven into an argumentation structure represented by an argument map (argument diagram). So we can use the familiar argumentation tools to evaluate the whole sequence of connected argumentation to get a big map showing all the supporting and attacking arguments for each of the proposals.

3. FORMAL DIALOGUE MODELS OF DELIBERATION

The seminal MHP model (McBurney, Hitchcock and Parsons, 2007) has three stages: an opening stage, an argumentation stage and a closing

stage.¹ During the opening stage the issue is settled concerning the choice to be made. During the argumentation stage, there are four kinds of distinct intervals. During the first interval the agents seek for information concerning the circumstances of the case where the decision is to be made. During the second interval, the agents put forward proposals offering potential solutions to the problem that is to be solved in the deliberation dialogue. During the third interval, the agents consider and revise the proposal that has been put forward. During the fourth interval the agents recommend a particular proposal as the one best suited to solve the problem or to make the best decision based on the information that has been collected and assessed. The third stage of the dialogue is the closing stage where the agents reach agreement on what action to take, based on the evidence collected and the recommendations made during the argumentation stage.

An interesting problem with applying the MHP model to realistic cases of deliberation is that the knowledge bases that the agents have tend to be incomplete, and may need to be updated once new information comes in. For this reason Walton, Toniolo, and Norman (2016) proposed a model in which an open knowledge base enables information about changed circumstances to come in. During the argumentation stage there is a cyclical flow of argumentation as new knowledge comes that requires re-evaluation of proposals.

According to this revised model of deliberation dialogue, an additional feature is a knowledge base that is continually collecting new information about the circumstances as the agents are deliberating. In the Walton, et al. model, this information is used to continually update the knowledge base as new circumstances are retrieved. Naturally, as new knowledge comes in, this will affect the framing of the choice to be made, which may have to be updated as some options turn out to be unrealistic while others are supported by the new evidence. In the example dialogue in Section 2, the Ann's solution to the problem was based on new information that came in.

The argumentation stage of the revised model is comparable to that of the MHP model. In the first interval, where the agents find the circumstances of the decision to be made, new information continually streaming in from the updated knowledge base affects the other three intervals during the argumentation stage where proposals are put forward, revised and evaluated.

Based on this reconstruction of the argumentation stage, the revised model moves to a closing stage in which the best proposal is

¹ Subsequent models of deliberation dialogue include (Kok, Meyer, Prakken, & Vreeswijk, 2011), (Medellin-Gasque, Atkinson, McBurney, & Bench-Capon 2011) and (Walton, Toniolo, & Norman, 2016)

accepted as the course of action best suited to the findings carried out in the argumentation stage.

4. CONTROL OF AN INTELLIGENT AGENT

To move toward providing a framework defining moral responsibility in section 5, we introduce the technical term 'control' to stand for the capacity of an agent to act, as represented by the set of capabilities of an IAA defined in Section 1.

Control, in this sense of the term is *"the capacity to intervene in the course of events so as to be able both to make something happen and to preclude it from happening, this result being produced in a way that can be characterized as in some sense intended or planned or foreseen"* (Rescher, 1969, 329). On this view, a rational agent has control over its actions (or refraining from actions) of a sort that can change its circumstances. It can set goals for itself, direct its actions based on these goals, and can retract or modify its goals, for example if it sees that its goals conflict.

To extend the notion of control beyond the account of the capacities of an agent in Section 1, we can add seven further capabilities relating to goals and actions by expressing them in the language of control.

- (1) An agent has control over carrying out actions (or refraining from actions) of a kind that can change its circumstances.
- (2) An agent has goals, can set goals for itself, and direct its actions based on these goals.
- (3) An agent can retract or modify its goals, as it might do if it sees that its goals conflict.
- (4) An agent can grasp how actions to achieve a goal fall into an ordered sequence where some actions are required to carry out others.
- (5) An agent can organize goals and actions into a hierarchy of levels of abstraction.
- (6) An agent will generally keep trying to achieve a goal even if it has previously failed (plasticity), unless it has reasons to stop trying.
- (7) An agent will not continue trying to carry out an action that it knows is impossible.

These capabilities can be formulated even more precisely by classifying the different types of control that can be distinguished. A classification system has been drawn up (Walton, 1974, 163), distinguishing six types of control: (1) complete positive control, (2) complete negative control, (3) positive partial control, (4) partial negative control, (5) full partial control, and (6) complete full control. The only kind of control not

defined yet is that of full partial control. Full partial control can be explained by looking at Figure 2.

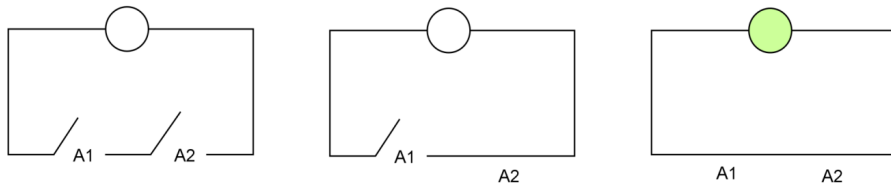


Figure 2: Series Circuit

Figure 2 represents the kind of case where agents A1 and A2 jointly, but not individually, have both positive and negative control over the outcome. An example of full partial control would be one where A1 and A2 are separately at the mercy of A3 with respect to the outcome, but where A1 and A2 can team up and get control of the outcome from A3. Individually, A1 and A2 are powerless to produce or prevent the outcome which is fully controlled by A3, yet jointly they can either produce or prevent the outcome. The existence of this type of control suggests the usefulness of modeling control as a teleological notion that needs to be defined within a framework of multiagent deliberation.

Next let us look at Figure 3. As shown by the two right circuits, A1 can keep the light on whether A2 turns her switch off or on. The only way for the light to be off is if both agents keep their switch in the open position, as shown in the left circuit.

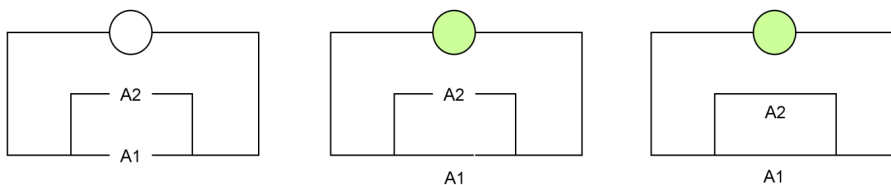


Figure 3: Parallel Circuit

Each agent, for example A1, can illuminate the light by closing his/her switch, thus exercising positive control over the outcome that the light is on, but A1 cannot make it so that the light is off unless A2 also decides to keep her switch open. Thus although A1 has positive control, he lacks full control because he lacks negative control. Neither agent individually has negative control. But each has positive control. Each has partial control. If they act together, they can exert positive control on whether the light is on or off (Walton, 1974, 164).

However, the requisite notion of control has not yet been defined completely enough to handle all problem cases. To conclude this section we set out two problems for defining the notion of control in this technical sense more fully.

The first example is posed by asking the question: does the moon control the tides (Rescher, 1969, 332)? Rescher answers that, in the proper meaning of the term ‘control’, it does not. Although he concedes that the movements of the moon determine the ebb and flow of the tides, he asserts that it is not proper to say that in the sense of the word ‘control’ he has in mind, the moon controls the tides. We agree with Rescher that in order to retain the intuitive idea of control, there has to be some aspect of deliberative agency or goal-directed action on the part of a controller present in the background.

The second example is the case of the berserk traffic light. Rescher argues this is not really an instance of control because *“it is not possible to retain the intuitive idea of control without retaining some aspect of deliberative agency or purpose contrivance on the part of the controller”* (Rescher, 1969, 332). He argues that the traffic light may still determine the flow of traffic, but the flow is no longer a controlled one, so in his (and our) sense of the word, the terminology of control has become inappropriate.

5. MORAL RESPONSIBILITY OF INTELLIGENT AGENTS

An IAA is only morally responsible for actions that actions it controls (carries out voluntarily, could have done otherwise). Such an agent must be autonomous (have self-control). Traditionally in philosophy this factor is called “free will” (a contested term. One way to reframe this notion so it can be made more precise for application to multiagent systems is to say that such an agent can control its actions). It is generally assumed in moral and legal philosophy that moral responsibility is *“the status of morally deserving praise, blame, reward, or punishment for an act or omission, in accordance with one’s moral obligations”* (Eshleman, 2016, 1).

An intelligent rational ethical agent is an IAA that is committed to social (ethical) norms specifying that certain actions, or kinds of actions, are obligatory, permitted or forbidden in a group it is part of, in addition to the defining features of an IAA given above. This means that to have a formal deliberation system in which to frame ethical judgments about responsibility, deontic logic has to be brought in.

An open question, however, remains on how a deliberation model for an ethical IAA should be designed to yield a fair, responsible and transparent deliberation, particularly if that agent’s role is to

deliberate on behalf of a user in a team of agents or a mixed team of agents and users.

Desirable properties in our dialogue model include the ability to explain why a decision was taken, walking back through the dialogue exchange. Key information needs to be exchanged to identify a new option, as well as the selection of a new option. Reasons why a particular dialogue step was taken need to be allowed, according to the dialogue protocol formalization. We note that with the use of argumentation-based deliberation dialogue, dialogues are more transparent through reasons supporting various alternatives (Yu et al, 2018). Argumentation-based explainable AI (e.g. Fan and Toni, 2015) can be used to compute a set of arguments that form relevant explanations to the acceptability of an argument. Tintarev and Kutlak (2014) propose a system of dialogue to better understand the steps of a plan for example “Why does the system NOT say that I should do Y?” The user can ask why an option is rejected.

Consider a follow up to our example in Section 2. Ann and Bob agree that they will go to the Greek restaurant, but next they have to decide how to get there. Ann suggests that the fastest way to get there is to take the tube to the place, but in the end they decide to walk to the main square and then take the bus from there because the tube is too busy. Assume that Charlie joins the discussion later. If so, Charlie should be provided with an explanation on why they are not taking the fastest route.

The second desirable property is that of fairness, which requires that protocols are designed not to discriminate against agents. From one side, agents should be allowed to exchange actions and plans that better represent their interests and that of the group. Our focus however is to understand whether this is always possible given a specific protocol. By using a group turn-taking algorithm, agents can eventually voice all the proposals that they have available, provided that they can continue to discuss previously moved proposals, skip a turn, or advance new proposals.

However, we noticed that this function together with other constraints might prevent agents to exchange proposals or information leading to an unfair situation. For example, in Toniolo, Norman, and Sycara (2012), adopting components from Kok et al’s (2011) dialogue framework, an agent can only make a relevant move in a dialogue. A relevant move is one that changes the acceptability status of a proposal, but this may prevent an agent from stating other proposals or further information. The dialogue protocol rules may then prevent agents from being able to state crucial information about an action because the statement no longer contributes to changing the acceptability of the proposal. However, this information may be essential to identify a

different proposal. An agent may then be held responsible later if the action or plan chosen causes negative consequences. In our example, assume that Bob has also a different reason for not wanting to go to the Japanese restaurant: not only is it far, but it can only be reached on foot, and his knee is painful. Assume that Bob shares this second reason, instead of stating that he wants to go home early. Note that in this example, only one of these reasons would be considered relevant. Ann would not receive the critical information that the place needs to be close by. Hence the dialogue could take longer to explore the space of proposals with closer locations, and might end with a less favorable option or end with no acceptable option.

6. ELEMENTS OF MORAL RESPONSIBILITY

Aristotle (1969) suggested that knowledge is an element of moral responsibility (*Nichomachean Ethics*, 1110 b1 17) when he wrote that everything that is done by reason of ignorance is nonvoluntary. Aristotle argued that for a man to have acted voluntarily in the ethical sense of the term, he must know what he was doing when he acted (1110 b1 18). He also holds that an action can be a candidate for praise or blame only if it was voluntarily undertaken by its agent.

This approach suggests a way of modeling the inferential structure of the sequence of evidence-based argumentation used in legal and ethical cases to arrive at a conclusion about how to assign praise or blame. We argue that an agent acts voluntarily only if the action he carried out was under his control at the time.

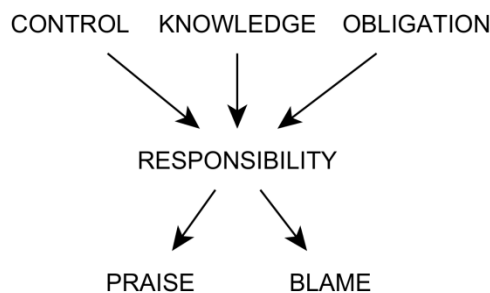


Figure 4: Factors for Arguing from Responsibility to Praise and Blame

On this approach, the three elements required to draw a conclusion about the responsibility of an IAA are control, knowledge and obligation, as shown in Figure 4. However, after some discussion at the

ECA conference in Groningen, we became convinced that in many cases, the notion of causation has to be factored in as well. Our initial reaction was that the concept of causation is too complex and not always needed for judging ethical responsibility. However, we accepted that a partial definition of causation could be used based on the INUS conditions.² On this approach, one event or action *A* *causes* another event or action *B* if and only if *A* is a member of a set of necessary conditions that, taken together, provide a sufficient condition for the occurrence of *B*. However, following the theory of causation in law of Hart and Honoré, such a selected event is generally a voluntary (human) action or an event or action that is “abnormal”.

The next question is how to define the concept of a voluntary action. It is a contested concept and there is much literature on it in law, philosophy, and other fields. But H. L. A. Hart had a way around this. He saw remarkably (in 1949) that voluntariness is best defined in law as a defeasible concept (Hart, 1949, 180). That is, instead of seeing voluntariness as some elusive internal event or state in the human mind, he wrote that it should be defined by excluding a number of other concepts. This means that it serves to *exclude* a heterogeneous range of cases such as physical compulsion, coercion by threats, accidents mistakes, etc. In today’s terms, he saw it as a defeasible concept. This insight anticipates the later AI view that case-based reasoning of the kind used in ethical and legal reasoning is inherently defeasible.

² INUS conditions are insufficient but non-redundant parts of a condition which is itself unnecessary but sufficient for the occurrence of the effect according to the account of (Mackie, 1974).

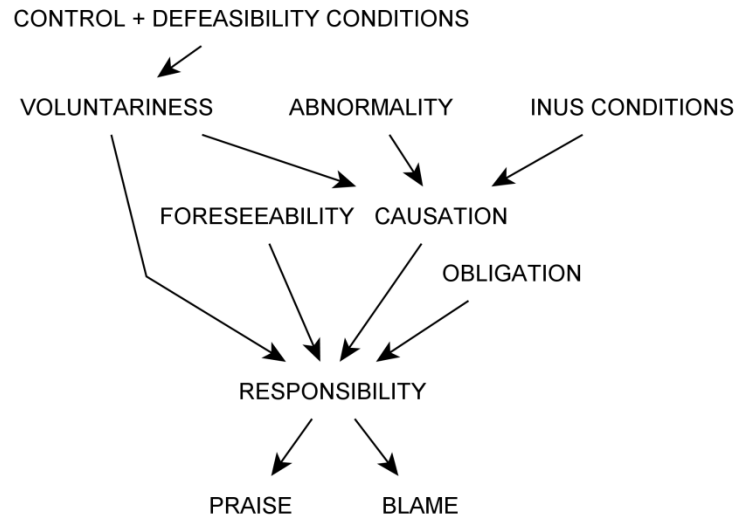


Figure 5: Factors for Arguing from Responsibility to Praise and Blame

What advantage can we derive from the insights of these early British analytical philosophers? To accommodate them we could modify the attempted defining conditions on how responsibility should be arrived at shown in Figure 5 as follows.

As indicated in Figure 5, we distinguish four basic components of responsibility. These are voluntariness, foreseeability, causation and obligation. We define causation in an admittedly simplistic way by using the INUS conditions, leaving the concept open to further refinement. Although causation is not always required to be considered, it is important in some cases, and therefore useful to include.

Foreseeability fits in well with the account of capabilities of an IAA listed in Section 1. In such systems, a rational agent has only a bounded rationality: it is aware of some (but not all) of the consequences of the actions it carries out or is contemplating carrying out. How foreseeable such a rational agent is expected to be in legal and ethical settings is variable. It is circumscribed by the granularity of the common sense reasoning that needs to be applied to the given situation in which the agent is situated. Our proposal then is that when the notion of ethical responsibility is redefined in this way, it would integrate formal models of deliberation with requirements for defining the notion of responsibility in a manner suitable for use in artificial intelligence.

Responsibility for omissions should be considered as one of the principles for protocol design based on the scheme for blame for omissions shown below.

Argumentation Scheme for Blame for Omissions

Major Premise: Agent A1 failed to carry out action S1.

Minor Premise 1: A had control over carrying out action S1.

Minor Premise 2: A1's failure to carry out S1 had negative consequences (NC).

Conclusion: A is to blame for NC.

Critical Questions

CQ1: What kind of control was involved, such as full control or partial control?

CQ2: Did A1 have knowledge about the reasonable likelihood of NC?

CQ3: Did A1 have an obligation to carry out S1 or otherwise to prevent NC?

CQ4: Did A1's carrying out S1 run into conflict with some other obligation of A1?

CQ5: Could NC have been prevented by other agents who were involved?

If we consider this scheme in the context of an agent deciding which argument or proposal to move forward in the deliberation, an agent may fail to inform another agent about an action T1 that could have been carried out at a particular time. If we assume that this lack of information leads to carry out T2 instead, which is revealed to yield negative consequences, the scheme above can be used to reason about whether A1 is to be blamed for these consequences.

7. CONCLUSIONS

We have presented a typology of deliberation dialogue that can be used to model ethical and legal responsibility in agent deliberation. Fairness, responsibility and transparency in AI decision making, among other properties, are essential in today's design of agent systems. We have shown that argument-based models of dialogue are useful for achieving more transparent decisions and that responsibility has many components that should be considered, including control and obligations. Protocols (and related constraints) should be further studied to consider how to block unfair moves in deliberation dialogues.

REFERENCES

Aristotle (1984). *Nicomachean Ethics* (trans. W. D. Ross, revised by J. O. Urmson). In Jonathan Barnes (ed.), *The Complete Works of Aristotle*. Princeton: Princeton University Press.

- Castelfranchi, C. (2000). Artificial liars: Why computers will (necessarily) deceive us and each other. *Ethics and Information Technology*, 2(2), 113-119.
- Eshleman, A. (2016). Moral Responsibility, *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.). <https://plato.stanford.edu/archives/win2016/entries/moral-responsibility>.
- Fan, X., and Toni, F. (2015). On computing explanations in argumentation. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- Hart, H. L. A. (1949). The Ascription of Responsibility and Rights, *Proceedings of the Aristotelian Society, New Series*, 49(1948-1949), 171-194. Published by: Wiley on behalf of The Aristotelian Society. <http://www.jstor.org/stable/4544455>.
- Hart, H. L. A. and Honoré, A. M. (1959). *Causation in the Law*. Oxford: Clarendon Press.
- Kok, E. M., Meyer, J.-J., Prakken, H. and Vreeswijk, G. (2011). A Formal Argumentation Framework for Deliberation Dialogues. In *Argumentation in Multi-Agent Systems*, ed. McBurney, P., Rahwan, I. and Parsons, 31-48. Springer Berlin Heidelberg.
- Maudet, M. and Chaib-draa, B. (2002). Commitment-based and Dialogue-based Protocols: New Trends in Agent Communication Languages: <http://www2.ift.ulaval.ca/~chaib/publications/maudet-chaib-ker-revised2.pdf>
- Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation*. Oxford: Clarendon Press.
- McBurney, P., Hitchcock, D., and Parsons, S. (2007). The Eightfold Way of Deliberation Dialogue. *International Journal of Intelligent Systems*, 22(1), 95-132.
- Medellin-Gasque, R., Atkinson, K., McBurney, P., and Bench-Capon, T. J. M. (2011). Arguments over co-operative plans. In *Theory and Applications of Formal Argumentation, Lecture Notes in Computer Science*, 7132, 50-66. Springer Berlin Heidelberg.
- Prakken, H. and Sartor, G. (2002). The role of logic in computational models of legal argument: a critical survey. In *Computational Logic: Logic Programming and Beyond, Lecture Notes in Computer Science*, 2408, 342-381. Springer Berlin Heidelberg.
- Rescher, N. (1969). *The Concept of Control. Essays in Philosophical Analysis*. Pittsburgh. Pittsburgh University Press.
- Shi, H., Ross, R., Tenbrink, T. and Bateman, j. (2010). Modelling illocutionary structure: Combining empirical studies with formal model analysis. In *Proceeding of the 11th International Conference in Computational Linguistics and Intelligent Text Processing*, 340-353.
- Tintarev, N., and Kutlak, R. (2014). SASy: Making Decisions Transparent with Argumentation and Natural Language Generation. In *Proceedings of the International Conference on Intelligent User Interfaces*, 29-32.
- Toniolo, A., Norman, T. J., and Sycara, K. (2012). An empirical study of argumentation schemes for deliberative dialogue. In *Proceedings of the Twentieth European Conference on Artificial Intelligence*, 242, 756-761.

- Walton, D. (1974). *Control, Behaviorism*, 2(2), 162-171.
- Walton, D., Toniolo, A. and Norman, T. J. (2016). Towards a Richer Model of Deliberation Dialogue: Closure Problem and Change of Circumstances, *Argument and Computation*, 7(2-3), 155-173.
- Wooldridge, M. (2009). *An Introduction to MultiAgent Systems*, 2nd ed., John Wiley & Sons.
- Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V. R., and Yang, Q. (2018). Building ethics into artificial intelligence. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 5527-5533.

Disagreement in Aristotle's *Topics* and its Latin Medieval reception

GUSTAVO FERNÁNDEZ WALKER

FLoV, University of Gothenburg
gustavo.fernandez.walker@gu.se

ANA MARÍA MORA-MÁRQUEZ

FLoV, University of Gothenburg
ana.maria.mora.marquez@gu.se

This paper aims to shed light on the Aristotelian notion of dissent and its understanding in the medieval reception of Aristotle's dialectics. Dissent is fundamental in Aristotle's *Topics*, although its exact nature is not clear. Specifically, it is unclear whether dissent is essentially related to an epistemic quality of the premises involved, or whether it simply amounts to social disagreement about them, regardless of their epistemic quality. Medieval scholars clearly opt for the former reading. We aim to show that this owes to their understanding of logic as a tool for science.

KEYWORDS: Aristotelian Dialectics, Medieval Dialectics, Aristotelian Dissent, Medieval Argumentation, Ancient Argumentation

1. ARISTOTLE'S TOPICS

In his *Topics*, Aristotle sets out to provide a method for the dialectical discussion. Although the dialectical discussion has several uses (cf. *Top.* I.2.101a25–28), it is clear¹ that the immediate use the method developed in the *Topics* intends to regiment is the gymnastic discussion between a questioner and an answerer, possibly in the presence of a third party.

¹ Mainly from the advice to questioner and answerer given by Aristotle in *Topics* VIII. This is Primavesi's position in Primavesi (1996), with which we side. See also Allen (2007). Owen (1961), Barnes (1980), and Bolton (1990) hold that the *Topics* provide a method for philosophical discussion.

The gymnastic dialectical discussion is structured as follows:

- (a) A questioner, *Qu*, presents a problem (*problema*) to the answerer, *An*, of the form 'Is *p* the case or not? Where '*p*' is a problematic proposition. Let's call *t* the thesis (i.e., either '*p*' or ' $\sim p$ '), the statement to which *An* commits himself. *Qu*'s objective is to force *An* to concede $\sim t$, the contradictory of his thesis.
- (b) *Qu* must do that by introducing questions (*protaseis*) of the form 'Is *q_n* the case?', which *An* can concede or reject. If *An* concedes, '*q_n*' becomes a premise to which *An* is henceforth also committed.
- (c) *Qu* wins whenever she has led *An* to accept a set of *q_n*s that syllogistically imply $\sim t$, so that *An* will be forced to also concede $\sim t$ and be in contradiction with his initial position. She loses, and hence *An* wins, if she fails to do so over a given amount of time.

Let's call (a) the opening stage of the dialectical discussion, (b) its interrogative stage, and (c) its concluding stage.

Aristotle's method in the *Topics* aims to improve the quality of such a dialectical discussion through a systematic understanding of its opening and interrogatives stages in a way that enhances the training aspect of the discussion. So, a fundamental part of the method includes a determination of suitable dialectical problems and premises given a gymnastic aim.²

In *Topics* I.11 Aristotle describes the dialectical problem as:

[(i)] a point of speculation, [(ii)] directed either to choice and avoidance or to truth and knowledge [...] [(iii)] about which people either have no opinion, or the public think the opposite of the wise, or the wise think the opposite of the public, or each of these groups have opposed opinions within itself.³ (Smith, 1997, p. 10)

This characterisation⁴ delimits the dialectical problem in terms of (i) the puzzling nature of its proposition, (ii) its subject matter, and (iii) its socio-epistemic character: A problem is the questioning of a puzzling proposition, most commonly practical or theoretical, with respect to

² Cf. Mora-Márquez (forthcoming).

³ *Top.* I.11.104b1–5: Πρόβλημα δ' ἐστὶ διαλεκτικὸν θεώρημα τὸ συντεῖνον ἢ πρὸς αἵρεσιν καὶ φυγὴν ἢ πρὸς ἀλήθειαν καὶ γνῶσιν, ἢ αὐτὸ ἢ ὡς συνεργὸν πρὸς τι ἕτερον τῶν τοιούτων, περὶ οὗ ἢ οὐδετέρως δοξάζουσιν ἢ ἐναντίως οἱ πολλοὶ τοῖς σοφοῖς ἢ οἱ σοφοὶ τοῖς πολλοῖς ἢ ἑκάτεροι αὐτοὶ ἑαυτοῖς.

⁴ For other analyses of Aristotle's *problemata*, see also Rubinelli (2009, pp. 4–5); Slomkowski (1997, pp. 15–18).

which there is either no general opinion or social disagreement.⁵ Let us focus on (i) and (iii).

The puzzling nature of the problematic proposition is a necessary condition for it to be worth an inquiry at all. Aristotle's characterisation of the puzzling proposition, the *theorema*, points to a proposition that warrants investigation insofar as (i) it does not come across as evident (*phaneron*) to everyone or the majority, and (ii) involves a difficulty (*aporia*):

No one in his right mind would hold out as a premise what nobody thinks or make a problem of what is evident to everyone or to most people, since the latter contains no puzzle while nobody would concede the former.⁶ (Smith, 1997, p. 9)

The aporetic aspect of the problematic statement (ii) suggests that something related to its subject matter makes it genuinely puzzling and hence worth an inquiry. For instance, 'the soul is unperishable' would be aporetic because of the difficult epistemic access we have to its subject, the soul. This is something that all problematic statements have in common. In the *Topics*, however, Aristotle focuses on the aspect that makes a problematic statement specifically dialectical *for the gymnastic purpose*: (i) that it does not come across as evident (*phaneron*) to everyone or the majority, i.e., its possibility of being the object of social disagreement. In other words, that a statement is puzzling for individual thinkers, and hence a matter of disagreement between them, is of no relevance for the gymnastic discussion.

This aspect draws attention to certain socio-epistemic attitudes towards the problematic proposition, so Aristotle introduces (iii): The problematic proposition suitable for the discussion must be either one about which there is no social stance at all or one about which there is disagreement, i.e., between the many and the wise, or between sub-groups within the many, or between sub-groups within the wise (cf. Brunschwig, 1967, p. 127). Moreover, such an aporetic proposition yields a good dialectical problem as long as the extent of its puzzlement is manageable given the length of the dialectical exercise, for:

Nor ought one to inquire into that the demonstration of which is near to hand, or those the demonstration of which is

⁵ Cf. Mora-Márquez (forthcoming).

⁶ *Top.* I.10.104a5-7: οὐδεὶς γὰρ ἂν προτείνειε νοῦν ἔχων τὸ μηδενὶ δοκοῦν οὐδὲ προβάλοι τὸ πᾶσι φανερόν ἢ τοῖς πλείστοις· τὰ μὲν γὰρ οὐκ ἔχει ἀπορίαν [...].

excessively remote. For the former present no difficulty, while the latter present too much for exercises.⁷ (Smith, 1997, p. 10)

But mostly, suitable dialectical problems involve problematic propositions about which there is already disagreement between social groups: (a) between the many and the wise; (b) between sub-groups of the many; or (c) between sub-groups of the wise.

How can we characterise the kind of disagreement Aristotle has in mind here? First of all, that it owes to the puzzling nature of a proposition indicates that it cannot include fabricated disagreement:⁸ Disagreement must be based on a plausible proposition whose truth-value is, however, not easily determinable. Second, that it obtains between the many and the wise, that is, well-demarcated social groups (in the Greek *polis*), also indicates that it must be social: Disagreement between individuals is of no relevance in this context. Finally, that the two well-demarcated social groups separate people with different epistemic backgrounds indicates that value is given to some degree of epistemic diversity: The opinion of the many has the same weight as that of the wise in determining whether a practical or theoretical proposition makes a good dialectical problem.

The two last entailments are better supported when we contrast the dialectical problem with the dialectical premise. Contrary to the dialectical problem, which is characterised in terms of social disagreement, the dialectical premise is characterised⁹ in terms of social acceptability – on its proposition being an *endoxon*:

A dialectical premiss is [(i)] the asking of [(ii)] something acceptable to everyone, most people, or the wise (that is, either all of them, most of them, or the most famous), provided it is not contrary to opinion (for everyone would concede what the wise think, so long as it is not contrary to the opinions of the many).¹⁰ (Smith, 1997, p. 9)

⁷ *Top.* I.11.7–9: οὐδὲ δὴ ὧν σύνεγγυς ἢ ἀπόδειξις, οὐδ' ὧν λίαν πόρρω· τὰ μὲν γὰρ οὐκ ἔχει ἀπορίαν, τὰ δὲ πλείω ἢ κατὰ γυμναστικήν.

⁸ An example of such “fabricated” disagreement can be found in those who doubt the validity of the principle of non-contradiction; see *infra*.

⁹ Other analyses of the premise are found in Brunschwig (1967, pp. xxxvi–xxxvii); Primavesi (1996, pp. 34–35); Slomkowski (1997, pp. 19–35); Smith, (1997, pp. 77–80).

¹⁰ *Top.* I.10.104a8–12: ἔστι δὲ πρότασις διαλεκτικὴ ἐρώτησις ἐνδοξος ἢ πᾶσιν ἢ τοῖς πλείστοις ἢ τοῖς σοφοῖς, καὶ τούτοις ἢ πᾶσιν ἢ τοῖς πλείστοις ἢ τοῖς μάλιστα γνωρίμοις, μὴ παράδοξος· θεὶ γὰρ ἂν τις τὸ δοκοῦν τοῖς σοφοῖς, εἰ μὴ ἐναντίον ταῖς τῶν πολλῶν δόξαις ᾗ.

The dialectical premise is an interrogative statement of the form ‘is *q* the case?’, where *q* must be an *endoxon* for the premise to have the highest odds to be conceded by *An*. There is disagreement between interpreters of Aristotle’s dialectics about the semantic and epistemic qualities a statement must have in order to be an *endoxon*.¹¹ We want to defend the view that the *endoxon* requires no epistemic quality other than to be plausible, and that it is characterised in terms of its social acceptability.

From the semantic perspective, the dialectical premise must meet no requirement, as its truth-value is irrelevant as long as it is *acceptable*. But how is this acceptability understood? As already said, a necessary condition for a proposition to be acceptable is that it be plausible, but, as we have seen, this is also a condition for it to yield a problem, so plausibility cannot be the defining feature of the *endoxon*. We submit that the defining feature of the *endoxon*, in opposition to that of the problem, is its being the object of wide social agreement. This means that an *endoxon* is: (a) in fact accepted by everyone or by most people, or by every wise or the majority thereof, as long as the many do not disagree with it, or (b) derived in some specific ways from propositions as in (a). A proposition that is accepted by the wise, but not by the many, does not make a good dialectical premise, not because the many are more likely than the wise to be right about it, but because *An* can justifiably reject it on the basis of such a disagreement.

To sum up, the dialectical discussion in the gymnastic setting (i.e., the one for which the *Topics* provides a method) trains the participants in the art of arguing about matters of societal relevance by putting under scrutiny a genuinely puzzling proposition which is the object of disagreement between well demarcated and diverse social groups. If *Qu* wins the discussion by skilfully introducing the right acceptable premises, she will have incidentally shown that *An*’s thesis is inconsistent with a set of opinions widely accepted by the relevant social groups. The purely gymnastic use of dialectics, however, is unrelated to truth-determination or epistemic-enhancement purposes with regard to individuals, even though other uses of dialectics can have those purposes in mind.

2. LATIN MEDIEVAL RECEPTION

Latin medieval readers of the *Topics* largely focused on the epistemic-enhancement potential of the dialectical discussion, mainly, we submit,

¹¹ For other interpretations of the *endoxon*, see Bolton (1990); Brunschwig (1967, pp. 113 – 114); Karbowski (2015); King (2013); Reinhardt (2015). Bolton and Reinhardt hold that the premise in the *Topics* is characterised by an epistemic quality; Karbowski rebuts such a reading.

because they took it to be one of the most efficient tools for the production and dissemination of scientific knowledge.¹²

Boethius of Dacia, a Danish master of Arts, active at the University of Paris during the 1270s, explicitly addresses this issue in the prologue of his commentary on the *Topics*, stating that "the dialectician does not consider truth itself (...) but rather shows the method that a given professional must apply in the pursuit of truth within a particular subject" (Green-Pedersen & Pinborg, 1976, pp. 8–9).¹³ Thus understood, then, the dialectical discussion does not seem to have a proper subject matter, but becomes rather a method used within the particular sciences or disciplines that provide the premises and problems of dialectical discussions. Consequently, their understanding of the disagreement and agreement involved in problems and premises relies on the epistemic qualities of scientific propositions and the epistemic attitudes of individual thinkers within the particular sciences.

This presented a problem for medieval commentators, since the dialectician and the scientist would deal with the same problems, thus rendering one of them superfluous (Green-Pedersen & Pinborg, 1976, pp. 46–49). On q. 36 of the first book of his commentary on the *Topics*, Radulphus Brito, a French master of Arts, active at the University of Paris from at least the 1290s, asks whether every premise and every problem are dialectical (*utrum omnis propositio et omne problema sit dialecticum*). He argues that they are not, but admits that if we consider only the way in which they are formulated, namely the *modus interrogandi* in the case of the premise and the *modus quaerendi* in the case of the problem, any proposition *p* could become a dialectical premise ("Is *p* is the case?"), or a dialectical problem ("Is *p* the case or not?").¹⁴ But this, according to

¹² The study of the medieval reception of Aristotle's *Topics* is still in many ways a work in progress, given that the vast majority of the sources remain unedited. Moreover, the place given to the *Topics* in the University curricula varied throughout the Middle Ages (Green-Pedersen, 1984, pp. 87–93; Ebbesen, 1997, p. 337). In the following pages, then, we will focus on some key aspects of that tradition. We will give a transcription of the manuscript when using an unedited source.

¹³ Boethius of Dacia, *Quaest. Top.*, Prooemium: *Dialecticus ipsam veritatem non considerat (...) sed docet modum, quem artifex specialis debet materiae speciali applicare ad veritatis inquisitionem*. The anachronism "a given professional" for *artifex specialis* indicates here a person who masters a particular scientific discipline.

¹⁴ Radulfus Brito, *Quaestiones super librum Topicorum*, q. 36, ms. Paris, BNF lat. 11132, f. 24vb: Every proposition in which the assent of the answerer is required is dialectical insofar as its formulation is concerned. But in any proposition, it is possible to ask for the assent of the answerer. Therefore, <every proposition is dialectical.> The major premise is evident, because in a dialectical proposition the assent of the answerer is required, and therefore every proposition thus

Brito, cannot be the sole criterion to determine whether a premise or a problem is dialectic, because dialectic would be either superfluous or virtually omnipresent. For this reason, Brito argues, a problem is dialectical in virtue of the way in which it is solved (*terminatur*), that is, through common and probable reasons.¹⁵ As Brito points out, a problem is ethical in virtue of its subject matter, but it can also be dialectical insofar as it determined by means of common topical relations and probable premises.¹⁶

But what is meant here by “common” and “probable”? With “common”, medieval scholars referred to argumentative schemata that could be applied to any scientific issue regardless of its specific subject-matter. These schemata are divided according to Aristotle’s division of the predicables (i.e., accident, genus, definition, and proprium), presented and discussed in the central books (II-VII) of the *Topics*.

With “probable”, they qualified (through the lens of Avicenna) a proposition which is conceded as true with the fear of its contradictory

asked is dialectical insofar as its formulation is concerned. The minor <premise> is clear, because the assent of the answerer can be asked about any proposition, be it true, false, necessary or probable [*Omnis propositio in qua interrogatur de consensu respondentis quantum ad modum interrogandi est dialectica. Sed in propositione potest quaeri de consensu respondentis. Ideo, etc. Maior patet quia in propositione dialectica interrogatur de consensu respondentis, et ideo omnis illa propositio in qua sic interrogatur est dialectica quantum ad modum interrogandi. Minor apparet quia sive propositio sit vera sive falsa sive necessaria sive probabilis potest interrogari de consensu respondentis.*]

¹⁵ Radulfus Brito, *Quaest. Top.*, q. 36, ms. Paris, BNF lat. 11132, f. 24vb: If we speak of a proposition in the second way, that is, insofar as it can be judged as true or false in virtue of probable signs, then every proposition is dialectical, since every probable proposition, insofar as it is probable, is dialectical (...) If it is solved through common topical dispositions, such a problem is dialectical, because the dialectician, qua dialectician, considers common second intentions applicable to any subject-matter. Therefore, a problem thus solved is dialectical. [*Si loquamur de propositione secundo modo, scilicet ut iudicatur vera vel falsa per aliqua signa probabilis, sic omnis propositio est dialectica, quia omnis propositio probabilis, ut probabilis est, est dialectica. (...) Si terminatur per habitudines locales communes, tale problema est dialecticum, quia dialecticus per se considerat intentiones secundas communes applicabiles ad quamlibet materiam. Ergo problema ut sic terminatum est dialecticum.*]

¹⁶ Radulfus Brito, *Quaest. Top.*, q. 36, ms. Paris, BNF lat. 11132, f. 25ra: And when it is said that every proposition belonging to the natural or moral sciences is also dialectical, it is true insofar as they are solved through common topical dispositions; but if they are solved through the principles of their own science, then some propositions are natural, some moral, etc. [*Et cum dicitur quod omnis propositio naturalis vel moralis est dialectica, verum est ut terminatur per habitudines locales communes; si terminatur penes propria principia, sic quaedam sunt naturalia, quaedam moralia, et sic de aliis.*]

opposite being the case (Porro, 2015). This “fear of the contradictory opposite” (*formido oppositii*) is crucial in the medieval understanding of dialectical problems: A problem is dialectical if the question “Is it the case that p or not?” is decided in favour of one member of the contradictory pair ($p, \sim p$) with the fear of the other member being true, this fear being an individual attitude towards p or $\sim p$.

In the fourteenth century, John Buridan introduces another notion that helps demarcate dialectical problems and premises from non-dialectical ones, namely the notion of the margin of certainty (*latitudo certitudinis*):

Fourth conclusion: likewise, not all premises are dialectical. This is proved, because some premises are so evidently true, that the assent by which we accept them is, without qualification, not below the margin of certainty, such as the first principle. And such a premise is not dialectical, since for a premise to be dialectical it is required that it be accepted, without qualification, below the margin of certainty.¹⁷ (Green-Pedersen, 2008, p. 56)

For Buridan, what characterises the dialectical premise is that the assent given to it is below the margin of certainty (i.e. with the fear of its contradictory opposite being the case). The premises that are accepted above such a margin are those that are evidently true, and hence demonstrative. Below that margin, propositions are merely probable, and hence dialectical. It is noteworthy, then, that some propositions will *per se* force assent with a level of certainty above the margin, and some will not, so that the level of certainty that accompanies assent is at least partly due to an intrinsic epistemic quality of the propositions at stake (with the principle of non-contradiction as the usual example of self-evident proposition producing the highest degree of certainty).

It is also noteworthy that in these medieval accounts the difference between dialectical problems and premises, clearly introduced by Aristotle in terms of social disagreement, vanishes; for any scientific statement can become a dialectical problem or premise, depending on its formulation and a given enquiring context. For instance,

¹⁷ John Buridan, *Quaestiones Topicorum*, q. 13: *Quarta conclusio: similiter non omnis propositio est dialectica. Probatur, quia aliqua est ita evidenter vera, quod assensus, quo ei assentimus, non est infra latitudinem certitudinis simpliciter, sicut est primum principium; et talis non est dialectica ex eo, quod ad hoc, quod aliqua sit propositio dialectica requiritur, quod ei assentiatur infra latitudinem certitudinis simpliciter.* In the *sed contra* section of his discussion, Buridan acknowledges that one can deny the certainty of the first principle only sophistically.

in medieval psychology the proposition “Intellection is a kind of affection” (*intelligere est quoddam pati*) is, on the one hand, often introduced as premise of a discussion. On the other hand, it is introduced just as often as the problem to be discussed, notably in question-commentaries on Aristotle’s *On the Soul* that raise the question “Whether intellection is a kind of affection or not” (*utrum intelligere est quoddam pati*).¹⁸ So, a probable proposition (whether in terms of fear of the opposite or of margin of certainty) can be turned into a matter of disagreement, hence of inquiry, if formulated as a problem (notably through the apposition of ‘*utrum*’); or introduced as a premise, if formulated as such. The difference between a premise and a problem is, thus, functional. As a result, in the medieval dialectical practice any proposition which has not been proved through demonstration can still be turned into a problem and submitted to scientific scrutiny.

This brings us back to where we started: In the late Middle Ages the dialectical discussion is understood as a tool for the production and dissemination of scientific knowledge. As such, it is mainly a scientific discussion that takes place in a social institution, the medieval university, by means of an exchange between master and students following strict formal and contextual rules (i.e., ways of formulation, time limits, etc.). Many medieval readers of the *Topics* noted indeed structural similarities between the dialectical discussion and the different forms of the *quaestio* which give a format to some medieval scientific debates.¹⁹ In this framework, suitable problems and premises are crucially related to individual epistemic attitudes towards certain scientific propositions which are at least partly due to intrinsic qualities of those propositions. This focus on the scientific enterprise, in opposition to Aristotle’s focus on gymnastic training in the *Topics*, explains, then, the medieval departure from Aristotle’s characterisation of dialectical problems and premises in socio-epistemic terms.²⁰

¹⁸ For some medieval commentaries on Aristotle’s *On the Soul* that raise this question, see Mora-Márquez (2014).

¹⁹ Angelus of Camerino and the anonymous compiler of the Ripoll-Compendium (both from ca. the middle of the 13th century) are among the earliest commentators who equate a dialectical problem with the *quaestio*-form, in which the particle *utrum* is the mark of an opposition between two contradictory stances (Fernández Walker, 2017, pp. 321–322).

²⁰ There are of course medieval readings of the *Topics* that focus on gymnastic training (it is, after all, one the three goals explicitly stated by Aristotle at the beginning of the treatise). One example is the type of scholarly debate known as *obligatio*, which some medieval authors (Boethius of Dacia and the author of the commentary transmitted in ms. Firenze, BNC, Conv. soppr. B.4. 1618, pp. 95a–151b, here p. 145b) find directly derived from the rules stated by Aristotle in *Topics* VIII. For a discussion on Medieval *obligationes*, see Dutilh Novaes,

The scientific use of the dialectical discussion is still at the centre in the late fourteenth century, and clearly illustrated in the commentary on the *Topics* by Hartlevus de Marca, a follower of Buridan and the first rector of the University of Cologne. Hartlevus' characterisation of the dialectical problem is as follows:

In this question, it must be noted first that 'problem' can be understood in three ways: in a common way, in a proper way, and in a more proper way. Taken in a common way, [a problem] is the enunciation of a proposition and its contradictory in the form of a disjunction. And, thus, one can build a problem from a first principle, for instance, "whether a whole is bigger than its part or not", just as a problem can be built from a false proposition, or a proposition evidently impossible. A problem in a proper sense is the enunciation of a *doubtful* proposition and its contradictory in the form of a disjunction, and thus a problem not determinable by human understanding is still a problem, for instance "whether the number of all the stars is even or not". But more properly, a problem is the enunciation of a *doubtful* proposition *determinable* by dialectical arguments and its contradictory proposed as a disjunction, and this is how 'problem' is considered here.²¹

In a first sense, *problema communiter acceptum*, 'dialectical problem' refers to the formulation, and it involves virtually any problem (cf. Radulfus Brito's *modus interrogandi*). This, as we have seen, is not a characterisation at all, since it involves all possible questions with the form "Is *p* the case or not?". This sense covers even the problems raised in the context of obligational disputations, which may be useful for gymnastic, but not for scientific purposes (he even introduces as an

Uckelman (2016); Dutilh Novaes (2011); Keffer (2001); Yrjönsuuri (1993). It is worth mentioning however that, even when the gymnastic approach of the *Topics* is acknowledged by medieval authors commenting on Book VIII, it always remains ancillary to the main goal of dialectic, namely epistemic enhancement.

²¹ Hartlevus of Marca, *Quaestiones libri Topicorum*, I, q. 9, ms. Erfurt, Universitätsbibliothek, Dep. Erf. CA. 4° 270, f. 88ra: In ista questione est primo notandum quod problema capitur tripliciter: communiter, proprie et magis [proprie]. Communiter acceptum [problema] est enuntiatio propositionis cum suo contradictorio disiunctive. Et sic ex primo principio potest fieri problema, scilicet an totum est maius sua parte vel non, sicut etiam ex propositione falsa vel evidenter impossibili potest fieri problema. Problema proprie acceptum est enuntiatio propositionis dubie cum suo contradictorio disiunctive, et sic problema quando non est terminabile per humanam considerationem est problema, scilicet an omnia astra sunt paria vel non. Problema magis proprie est enuntiatio propositionis dubie terminabilis per argumentationem dialecticam cum suo contradictorio disiunctive proposito, et sic accipitur hic problema.

example the problematization of the axiomatic statement "the whole is bigger than its part"). Thus, Hartlevus introduces a proper sense, *problema proprie acceptum*, which covers all problems involving probable statements, including also those not determinable by human understanding (here the example is the number of stars being odd or even). But problems that cannot be possibly determined within the limits of human reasoning are scientifically irrelevant. So, he arrives at the most proper sense, *problema magis proprie acceptum*: Dialectical problems are most properly those involving exclusively scientific propositions, and which can be determined through probable premises. In other words, a problem is most properly dialectical when it is relevant for the current scientific practice.

3. CONCLUSION

A quick glance at medieval readings of Aristotle's *Topics* shows that, for the most part, commentators stressed the importance of dialectic for the scientific practice. While in Aristotle a social disagreement was at the bottom of any dialectical problem, in the Middle Ages any proposition not yet proved through demonstration was susceptible of problematic treatment, regardless of the social stance towards it.

This does not necessarily mean that Aristotle's text was misread. While we argue that Aristotle's main focus in the *Topics* was the gymnastic aspect of dialectic, medieval readers focused on the scientific use Aristotle himself considers in *Topics* I.2.101a25–26, where he says:

Next in order (...) would be to state the number and kinds of things our study is useful for. There are, then, three of these: exercise, encounters, and the philosophical sciences." (Smith 1997, p. 2)

Some Aristotelian scholars today also focus on dialectic as a scientific tool, among which Barnes 1980 and Bolton 1990 are notorious. We hope to have shown that their view has a venerable precedent in the medieval tradition of commentaries on the *Topics*, where dialectic is understood mostly as a method for the scholarly discussion, through probable arguments, about the *endoxa* of the medieval sciences.

ACKNOWLEDGEMENTS: We thank Knut and Alice Wallenberg Foundation, Sweden, who funded the research that led to this paper.

REFERENCES

Allen, J. (2007). Aristotle on the Disciplines of Argument: Rhetoric, Dialectic, Analytic. *Rhetorica*, 25, 87–108.

- Barnes, J. (1980). Aristotle and the Methods of Ethics. *Revue internationale de philosophie*, 34, 490–511.
- Bolton, R. (1990). The Epistemological Basis of Aristotelian Dialectic. In Devereux, D. & Pellegrin, P. (Eds.), *Biologie, logique et métaphysique chez Aristote* (pp. 185–236), Paris: Editions du CNRS.
- Brunschwig, J. (1967). *Aristote, Topiques. Tome I*. Paris: Les Belles Lettres.
- Dutilh Novaes, C. (2011). Medieval *Obligationes* as a Theory of Discursive Commitment Management. *Vivarium*, 49, 240–257.
- Dutilh Novaes, C. & Uckelman, S. (2016). *Obligationes*. In Dutilh Novaes, C. & Read, S. (Eds.), *The Cambridge Companion to Medieval Logic* (pp. 370–195), Cambridge: Cambridge University Press.
- Ebbesen, S. (1997). The *Ars Nova* in the 'Ripoll Compendium'. In Lafleur, C. (Ed.), *L'enseignement de la philosophie au XIIIe siècle. Autour du 'Guide de l'étudiant' du ms. Ripoll 109* (pp. 325–352), Turnhout: Brepols.
- Fernández Walker, G. (2017). Dialectical Problems in Paris and Cologne. Commentaries on *Topics I*.11. *Cahiers de l'Institut du Moyen-Âge Grec et Latin*, 86, 313–348.
- Green-Pedersen, N. (2008). *Johannes Buridanus: Quaestiones Topicorum*, Turnhout: Brepols.
- Green-Pedersen, N. (1984). *The Tradition of the Topics in the Middle Ages. The Commentaries on Aristotle's and Boethius' 'Topics'*, Munich: Philosophia.
- Green-Pedersen, N., & Pinborg, J. (1976). *Boethius Dacus: Quaestiones super librum Topicorum*, Copenhagen: GAD.
- Karbowsky, J. Complexity and Progression in Aristotle's Treatment of *Endoxa* in the *Topics*. *Ancient Philosophy*, 35, 75–96.
- Keffer, H. (2001). *De obligationibus. Rekonstruktion einer spätmittelalterlichen Disputationstheorie*. Leiden: Brill.
- King, C.G. (2013). False endoxa and fallacious argumentation. *Logical Analysis and the History of Philosophy*, 15, 185–199.
- Mora-Márquez, A. M. (Forthcoming). Elements of (Dialectical) Argumentation Theory in Aristotle's *Topics*. In Gazziero, L. (Ed.) *Le langage. Lectures d'Aristote*, Leuven: Peeters.
- Mora-Márquez, A. M. (2014). A List of Commentaries on Aristotle's *De anima* III (c. 1200–c. 1400). *Cahiers de l'Institut du Moyen Âge Grec et Latin*, 83, 207–256.
- Owen, G. E. L. Tithenai ta phainomena. In Mansion, S. (Ed.) *Aristote et les problèmes de méthode* (pp. 83–103), Leuven: Symposium Aristotelicum.
- Porro, P. (2015). Il timore dall'altra parte. Il ruolo della *formido* nei dibattiti scolastici sull'assenso (da Tommaso d'Aquino a Pietro Aureolo). *Archivio di Filosofia*, 83(1–2), 209–220.
- Primavesi, O. (1996). *Die Aristotelische Topik. Ein Interpretationsmodell und seine Erprobung am Beispiel von Topik B*. Munich: Beck.
- Rubinelli, S. (2009). *The Classical Technique of Constructing Arguments from Aristotle to Cicero*, Dordrecht, Springer.
- Slomkowski, P. (1997). *Aristotle's Topics*. Leiden: Brill.
- Smith, R. (1997). *Aristotle. Topics I and VIII* translated with a commentary. Oxford: Clarendon Press.
- Yrjönsuuri, M. (1993). Aristotle's *Topics* and Medieval Obligational Disputations. *Synthese*, 96 (1), 59–82.

MANUSCRIPTS

Ms. Erfurt, Universitätsbibliothek, Dep. Erf. CA. 4° 270, ff. 76ra–117va.

Ms. Firenze, Biblioteca Nazionale Centrale, Conv. soppr. B.4. 1618, pp. 95a–151b.

Ms. Paris, Bibliothèque nationale de France, lat. 11132, ff. 11ra–52ra.

Conductive Arguments. Towards a solution of the riddles

HARALD R. WOHLRAPP

Philosophisches Seminar, Universität Hamburg, Germany
wohlrapp@uni-hamburg.de

Five kinds of dissent about Conductive arguments are presented, the dissent about the appropriate approach being the most fundamental. It can be resolved under a “dialectical-pragmatist approach” – a combination of dialectical and pragmatist philosophies. In the light of that approach three steps of conceiving the operation of “balancing (outweighing) considerations” are proposed: (a) Ranking instead of weighing, (b) Unification of perspectives, (c) Focus on a particular case.

KEYWORDS:

Allen, dialectic, Govier, particular case clause, Possin, pragmatism, unification of perspectives

1. INTRODUCTION

In his two books (Wellman, 1971, 1975) Carl Wellman had given a highly enigmatic definition of “Conductive Arguments” (hereafter: CA), together with some theoretical suggestions and extensive discussions of examples. Since then theorists of argumentation have forged ahead to come to terms with his ideas. The discussion was entirely shaped by the way in which Trudy Govier had apprehended the CA (Govier, 1987). Her pattern was accepted by a lot of followers but then also some criticism was uttered and with further theoretical endeavor a large field of controversial views emerged, partly elaborating and partly criticizing Govier’s treatment or even the whole idea of the CA.

We can identify five kinds of notorious disagreements about CA:

- Wellman-Exegesis (3 patterns, convergency, independence, particular case clause)
- Counter-considerations (premises, counter-arguments, internal vs. external parts of the argument)
- O-B Principle (Scheme for inference from Pros/Cons to conclusion, warrant)
- Assessment of CA (strength/ force/ importance of arguments,

- the mysterious operation of “outweighing” them)
- Appropriate Approach (product vs. process; monological vs. dialectical)

2. THE STRUGGLE ABOUT THE APPROPRIATE APPROACH

2.1 *Product vs. process approach*

“Product approach” means to conceive arguments as PPC- patterns:

Premise group – Inference – Conclusion

“Process approach” means to conceive arguments as sequences of moves in which (parts of) arguments, related to a thesis, are forwarded, considered and eventually lead to a conclusion.

Thesis – A_1 – A_2 – ... A_n – Conclusion

Product view and process view are *prima facie* opposing ways to conceive arguments. In fact, even in the process view each move is based on an evaluation of the respective (part of) argument; and can therefore be seen as a “product”. If, however, among those arguments there are some that are meant to “converge” (accumulate, summarize), then a transition to a final conclusion can be problematic (an inference scheme seems to be necessary – and is not at hand). If there are even some arguments that speak against the thesis, then the inference to a conclusion is, theoretically, a full-blown riddle.

In daily practice we are used to ignore that riddle and produce a conclusion via “outweighing” the pros against the cons. It is, however, not clear at all what kind of operation this is.

In Wellman’s work there are prominent traces of both approaches. He speaks about “outweighing”, even of a “heft-weight” that arguments may have; but at the same time he claims that only the “thinking trough” of the Pros and Cons (again and again) can make us feel the “logical force” (also: the “psychological force”) of the arguments – so that eventually a conclusion appears obvious.

Govier had shaped the discussion with naturally taking up a product view. Moreover, she used the “outweighing” operation in an intuitive way, appealing to the judgement forces of the arguers. This solution goes for everyday reasoning (take a taxi to be in time, even if it is very costly), but not for the serious questions of our life (abortion, euthana-

sia, economic inequality).

Govier's pattern was criticized, also as focusing on the product view. In the late 80ies and early 90ies several critical articles, notably by the author of the present paper, were published. In the late 90ies, Ralph Johnson, so far an adherent of the product approach, began to demand a "dialectical tier" for any argumentative endeavor.¹

And while Johnson was recognized by Govier and (at least partly) integrated², Wohlrapp was ignored and eventually misjudged as being an adherent of a pure process approach.³

2.2. *Monological vs. dialectical approach*

These two approaches are also *prima facie* opposing each other. In current argumentation theory "dialectical" means no more than "dialogical"⁴, viz. a setting with a proponent and a (critical) opponent. At the same time the concept of a dialogue is occupied by all kinds of "dialogue games" which are governed by prefabricated rules and, thus, a static setup (this began with Hamblin⁵, continued with Barth/Krabbe⁶, then Krabbe/ Walton⁷ etc.). In my opinion this kind of "dialogue" is of little use in the concept of CA.

Thus, it seems that the controversy between a monological and a dialectical view ("dialectical" in this sense) is indeed a variant of the opposition between product and process approach.

A new proposal to see CA in a dialectical (viz: dialogical) form stems from Yu/ Zenker. They seem to be determined to conceive CA as a process; but what they propose in the end is indeed a pattern for a

¹ Johnson (2000), pp. 164ff: "Rethinking the Nature of Argument".

² Govier (1999), Chap. 12: "Becoming dialectical. Two tiers of argument appraisal?"

³ See Govier (2017), pp. 213ff, see Blair (2017), Part 7.2. Blair's criticism is obscure, because he was the editor of Wohlrapp (2017). There, in Part III, contradicting Govier, there is a long explanation of why a combination of product and process approach is necessary. The latest example for erroneously ascribing Wohlrapp's theory a pure process view is Yu/ Zenker (2019), p. 36.

⁴ „Pragmadiagnostics“ is a combination of a dialogical setting with speech-act pragmatics. It lacks profound dialectical thinking as well as profound pragmatist thinking; see my criticism in Wohlrapp (2014), LIII-LVIII.

⁵ Hamblin (1970).

⁶ Barth/Krabbe (1983).

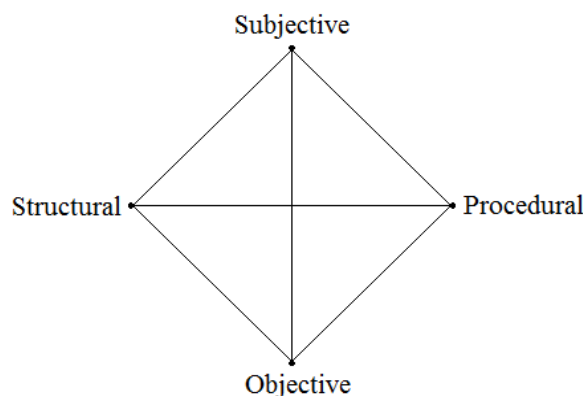
⁷ Krabbe/ Walton (1995) (Commitment in Dialogue)

product, reached in a way whose exact comprehension, they believe, „goes beyond what *argumentation* scholarship should offer“⁸.

2.3. Dialectical-pragmatist approach

This approach has been developed in the building of a philosophical foundation of argumentation theory. It is a combination of elements of dialectical with pragmatist philosophy. Compared with the most influential schools (Informal Logic and Pragma-dialectics) it covers two fairly understated sides of argumentation practice, namely dynamics and subjectivity. Argumentation practice is then to be seen in two pairs of polar dimensions.

Diagram: Square of dimensions:



My claim is now that with thinking in terms of this approach the problems of CA can find a reasonable solution. In order to show this, I will start with very shortly characterizing that approach. Due to the necessary terseness, I will here confine myself with two invitations to adopt a certain thought style (more is in Wohlrapp, 2014).

Pragmatist feature:

Anchor the reliability of all theoretical results (their correctness, truth, appropriateness) in action faculties! Knowledge is verbally stylized action potential. Thus, argumentation is a verbal practice embedded in non-verbal activities of human life. It aims at maintaining our orientation in the world beyond established knowledge.

Dialectical feature:

Be aware of the reflective character of all human intellectual activity! Reflection is not simply self-reference but it is the awareness of one-self

⁸ Yu, S./ Zenker, F., A (2019), p. 61.

as being able to increasing levels of self-reference and to, thus, becoming a self-conscious subject (an individual). Dialectical thinking in argument leads to considering theses and arguments in a true dialogue (with oneself and/or others) in which all concepts, rules, views, and persons are kept open to change (as far as possible).

3. CONCEPTION OF CONDUCTION

This part is concerned about the consequences of the Dialectical-Pragmatist-approach for the conception of the CA. The most prominent consequence is:

3.1. Reconsideration of product-process distinction

The common confrontation of the two separated views is erroneous, both are necessary and interdependent phases of argumentative practice. As such they deserve proper designations.

My terms are here:

Process phase: "Discussion"; Product phase: "Inferential Setup".

I will shortly characterize both.

3.1.1 Discussion

A thesis (Th) is doubted or contested and dialogue partners (P and O, not necessarily two persons) utter arguments ($A_1 \dots A_n$) to justify it (Pros) or to criticize it (Cons). All Arguments have the form:

$$A_m \text{ (if correct/acceptable/true)} \Rightarrow \text{Th is (not) c./a./t.}$$

Example:

You promised to take your son to the movies \Rightarrow you should do so.

Sometimes the material implication (\Rightarrow), is not sufficiently based.

(This example: Are promises really categorical commitments – even if the situation has considerably changed?).

If so, we may mobilize further arguments. This may lead to the problem of convergence (accumulation) of arguments. In my opinion its solution has to be seen in an improvement of the theoretical basis about the contested matter – yet, this is not treated in the present paper.

Discussion process exhibits some remarkable features:

a) "Retroflexive structure"

In the process the arguments are not simply collected and listed, but considered (how is this justified?) and confronted with their possible

counterparts. This may encourage or even necessitate all kinds of modifications (incl. droppings) of arguments; and even of the thesis itself: Thus, we get a continuous forwards and backwards reasoning in which (1) the relevant arguments are identified and given an appropriate form and (2) an appropriate variant of the thesis may be found.

b) “Frame” differences

Frames have not only a rhetorical, they also have an epistemological meaning – as they raise bounds and limits for the consideration (even the perception) of any issue. The contents of Pro/Con arguments can often come in different frames – a difference that has to be overcome⁹.

c) “Maieutic function”¹⁰

In discussion we can become aware of what we “really believe”. In a sense we could not know it before: Because I have my beliefs on my personal background system and I do not know how they look in your perspective which I now get confronted with. Thus, in discussing I can learn to see my beliefs in a new and more comprising way.

3.1.2 Inferential Setup

At any move of the discussion a break is possible and we can ask: What is the present state of the discussion (what is its product so far?)– can we summarize what we have achieved until now?

This “state of discussion” consists of a list of Pros and Cons – not of any possible arguments, but of only the germane ones, that have withstood the discussion so far. Now a judgement is queried if from these arguments the thesis is attainable – maybe in another modified version – as a conclusion.

(Distinction: “Thesis” is the vantage point, “Conclusion” the wanted result)

T
P/C
C

Realistically observed, the cases are rare in which a unified conclusion shows up. Usually we hang around with certain Pros and Cons. What is mostly produced then is not a materially derived conclusion but it is a “decision” – be it authoritative (parent, teacher, boss) or democratic

⁹ See Wohlrapp (2014), Chap. 5.

¹⁰ Walton, D. /Krabbe, E. C. (1995)

(majority vote).

Very often this kind of decision is then proclaimed as being the result of “carefully outweighing” all the arguments, yet, nobody could tell what this refers to. Theoretically this is a complete riddle.

4. TOWARDS A SOLUTION OF CONDUCTION

My proposed solution for the riddle of conduction has three elements:

- Ranking instead of weighing
- Unification of perspectives
- Focus on particular cases

4.1. *Ranking instead of weighing*

“Outweighing” arguments is affiliated to speaking about a “weight” (strength, force, importance) of arguments. This is clearly a metaphor but its signal is, that there is some mystical unit (Hitchcock: “graviton”¹¹) or standard against that we can come to an answer about which argument can trump which other argument.

Wellman, in some of his attempts, spoke about “heftweighing” the arguments as a preparation of their outweighing.

Tom Fisher has tried to take this literally¹². He refers to the fact that for many issues we use “non-numerical, comparative quantitative categories” (e.g. the weight of a thing: light, medium heavy, very heavy etc.) At least he is confident, that they might function more or less as degrees.¹³ From here, I think, one step further could lead us to a tenable answer.

I was inspired for that step by Fred Kauffeld. He argues that there is no weight (strength, force, relevance etc.) “in” the arguments by which we could compare them. Still, we are often able to ranking them according to our preferences: We use to have “paramount concerns” that, by their very nature, “dominate other considerations”.¹⁴

The striking difference between the two modes of comparing is this: Ranking does not involve a *metrical* comparison, but *only a topological* one.

¹¹ Hitchcock (2016): Chap. 8 “Weighing”.

¹² Fischer (2011).

¹³ Govier had asked „What are these degrees anyway? There is no answer.” See Fisher (2011), p. 89.

¹⁴ Kauffeld (2011), 166

The latter relation is closer than the former to what we are really “doing” when we search for a “balance of considerations”. Thus, pragmatic thinking is the door-opener for this replacement.

Still, an attempt to produce a ranking between several arguments (especially Pros against Cons) results in a more or less subjective conclusion. This is *prima facie* not really satisfying, because, as Tom Fischer is rightly reminding: argument “constitutes par excellence the process of moving from subjectivity to objectivity” (Fischer, 2011, p. 101)). The intention to fulfill that demand leads me to the second element of the present proposal.

4.2. Unification of perspectives

In the discussion about arguments – and their possible conduction – we participate on the basis of our subjective systems. There is no other possibility. Why: Even the most serious attempt to judge objectively is done on the basis of a (subjective) conception of “objectivity”.

Insofar, an argument with its ranking over/under other arguments is always an “argument for me”. The predicate “argument” appears to be a triple relation:

Argument A – for/against thesis Th – considered by subject S

As I said above, the ranking refers to a “preference system”. This is a term established e.g. in decision theory. For our theorizing about subjectivity in argument we should, however, be aware, that a person usually does not have such a thing. We are not able to say in a general way if we prefer X over Y – e.g. a safer car over a faster or a more beautiful one. (The *ceteris paribus* clause is an illusion.)

Concerning the possession of an elaborated preference system: For one this is at best in some cases so but not in all. For the second we do not know it precisely, even if we can make an ad-hoc decision. And for the third: its details heavily depend on the particular case (see below).

I have said above, under the heading of the “maieutic function” that in discussion we may get to know better, what we believe – and, thus: who we are. Particularly the debate over an important question (organ donation, climate politics) can contribute to greater self-knowledge and stabilization of the subject system.

What is needed here: A unification of the subjectivities of the arguers in the light of a disputed question. As long as this unification is not achieved, the rankings of the participants remain private and more or less tentative.

Let me consider two examples about the achievement of that kind of unification:

Movie example:

You should take your son to the movies.

Pro: You have promised it.

Con: Your lawn needs to be cut urgently.

A resolution may be found because a sound family, by its internal cohesion, is sufficiently united to find a solution for their dissent; e.g. the son is promised to be taken *two times* to cinema at the next occasion, or the wife is promised that the lawn will definitely be cut *before the weekend*....

Grading example:

Christian Kock (Kock, 2011, p. 70) speaks about problems, when a team among his academic colleagues had to decide about a grading for a student's exam. There were arguments Pro (richness of information, originality) and arguments Con (bad organization, bad spelling). Yet, the discussion showed that their impact on the grade was differently judged. E.g. for one the bad spelling was vital, while his colleague was nearly not fussed about it.

I think one has to look closer into the circumstances of the case. Maybe an agreement about preferences could be found; e.g. of putting the spelling below the originality of the content: if the candidate was a foreigner.

(In my university the uses are, that, if no agreement is attainable, a third judge is asked and then the majority decides.)

Result: If a unification of the subject is available, a preference relation is either at hand or it can be established by cooperative deliberation. If not, then "conductive argumentation" is *no more an argumentative endeavor*. Then a "decision" is necessary (authoritative or via majority vote).

4.3 Focus on "the particular case"

This is, in my opinion, the most important element in a reasonable treatment of conduction. Insofar, it is a sad or ironic fact, that Trudy Govier, at her very start with that topic, had firmly refrained from this "particular case clause". (I will not discuss her reasons for this, I think they were somehow shortsighted. For criticism and some hints about a different view on that clause see Possin (2016). Still, I think that his differentiation between type and token is only one hint to a more adequate treatment of that clause.)

In the subsequent endeavor of argumentation scholars it was, again and again, put aside, and even those who made “case studies” about conduction (Allen, Kauffeld, Walton), were not necessarily aware of the role of the particular case clause. Allen had at least clearly stated that assessing conductions for reaching a juridical verdict “requires judges to remain rooted in the particularities of the case at hand”¹⁵.

The reason why we need the reference to a particular case is this: In a reasonable discussion we do not simply collect the arguments and then ponder the group of Pros against the group of Cons, but we consider each single argument of one side and ponder it against arguments of the other side, trying to find out which ones are supported or restricted or even refuted by any other argument. If there are massive contradictions between Pros and Cons, we have to investigate deeper and deeper into how it could be possible to believe in their correctness (acceptability/truth). Like this the thesis will gradually show its full embedment in the subject systems of the participants. There, its acceptance or rejection will be justified with generic sentences and, the deeper we come, with ever more foundational principles.

If a Pro/ Con setting turns out to be justified with contradicting basic principles in the subject systems of the participants, then the only way out of that “deep disagreement” is to differentiate the principles with regard to the particular case (that gave rise to the struggle about the thesis). Still, there is never a guarantee that enough differentiations are found or constructed so that the principles can become *compatible with regard to the particular case*. Even though, this is, I think, the only reasonable way for overcoming the split of the subjectivities and, thus, eventually produce a unified conclusion. I will illustrate what I mean, first, again, with the movie example; and then, very shortly, touch Derek Allen’s meritorious analysis of a conductive argumentation for a judgement of Canada’s Supreme Court.

4.3.1 Movie-example

Thesis: You should take your son to the movies.

P1: You promised it.

P2: You have nothing better to do.

P3: The movie is ideal for children.

C1: The lawn needs cutting.

C2: The boy has not yet made his homework for school.

¹⁵ Allen (2011), p. 189

A discussion of these arguments may discover the following general rules being used in their justification:

P1: Promises have to be kept.

P2: We should fill empty time with valuable activities.

P3: The film was recommended by school (school recommendations are good).

C1: The lawn has to be cut regularly, it's high time already.

C2: Homework has to be made right after lunch.

As I said before, the usual internal cohesion of a sound family provides some basic *unity of perspectives* to the extent that it should be possible to find differentiations with regard to this case and, thus, mitigate the rigidity of any of those rules. If the family can agree that the condition of the lawn is urgent and that homework has already been treated too carelessly last week, then the two Con arguments will clearly trump over the three Pro arguments in the following way:

The general rule which underlies P1 can be enriched with the possibility of exceptions for relevant changes of the situation. After this, the miserable condition of the lawn could be accepted as a relevant change. The second Pro proves flatly invalid with regard to the lawn urgency. And the third Pro will be refuted by the consideration, that, under the given circumstances, homework is more important than a movie (and also a recommended one).

4.3.2 Supreme Court example

Derek Allen has published an article in which he analyses a very important and interesting case of argumentative conduction in a judgement of Canada's Supreme Court¹⁶. In 1984 the Albertan school teacher James Keegstra was convicted for communicating anti-semitic propaganda to his students. The legal basis of the verdict was the "hate law" (s. 319(2)). Keegstra then appealed to the Supreme Court claiming that the *Canadian Charter of rights and freedoms* (under s. 2 (b)) guarantees his freedom of expression and that this right was infringed by the condemnation. Thus, the Court had to reconsider and prove the constitutional validity of the "hate law".

Allen now shows meticulously (and, of course, with far more details than I can display here) how the Court is arguing "conductively": The principle of freedom of expression had to be reconsidered. Yet, the reconsideration had to be done not in a general, abstract manner, but

¹⁶ Allen (2011).

with regard to the particular case, i.e. to the “free” expression of Keegstra’s highly problematic teachings.

It is by no means self-evident that utterances which are heavily disturbing for others do not enjoy the protection of the *Charter*. Thus, *prima facie* there is a big Pro argument for Keegstra’s complaint and one of the judges indeed saw the hate law in principle infringing the *Charter*’s guarantee (of s. 2(b)). The leading question was “whether the expression prohibited by the hate law is closely linked to the values underlying the *Charter* guarantee of freedom of expression” or not (Allen, 2011, p. 179).

Therefore, a re-examination of those values was provided: “attainment of truth and the common good; individual self-fulfillment and human flourishing; participation in social and political decision making” (Allen, 2011, p. 180). Surprisingly, this re-examination allowed to build an even more specific Pro argument for Keegstra’s case because his teachings were seriously meant to encourage political participation of the students. But then the character of the incriminated utterances was characterized as standing in “a most extreme opposition to the idea that members of identifiable groups (here: Jews, H.W.) should enjoy the self-fulfillment aspect of the free expression guarantee” (Allen, 2011, p. 180).

Despite the unanimous evaluation of those utterances the hate law’s constitutionality was differently judged (against the requirement of the “Oakes test”), so that the Court finally concluded (in a 4-3 decision) “that the hate law was constitutionally valid” (Allen, 2011, p. 175). Keegstra’s appeal was quashed.

The structure of the case is the following:

Two laws – the guarantee for freedom of expression and the prohibition of hate speech – are seemingly contradicting each other in view of this particular case. Thus, the possibility arises that the “hate law” be constitutionally invalid – and Keegstra’s attorney tries to defend the respective thesis. There are arguments for and arguments against this thesis. But then two closer investigations are undertaken. First, the right of freedom of expression is examined with regard to the specific values that it is supposed to represent; and second, the incriminated expression is reconsidered to find out if it is compatible with those values. The answer was, that it is not.

We should clearly see that in the Court’s judgement the Charter freedom of expression is not restricted. Rather it is reformulated in a more differentiated manner so that it becomes evident: Mr. Keegstra’s kind of expression is not protected by the Charter right.

It is important to consider the nature of the decision and to correctly

assess the relationship between the general normative outline of the Charter right on one hand and the prohibition of Keegstra's teaching on the other:

It was not James Keegstra's individual expression, but the "kind of" his expression, which was judged as not being protected. Nevertheless, it is clear, that the "balancing of considerations" which led the Court to its verdict, depended strictly on the particular case.

5. CONCLUSION

Some of the main problems of the CA have been designated. The dispute about the appropriateness of approaches (product-process-dialectical) could be mitigated or even overridden by offering a dialectical-pragmatist approach, that enables the inclusion of two additional dimensions of argumentation (dynamics and subjectivity).

After this, the approach was applied to the riddles of conduction. It consists of three elements:

- Replacement of "weighing" with "ranking",
- Unification of subjective perspectives
- Differentiations in view of the particular case.

If these considerations are respected a satisfying elucidation of the dubious "conductive argument" should come into reach.

REFERENCES

- Allen, D. (2011). Conductive Arguments and the Toulmin Model: A Case Study, in: Blair/ Johnson (2011), 167-190.
- Barth, E. /Krabbe, E. (1982) *From Axiom to Dialogue. A philosophical study of logic and argumentation*. Berlin/ New York: de Gruyter.
- Blair, A. (2017). In Defence of Conduction: Two Neglected Features of Argumentation. In: Proceedings of ECA, Zürich 2016
- Blair, A./ Johnson, R. (eds.). (2011). *Conductive Argument. An overlooked type of defeasible reasoning*, London: College Publications.
- Fisher (2011): Tom Fisher, Weighing considerations in Conductive Pro and Con Arguments. In: Blair/ Johnson (2011), 86-103.
- Govier, T. (1987). Two unreceived views about reasoning and argument. In: Govier, T., Problems in Argument Analysis and Evaluation, Dordrecht: Foris, pp. 55-80.
- Govier, T. (1999). *The Philosophy of Argument*, Newport, Virginia: Vale
- Govier, T. (2017). Issues on Logicism and Objectivity. In: *Informal Logic* 37, 3 (2017), pp. 211-222.
- Hamblin, C. L. (1970). *Fallacies*. London: Methuen.

- Hitchcock, D. (2016). All things considered. In: D. Mohammed & M. Lewiński (eds.), *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation, Lisbon, 2015*. London: College Publications, Vol. I, pp. 165-180.
- Kock, Ch. (2011). Why argumentation theory should differentiate types of claim. In: Blair/ Johnson (2011), 62-73.
- Possin, K. (2016). Conductive Arguments: Why is this still a thing? In: *Informal Logic* 36,4 (2016), pp. 563-593.
- Walton, D. /Krabbe, E. C. (1995). *Commitment in Dialogue*, New York: SUP
- Wellman, C. (1971). *Challenge and Response: Justification in Ethics*, Carbondale, ILL: Southern Ill. U.P.
- Wellman, C. (1975). *Morals and Ethics*, Englewood Cliffs: Prentice Hall.
- Wohlrapp, H. (2014). *The Concept of Argument. A philosophical foundation*, Dordrecht: Springer.
- Wohlrapp, H. (2017). Replies to Commentators on *The Concept of Argument*. In: *Informal Logic* 37, 4 (2017), pp. 211-222.
- Yu, S./ Zenker, F. (2019). A Dialectical View on Conduction: Reasons, Warrants, and Normal Suasory Inclinations, in: *Informal Logic*, 39, 1 (2019), pp. 32-69.

A ludological perspective on the shape of argument: Collaborative assent to dissenting opposition

MICHAEL A. YONG-SET

Argumentation Studies PhD, University of Windsor, Canada

yongset@uwindsor.ca

This paper attempts to extend Yong-Set's ludological approach to give a better philosophical articulation of the nebulous tension found in the activity of argumentation. By harnessing the resources of ludology – the academic and critical study of games *qua* games – we can use the framework of 'player-opposition shapes' to help understand the appropriate forms of opposition an antagonist's dissent should take to accomplish the collaborative goals embraced when assenting to participate in various types of argumentation.

KEYWORDS: games, ludology, opposition, collaboration, players, dialogue types

1. INTRODUCTION

There is an intuitive but difficult to articulate tension in the activity of argumentation: in some nebulous way, opposing, disagreeing with, or interfering with the proponent of an argument is not only helpful but necessary. This paper attempts to articulate and address this tension through the cross-application of ludological approaches to argumentation. I argue that we can improve our understanding of argumentation by harnessing the insights of *ludology* – the academic and critical study of games *qua* games, especially in terms of player experience, systems design and the socio-cultural dynamics of gaming. Whether games share a familial relation with argumentation or whether games are merely analogous to argument, it remains the case that we can enrich our understanding of argument by better understanding games.

When analyzed through a ludological lens, it is apparent that all games necessarily contain a minimal degree of both 'collaboration' and 'opposition.' Even the most co-operative or solitary games contain opposition and even the most adversarial games contain collaboration in some minimal degree. But it is the particular arrangement and degree of these elements that differentiates various types of games from one

another. It is useful to differentiate literal games in terms of how the players relate to the other elements of the game system – whether they stand in relations of opposition or collaboration. These ‘player-opposition shapes’ provide a framework by which an analyst can give preliminary form to the otherwise nebulous phenomenon of ‘gameplay.’ Analogously, I argue that taking a ludological perspective can give shape to the otherwise nebulous activity of argumentation.

2. ON THE CONCEPT OF GAMES

The Wittgenstein of the *Philosophical Investigations* is suggestive of an interesting but mysterious notion: if one can see how games work, one can see how natural language works. This sentiment eventually developed into a perspective that has come to be known as the ‘language-game’ theory of language. There are a variety of interesting challenges in taking up the language-game theory of language – most of which are beyond the scope of this paper to address but that I suspect can be well met. One of the most significant difficulties of that theory is as follows: one can look upon games and nevertheless fail to understand how games work, or how the workings of games relate to language use. This confusion is sometimes compounded when Wittgenstein makes exhortations such as:

Don’t say: “[all games] *must* have something in common, or they would not be called ‘games’” – but *look and see* whether there is anything common to all. – For if you look at them, you won’t see something that is common to *all*, but similarities, affinities, and a whole series of them at that. To repeat: don’t think, but look! (Wittgenstein, 1953, S.66).

It is understandable that some may find this advice unhelpful – one may have looked upon children playing games or reflected on one’s own experience in partaking in activities said to be ‘games’ and still been unable to see what Wittgenstein is getting at. *Prima facie*, we are expressly being told not to think about the issue – rather, we should understand the nature of games simply by looking. On the worst of interpretations, it might sound as though games were fundamentally ‘non-cognitive’ and therefore incompatible with any reflective, philosophical or theoretical activity – such as producing analyses of language use or argumentation.

There is a simple way to dissolve this confusion. It will not be *a priori* thinking – thinking that is prior to or independent from experience – that will reveal the nature of games. Rather, we must go out into the world of real, occurring practices to gather evidence and experiences – we must look at, engage in and study various instances of

activities we already believe to be 'games.' And from this collection of empirical experience, we may then begin thinking about, reflecting upon, and constructing frameworks that can help us characterize the matrix of affinities, similarities and features that tend to unite the many activities into one family of 'games.' There should be no problem in principle in attempting to use an understanding of games to enhance our understanding of argument.

Of course, I will not pretend that the notion of using games to further our understanding of argument is new. Many other authors have similarly had the idea that we can better understand one thing – arguments – in terms of the other – games. There is *prima facie* plausibility that the two activities of gaming and arguing share enough relevant functional similarities such that some have even gone as far as suggesting that we reform the very way in which we conceive of 'argument' along the lines of a game, such as tennis or chess (Cohen, 1995; Lakoff and Johnson, 1980; Ritchie, 2003). Ralph Johnson and Tony Blair – two of the seminal figures in the informal logic and natural language argumentation movement – remark, "an argument understood as *product* [...] cannot be properly understood except against the background *process* which produced it [...] the appropriate analogy is a move in a chess game or a play in a football game" (Johnson & Blair, 1987, p. 45, emphasis in original). The prolific Douglas Walton and Erik Krabbe make the game comparison as well: "at this level, types of dialogue coincide with particular dialectical systems or dialogue games" (Walton & Krabbe, 1995, p. 67). Addressing the question of whether argument is related metaphorically, literally or analogously to games is not a question I seek to address in this paper. Rather, the simple point I would like to make is that many argumentation theorists have discerned a *prima facie* plausibility in mixing 'games' into our analyses of argumentation.

As one of the more well-known and well-developed 'game-infused' accounts of argument, Walton and Krabbe's theory of dialogue types warrants some further examination. Their suggestion is that what differentiates persuasion from negotiation, inquiry from quarrel and so on can be discerned by observing the differences that obtain among four axes of 'structural features' (1995). Roughly stated, dialogue types are differentiated based on: 1) the initial situation that prompts the engagement; 2) the main goal of the dialogue; 3) the participants' primary aims; 4) the side benefits that can be accrued by engaging in the dialogue. Their dialectical account of reasoning portends to include interpersonal elements as part of the analysis – this marked a significant shift away from prior accounts of argument that focused exclusively on timeless, impersonal claim-reason complexes. The ever-evolving Waltonian theory of dialogue types has much to recommend it; I believe

there is something very right-minded about differentiation along structural lines. Yet, I will argue that there is something incomplete – something close-yet-so-still-far – in their account of the types of ‘dialogue games.’

Walton and Krabbe write that the “structure of a dialogue can be thought of as a dialogue game, in the sense that participants take turns making moves, and that they have goals and strategies” (Walton & Krabbe, 1995, p. 67). Of course, as with all analogies, the comparison is necessarily partial and less than full identity. Their phrasing “in the sense” presumably delimits the scope of the analogy. They contend that *only* in these ways is a dialogue like a game; accordingly, it is only in the sense of turn-taking and goal-oriented strategy that games share relevant functional similarities to dialogues. As with all arguments from analogy, there are critical questions one would be well-served to ask when using a known concept to extend our knowledge of a less-known concept. Perhaps most important among the questions: do we, in fact, have a good understanding of the first analogue? If we were to misunderstand games, would we misunderstand argument?

I submit that ‘turn-taking’ and ‘goal-oriented strategy’ does not begin to exhaust the relevant functional similarities between game and argument. This is neither the boundary nor the limit of what we might fruitfully gain from attempting to use games to better understand argument. My point of departure from the others who have variously attempted positing game-infused accounts of argument is this: prior attempts to do so have largely operated with either an impoverished or overly-reductive concept of ‘game.’ This unnecessarily limited view of games thus produces an unnecessarily limited view of argument. Walton and Krabbe appear to have this sort of concept in mind, as they remark that “ultimately, the rules may define a “game” in the sense of mathematical game theory” (1995, p. 70).

If one were not interested in trusting an extemporaneous, anecdotal understanding of games, one might presume that the well-established and boldly named ‘game theory’ might be a reasonable source of guidance into what sort of thing a game is. But this presumption should be revoked on further examination. Roughly stated, ‘game theory’ is a theory of strategic decision-making, especially in terms of economic and mathematical models that originates in part from rational choice theories of action. Its conceptual resources are largely geared towards qualifying and quantifying a given situation; this is so one can calculate which decision most optimally satisfies one’s interests in that moment. Game theory primarily attempts to answer the question: “what is the best choice to make right now?”

Game theory is normative in this way: it presumes that there is a best – or perhaps a ‘least worst’ – choice in any situation; and that

individuals should always want to be finding and taking it. In other words, players of this 'game' of strategic choosing are supposed to be 'rational choice actors' who should be and always are seeking to maximize their interests. This approach also skirts into the realm of the idealized insofar as it involves theoretical modelling of abstracted and optimal decision-making. But unless one were to think that 'playing a game' reduces down or is fully analyzed into an idealized series of strategic decisions, 'game theory' – despite its name – actually tells us very, very little about real 'games' or its players. Game theory is not entirely useless in understanding games – but it is only useful for dealing with a very narrow slice of the broad family of concepts that are relevant to playing, understanding and designing actual games.

Standing in contrast to the narrow, reductive resources of game theory, we are better-served to consider the broad explanatory power and descriptive richness of *ludology* – the academic and critical study of games *qua* games, especially in terms of player experience, systems design and the socio-cultural dynamic of gaming. Ludology is a multi-disciplinary theory of games richly understood. Among other things, it is a discipline that seeks to understand how fun things are fun so that we can better appreciate- and purposefully create fun experiences. It is the ludologists – not the game theorists – who have taken up Wittgenstein's exhortation to "look!" at real games played by real players in real life and only then try to understand the complex network of elements, interactions, affinities and structures that allows games to successfully function as games. I argue that if we were to use this ludology-enriched understanding of games, our understanding of argument will be enriched. The first step, of course, will be trying to grasp what sort of thing a 'game' is.

3. A LUDOLOGICAL PERSPECTIVE

The following definition of 'game' was adapted from Fullerton et al. (2008) and enhanced with a few philosophical considerations of my own making. However, a few disclaimers are in order: 1) in the field of ludology, the definition of 'game' is roughly as settled as 'argument' is in argumentation theory; 2) the following is expressly and explicitly *not* intended to be a set of necessary or jointly-sufficient conditions. There will necessarily be interesting, marginal edge cases. This is unproblematic since the goal of this definition is not to create a rigorous, universal sorting metric; the goal is to say something helpfully insightful about a large family of activities. From a ludological perspective, a game is:

A first-order system of interacting, formal elements (including rules, objects, players, and goals) that gives rise to a second-order, emergent, possibility-space that situates the player(s) in a structured conflict in which they resolve uncertainty in unequal outcomes.

Elsewhere (Yong-Set, 2016), I have gone into greater detail about this definition of 'game' and its myriad of interesting components. For our purposes here, however, I will be focusing on unpacking and extending the aspect of 'structured conflict' to think in new ways about argument.

To make this case, I will import two ludological concepts. First is the 'lusory attitude.' Bernard Suits remarks that:

There has to be an explanation of that curious state of affairs wherein one adopts rules which require one to employ worse rather than better means for reaching an end [Suits, 1978, p. 38].

Stated briefly and succinctly, the lusory attitude is characterized by a player's willing adoption of constraints and inefficiencies for the sake of accomplishing some goal. Consider the example of golf for an illustration. In this game, the task goal is to get a small white ball into a small hole far in the distance. However, if this were the goal, why would one ever agree to follow a set of rules that prevents the use of hands to do so? The rules require that one eschews perhaps the most expedient method – using one's hands – and instead one must hit the ball with a stick to launch it through the air or over the ground. Why would anyone willingly desire to do things in this inexpedient, inefficient – and some might say – 'irrational' way?

The short answer is that, in a game, players willingly subject themselves to limitations and constraints that make things harder to do precisely because there is a certain kind of meaning, significance and joy in doing things the 'hard way' and not in some other way. When we choose to play together with others, all the players enthusiastically agree to be bound by these same rules of 'meaningful inefficiency.' This peculiar collaborative assent among players is the lusory attitude. Consonant with audience-centric, rhetorical approaches to argumentation, one can adopt a 'player-centric approach to game-design' (Fullerton et. al, 2008). On this view, 'players' are integral, formal elements of an interactive system that must also be understood to understand the emergent phenomena of gameplay.

The second ludology-inspired concept needed is what I will refer to as the 'degrees of opposition.' I take 'opposition' in its broadest of senses; something counts as 'opposition' if it makes it harder rather than easier to accomplish one's goal in some way. Opposition runs

opposite to one's intentions. Opposition can take many forms and come in different degrees. I distinguish between three degrees of opposition under the following stipulated headings: 1) conflict; 2) competition; and 3) confrontation.

The lowest degree of opposition I call 'conflict.' Using golf again as an illustrative example: 1) the player is 'in conflict' with the rules of the system. The stipulation that none can use their hands on the ball certainly makes it harder to accomplish the goal. These procedural rules – those that govern what counts as a move, how one transitions to different phases of gameplay, or how players are to resolve the outcome of interactions – all count as a form of resistance, friction, or constraint on the player. They conflict with or oppose the main goal in a minimal, but important sense. 2) Up one degree is 'competition' – when there is not enough to go around, scarcity frustrates the achievement of one's goals as well. This is the same sense of 'competition' we deploy when speaking of animals in the wild 'competing' over resources. Of course, animals – like players – can compete for a limited and scarce supply of resources without ever directly interacting with or fighting one another. Competition – even when indirect – is still a form of constraint that runs opposite to one's intentions. 3) The highest degree of opposition is 'confrontation.' In highly adversarial contexts – such as in a boxing ring or a criminal court room – one might encounter opposition in the form of another player's direct efforts to impede one's plans. In a confrontation, one player can take actions that directly interfere with the progress and aims of other players. One can manually and purposefully frustrate the other player's ambitions.

Bringing together the lusory attitude and degrees of opposition illuminates an interesting second-order character of games: all games fundamentally have a minimal degree of both collaboration and opposition. Even the most solitary or co-operative games with no adversarial players have a degree of opposition – there is minimally a system of rules and constraints that make achieving goals harder. Even the most adversarial, zero-sum games have a degree of collaboration – even if only in the mutual agreement to be the kind of opposition that the other players want, need and expect to make the victory meaningful. In games, players willingly agree to being opposed in some but not all ways. The results produced by a game system are meaningful only because they came about through an interactive process shaped by constraining, structured procedures that were willingly embraced by the participants. How a lack of explicit assent by tacit participants figures into the quality of a game's results is an important topic for another occasion.

4. THE SHAPE OF ARGUMENT

I submit that this ludological insight can nicely illuminate the murky intuition many have in regards to the apparent tension between challenging others' views and helping them in the process of argumentation. By leveraging ludology's resources, we can sketch a framework for analyzing 'game-types' in terms of the different shapes that the 'player-opposition' relation can take. In turn, I suggest that we can deploy a similar analysis to different argument-types. To demonstrate what this framework looks like, I will be using golf to illustrate three different game-shapes. In this analysis, there are three main types of elements: Player; System; and Environment, represented by P, S and E respectively.

The main opposition of golf is a conflict between the player and a combination of the rules and environment. This would be represented as: $P \text{ v } (S+E)$. As golf is a single-player game, it is straightforward how P is deployed here. However, it should be well-noted that it is not possible to bracket out 'the players' and their real attributes when analyzing games; there is no such thing as a game without players. The System, S, refers to the set of form-giving rules that define what legal procedures are and what constitutes an 'item' or status in the game among other things – in this case, what counts as a 'stroke,' what is a penalty, how one resolves penalties, what an 'albatross' is, what a 'round' is, how victory is determined, and so on. The Environment, E, of golf includes the course on which it is played, the wind, rain, sand and crocodiles that can hinder one's ambitions of getting that ball in that hole. Of course, some elements of the environment can also assist in achieving the goal – but I would argue that they are primarily oppositional. One may also argue that perhaps a digital crowd, a viewing audience or a heckling gallery might also count as an environmental element that can make things harder or easier. Nevertheless, the *primary* shape of the player-opposition relation that characterizes golf is: $P \text{ v } (S+E)$.

Those familiar with golf who watch it on TV may remark that there are many people playing golf at the same time – the leaderboards have dozens of players participating in the tournament against one another. So in what way is golf a single-player game? Simply put, golf is a game of many shapes. Many games can be played in a variety of different 'modes' while nevertheless still being 'the same game' for some intents and purposes. Of course, what differentiates the different game modes is, among other things, the differences among the player-opposition shapes. In tournaments or standard group play, the players are engaged in the second degree of opposition – they are participating in an indirect competition. In this form of the game, the primary player-opposition shape would be: $(\text{Player 1 v } S+E) \text{ v } (\text{Player 2 v } S+E)$. Not only

are the players each in a parallel conflict with the system and environment to get the ball in the hole they are *also* competing against other players by trying to get their ball in the hole in the fewest strokes possible. Since only one player can have the status of ‘having taken the fewest strokes,’ there is a relevant scarcity in what the players need to obtain the win condition.

It is less well known as a format, but there are also a few game modes for ‘team golf.’ The Ryder Cup – a biennial televised competition between North American and European teams of professional golfers – is among the better-known instances of this. For the most part, the competition involves a series of individual (Player 1 v S+E) v (Player 2 v S+E) matches between members of the opposing teams; and the final outcome of the Cup is determined based on which team wins more of those matches. However, some of the matches take place in a ‘foursome’ format. In this case, two players from each team join forces against two players from the other team. The team players take turns and alternate striking their team’s ball, and they can consult and talk with one another during the game. This is an instance of a game with a primary opposition-shape of: $[P1 \text{ v } (S+E)] + [P2 \text{ v } (S+E)] \text{ v } \{[P3 \text{ v } (S+E)] + [P4 \text{ v } (S+E)]\}$. Interestingly, foursome team golf is at once a single player conflict, a team-based collaboration, and an indirect competition. Arguably, with the right players and the right kinds of beverages, team golf can also become directly confrontational as well. In the above example, there are many elements that stand in a variety of relations to one another. Some are oppositional while others are collaborative. Similar to identifying the main operator in logical sentences, so too is it sometimes important to identify the main form of opposition in games. This short exposition canvassed only three possible player-opposition shapes; but these are in no way exhaustive or representative of the myriad of possible and fascinating player-opposition shapes one can design into games.

With some sense of how this analysis works for seeing features of literal game, I now turn to so-called ‘argument-games.’ In the Wittgensteinian spirit, if we can see how games work, we will see something about how arguments work as well. As previously stated, I concur with Walton and Krabbe in believing that it is fruitful to differentiate between types of dialogue – or types of ‘argument-games’ – by their structural and form-giving features. However, the details of my burgeoning framework of structural differentiation are guided by the integration of thick concepts taken from ludology rather than the thin concepts of game theory. To show the plausibility of this framework, I shall recast Walton’s dialogue types in this new light. This is laid out in Table 1 below:

Argument Type	Primary Game/ System Goal	Primary Shape of Player-Opposition
Inquiry	Discover 'truth' based on cumulative proof	$(P+P) \vee S$
Negotiation	Reach mutually advantageous agreement	$P \vee P$
(Rational) Persuasion	Instill / change belief in someone's mind	$(P+S) \vee (P+S)$
Quarrel	Vent emotions, inflict damage	$P \vee P$

Table 1 – Player-opposition shapes in goal-oriented argument

Inquiry can be construed as a co-operative game in which players work together to overcome an opposing system of constraints – namely, the standards of good inference and epistemic adequacy that hinder reaching just any conclusion – a $(P+P) \vee S$ game. The products of inquiry are meaningful candidates for being 'knowledge' precisely *because* they have presumably overcome certain forms of systematic opposition that are designed to exclude outcomes that do not reach certain standards of reliability, acceptability or epistemic warrant. A negotiation is a peculiar 'co-oppositional' game of imperfect information in which the roles and relations are fluid and fluctuate between helping and harming the other's goals – a $(P \vee + S)$ game.

Rational persuasion is a special kind of Player verses Player game in which the system of good inference rules acts as a perspicuous ally to both sides – $[(P+S) \vee (P+S)]$. If both participants are committed to playing this game and not some other, then both are willing to embrace the constraints placed upon them by the rules of good reasoning and epistemic conduct. This enables the possibility of one party making a 'legitimate' and binding move on the other's commitments. By specifying the range of acceptable argumentative moves and effectively codifying them into a rule-bounded system of inadmissible conduct and admissible procedures, the process of argumentation is crafted in such a way that we would deem its products meaningful and 'rationally persuasive.' Of course, an argument – like a game:

is not everywhere bounded by rules; but no more are there any rules for how high one may throw the ball in tennis, or how hard, yet tennis is a game for all that, and has rules too (Wittgenstein, 1953, S.68).

By contrast, the quarrel has no central role for inference-centred systemic constraints; it is a Player vs. Player (PvP) game in which the main goal is different from persuasion. It would not be fully accurate to say that ‘anything goes’ in a quarrel – but when compared to rational persuasion, far less is ruled procedurally inadmissible or inappropriate in crafting the process. In games, as in argument, players willingly agree to being opposed in limited but not fully-specified ways. What ways those should be depends on the goal of the argument-game mode.

There are two remarks that warrant mentioning. Firstly, above I canvass only four types of argument in Table 1. But I do not mean to suggest that these are exhaustive of all types of arguments nor that there are tidy and discrete boundary lines between these categories of analysis in practice. Walton identifies more types of dialogues than I have argument-types. These four categories likely account for a large portion of real-life argumentative dialogues, but this does not mean that all arguments fit into one of these four categories or that argumentation theorists should not be interested in those other forms. Secondly, at this preliminary phase of construction, I have not yet decided how or whether to represent the argumentation counterpart of E in this analysis. Perhaps E could be the distantly-engaged, indirectly-interacting, non-player-character audience; for example – a public speech or a newspaper editorial might be an example of a (P v E) argument-game. Alternatively, E could be the network of socio-economic forces and the dynamics of power relations that suffuse the social milieu in which real-arguments are invariably situated. Perhaps the E might stand for the shared cognitive environment of the sort expounded by Tindale (1999); it could be a backdrop of shared commitments, histories and beliefs that at times lends assistance to one’s aims while at others constitutes a status quo to be overcome. It almost seems to go without saying that these ‘environmental pressures’ are always operating in the background; so like a common term in math, it may divide out of the analysis without issue. But surely this is not correct for we have obtained plentiful evidence that reminds us that even when things go out of sight and out of mind, they do not always go out of effect.

5. CONCLUSION

I have argued that there is good reason to believe that our understanding of argument can be enhanced through an understanding of games. However, previous attempts at exploring this approach have been limited by an impoverished understanding of ‘games.’ By exchanging thin, reductive and extemporaneous concepts of ‘game’ with a thick, multi-faceted and enriched concept of ‘game’ developed by

ludologists, the true potential of such an approach will be better realized. One product of a ludological approach to argument is a conceptual framework that can help us analyze the structure of the opposition-collaboration relation among the participants and the other elements of the system. This in turn can help us better understand the appropriate forms of opposition an antagonist's dissent should take to accomplish the different kinds of goals embraced by arguers when they collaboratively assent to participate in different types of argument-games.

There are many ways to characterize, describe and analyze arguments – some of which are more useful for some purposes than others. Gilbert suggests that heuristic dialogues – such as inquiry – are about finding and making discoveries (2014, p. 40). Whether the dialogue is geared towards discovering propositions worthy of being called 'knowledge' or uncovering previously unknown motives and passions that support deeply held positions, heuristic dialogues involve finding out new information and new thoughts to think. Walton similarly notes that there are eristic dialogues – such as quarrels – in which the primary motif is navigating interpersonal strife or vanquishing an opposing interlocutor's reputation (1995, p. 78). At times, one could be forgiven for inferring that 'heuristic' and 'eristic' are two opposing ends of a spectrum that characterize the degree of adversariality in argumentative dialogue types. However, I would suggest as a final thought that, like 'opposition' and 'collaboration,' the relationship between 'strife' and 'discovery' may not be so simple. In games as in argument, the relationship may be more complex than diametric opposition or mutual-exclusion – and that relationship may come in a diversity of forms and shapes we would be well-served to explore.

When we argue, we collaboratively assent to be opposed by the rules and persons that make our results meaningful. If we keep this in mind, then perhaps we can better understand how we can help others by being the kind of dissenting opposition they need.

ACKNOWLEDGEMENTS: University of Windsor and CRRAR for much aid.

REFERENCES

- Cohen, D. H. (1995). Argument is war . . . and war is hell: Philosophy, Education and metaphors for argumentation. *Informal Logic*, 17(2), 177-188.

- Fullerton, T. & Swain, C. & Hoffman, S. S. (2008). *Game design workshop: A playcentric approach to creating innovative games* (2nd ed.). Burlington, MA: Elsevier Inc.
- Gilbert, M. A. (2014). *Arguing with people*. Toronto, ON: Broadview Press.
- Johnson, R. H. & Blair, J. A. (1987). Argumentation as dialectical. *Argumentation*, 1(1), 41-56
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago and London: The University of Chicago Press.
- Ritchie, D. (2003). ARGUMENT IS WAR – Or is it a game of chess?. *Metaphor and Symbol*, 18(2), 125-146.
- Suits, B. (1978). *The grasshopper: Games, life and utopia*. Toronto, ON: University of Toronto Press.
- Tindale, C. W. (1999). *Acts of arguing: A rhetorical model of argument*. Albany, NY: State University of New York Press.
- Walton, D. N. & Krabbe, E. C. (1995). *Commitment in dialogue: Basic concepts of interpersonal reasoning*. New York: State University of New York Press.
- Wittgenstein, L. (2009). *Philosophical investigations* (4th ed). (G. E. M. Anscombe & P. M. S. Hacker & J. Schulte, Trans.), (P. M. S. Hacker & Joachim Schulte, Eds.). Oxford: Wiley-Blackwell. Original work published 1953.
- Yong-Set, M. A. (2016). A ludological perspective on argument. *OSSA Conference Archive*. Retrieved from <https://scholar.uwindsor.ca/ossaarchive/OSSA11/papersandcommentaries/98/>

Towards the justification of warrant in a legal hard case: How to justify a legal rule by the backing

SHIYANG YU

Philosophy College, Nankai University, Tianjin, China
markyu0828@163.com

XI CHEN

Law School, Shenzhen University, Shenzhen, China
chenxi8601@126.com

Though Toulmin model is a heated topic, it is usually studied in the general sense, which oversimplifies some valuable issues, e.g., the process of justifying warrant by its backing. This paper, situated in the legal field, makes a start to this issue. This type of justification can be divided into two sub-justifications, justifying the existence and the interpretation of the legal rule, respectively. Focusing on the first sub-justification, we present two types of argumentation structures based on two main positions in legal philosophy.

KEYWORDS: backing, legal justification, legal hard cases, Toulmin model, warrant, legal philosophy

1. INTRODUCTION

Toulmin model has been an important issue in argumentation theory since its born. Its significance comes from its functional perspective, in which an argument is explained as constituted of six layouts with different functions, rather than simply by 'premise' and 'conclusion'. However, some authors doubt whether Toulmin has adequately discussed about his model. For instance, Goodnight points out, an additional inference should be added into Toulmin model to certify the choice of backing for a warrant. This idea somehow comes from Toulmin's 'field-dependence'-concept, which can be captured by "allowing for different backings of [the] warrant" (Prakken, 2005, p. 318). Goodnight's inference is useful, especially when some alternative backings have been found, and we need to show "an argument [is] certified by an appropriate, proper, or correct choice of backing" (Goodnight, 1993, p. 41).

Nevertheless, we disagree on Goodnight’s assertion that such inference is a “move from warrant to backing” (ibid.). We draw three figures to illustrate our disagreement. While Fig. 1 exhibits that people could get different backings from one (assumed) warrant, Fig. 2 shows the warrant can be justified by a backing, and Fig. 3 presents the backing 2, comparing with other alternatives, constitutes the best choice. Arrows in these figures also possess different meanings. In Fig. 1, each arrow refers to an *inference* from the warrant to a backing, while in both Fig. 2 and Fig. 3, each arrow represents a *justification* of the warrant by a backing. (But the arrows in Fig. 3 are less important, they only indicate that those backings in the dotted box are qualified candidates for electing the best.)

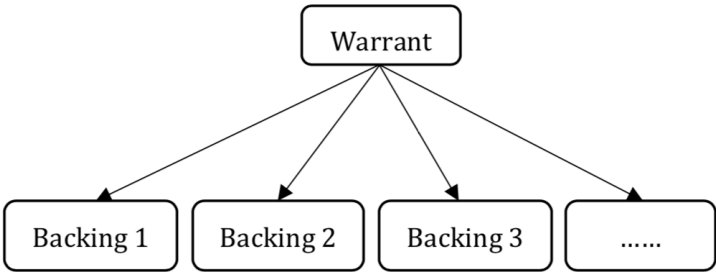


Figure 1 – Inferences from the warrant to backings

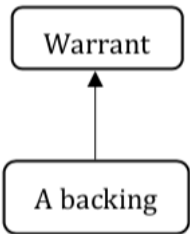


Figure 2 – Justification of warrant by a backing

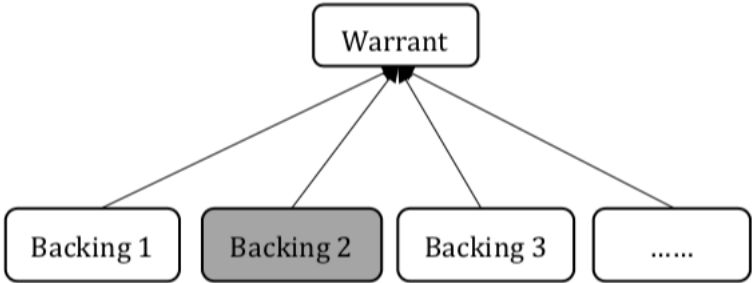


Figure 3 – An election of backings for best justifying the warrant

We use this set of figures to explain that, to justify our choice of the backing concerns more with Fig. 2 and Fig. 3 where the arrows go from the warrant to backings (rather than, like Goodnight suggests, going from the warrant to backings). Because after the proponent has asserted such a choice of backing, he implies, and needs to justify, the backing is a support to the warrant.

In this essay, we focus on Fig. 2. Thus, our intention is to show how to justify the warrant by a backing, or say, how to justify the supportive relation between backing and warrant. We choose legal field to exemplify such justification—since judicial process is parallel to the rational process set out by arguments (see Toulmin, 2003, p. 15). Moreover, for throwing more light on legal justification, we choice hard cases, rather than easy ones. Thus, our purpose in this essay is to present the justification of warrant by backing in the legal justification of hard cases (LJOHC).

We first introduce Toulmin model and its layouts (Sect. 2), then present the counterpart of such justification in legal field and divide it into two sub-justifications (Sect. 3). We conclude in Sect. 4.

2. TOULMIN MODEL AND ITS LAYOUTS

2.1 Toulmin's original model

Toulmin distinguishes six layouts in his model, which, respectively, are data, claim, warrant, backing, qualifier and rebuttal. The criterion to distinguish them is the function in arguments. We illustrate their functions by his famous example,

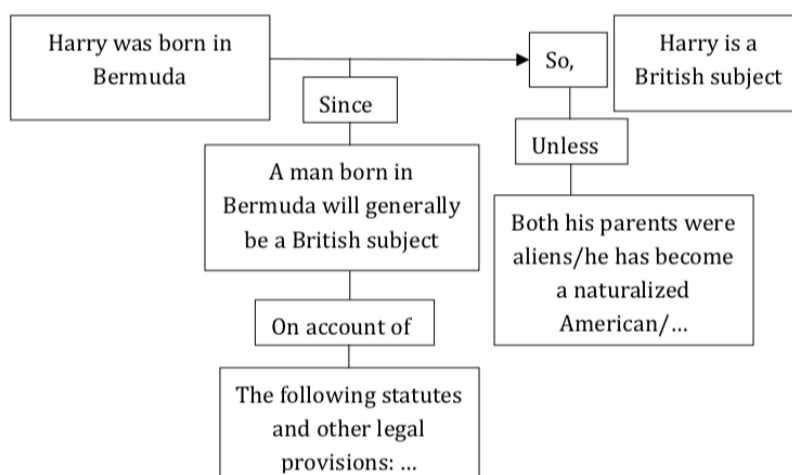


Figure 4 – Toulmin's "Harry is a British subject"—example

In this example, claim (C) is “Harry is a British subject,” “whose merits we are seeking to establish” (Toulmin, 2003, p. 90). Data (D) is “Harry was born in Bermuda,” which are “the facts we appeal to as a foundation for the claim” (ibid.).

Warrant (W) is “[a] general, hypothetical statement” that can “authorize the sort of step to which our specific argument commits us,” which is “[a] man born in Bermuda will generally be a British subject” in this case. It gets authority and currency from the backing (B)—“[t]he following statutes and other legal provisions: ...” in this example (see ibid., 96f.). Meantime, rebuttal (R) reveals some “exceptional conditions [...] might be capable of defeating or rebutting the warranted conclusion,” which can be either “[b]oth his parents were aliens” or by “[h]e has become a naturalized American” in this instance. Thus, “the strength conferred by the warrant on this step” is not absolute, and represented by qualifier—“presumably” in his example.

Based on such distinctions, components of an argument are reassembled in a functional view. In the next section, we introduce a revised model to Toulmin’s original model by Yu & Zenker (2019).

2.2 A revised understanding of Toulmin model

Yu & Zenker (2019) transform Fig. 5—Toulmin’s original model—into their revised Fig. 6.

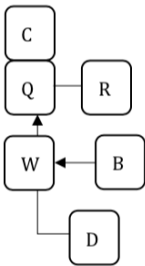


Figure 5 – Original Toulmin model

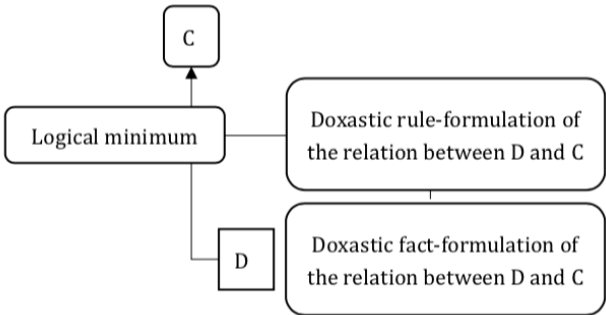


Figure 6 – Revised Toulmin model

The revised model makes three main changes to the original: Firstly, they consider that the essence of 'W' refers to a kind of substantial relation between 'D' and 'C', and the presentation of 'W' is a doxastic rule-formulation of the relation between 'D' and 'C'. Secondly, parallel to considering 'W' as a type of doxastic rule-formulation, they treat 'B' as a type of doxastic fact-formulation of the relation between 'D' and 'C'. It represents how people know, or what has been implied by the doxastic rule-formulation of the relation between 'D' and 'C'. Thirdly, 'Q' is omitted, since in their revised model, 'Q' has been accommodated into 'C', jointly being regarded as a new claim. In addition, 'R' is neither explicit. Because, in their view, 'R' points out the weaknesses not only of the 'W', but also of other components. For clarity, they omit 'R'-mark and use critical questions (CQs) to reveal these weaknesses.

So to speak, such a revised version is not only to reinterpret elements of Toulmin model, but also intends to expand this model. Our following discussion starts from their revised version.

3. THE JUSTIFICATIONS OF THE EXISTENCE OF RULE IN LJOHC

In this section, we divide the justification of warrant by backing in LJOHC into two sub-justifications, one for justifying the existence of a rule, the other for justifying the interpretation of the rule (3.1). Focusing on the former, we introduce another type of backing (3.2). However, such backing still needs further backings, which are decided by legal positions. We introduce two main positions in legal philosophy, and construct justifications of the rule under each position (3.3). Finally, taking natural lawyers' position, we improve their justifications (3.4).

3.1 Two main sub-justifications of warrant from backing in LJOHC

For presenting sub-justifications of 'W' by 'B' in LJOHC, we firstly have to identify the starting- and ending-points of the whole justification. We apply the revised Toulmin model to analyze a legal case: Mary has stolen a car, so the judge intends to claim a two years-sentence by law (please see Fig. 7).

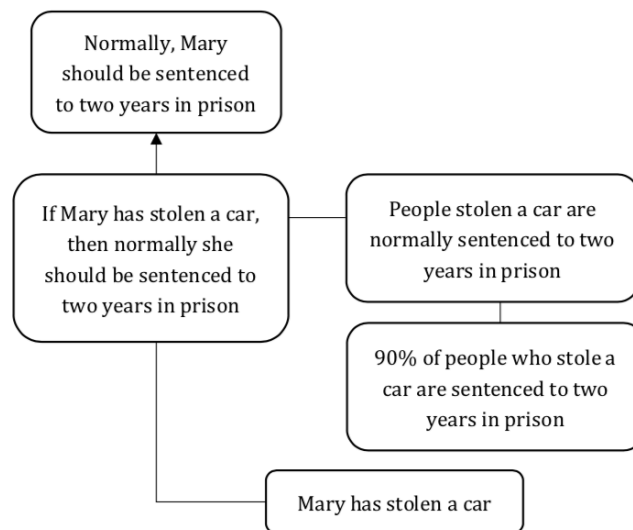


Figure 7 – Applying the revised Toulmin model to Mary-example

The ‘W’, in this example, is “People stolen a car are normally sentenced to two years in prison.” Although it seems like a legal rule, some people may consider it as a legal interpretation. This view is correct, especially in legal hard cases, since in which legal rules need to be interpreted before applied. Thus, we consider ‘W’ as legal interpretation in LJOHC.

Following Yu & Zenker (2019), the ‘B’ should be “90% of people who stole a car are sentenced to two years in prison,” which offers factual basis for the ‘W’. Thus, the justification of ‘W’ by ‘B’ in LJOHC is to justify the legal interpretation of legal rule by the factual basis of the interpretation.

Nevertheless, two problems need to be solved. Firstly, we did not see legal rule as a component yet, which is essential to legal justification. Secondly, we need to explain what can be called as a factual basis of legal interpretation of a rule.

For solving both problems, we need to consider the rule as an implicit component in the justification, which can be inserted between ‘W’ and ‘B’. Then, not only the legal rule becomes a component (first problem solved), but also the factual basis of legal interpretation of rule turns to be, more directly related with, the basis of rule (second problem solved). Thus, the whole justification can be divided into two parts. The first is to justify the legal interpretation of rule, advancing from the rule to ‘W’. The second is to justify the rule itself, advancing from ‘B’ to the rule. However, we still face a further problem, i.e., what is, or can be called as, the factual basis of a rule. Although “90% of people who stole a car are sentenced to two years in prison” is a fact, it is more like a social

(statistical) fact, describing a social, rather than legal, affair. Alas, situating in the legal field, it can not satisfy those people who maintain the factual basis should have been more legal. Further, it seems only to present the outcome of a rule, rather than to justify the validity of a rule. Thus, it is not able to be a *basis* of the rule. We solve this problem in the next section.

3.2 *An additional type of backing*

To reform the factual basis more legally, we need 'B' to express legal facts. Shapiro (2011, p. 25), a famous legal philosopher, defines legal fact as "a fact [that is] about either the existence or the content of a particular legal system," for instance, "the Bulgaria has a legal system" or "the law in California prohibits driving in excess of 65 miles per hour."

Shapiro claims that "legal fact [...] cannot be ultimate," which needs to be made "reference to other facts" about its "existence" (ibid., 26). For instance, although "authority-conferring laws" could "confer authority on officials," the former could also be called into question. Then "we might trace [...] back to the fundamental rules of the legal system" which, in United States, would in turn get help from "constitutional law" (ibid., 26). Since we could find no other legal facts on which constitutional law relies, "the Constitution is fundamental law" (ibid., 27).

Nevertheless, based on "an obvious query: what happens when we run out of legal facts upon which to rely?" (ibid.), we may ask what does the Constitution rely on? To answer this question, Shapiro borrows a pair of answers from two groups, i.e., legal positivists and natural lawyers. Their main difference is decided by the type of facts that are used to support the fundamental law. While "natural lawyers [...] hold that legal facts are ultimately [and simultaneously] determined by moral *and* social facts," "legal positivists [maintain] that all legal facts are ultimately determined by social facts alone" (ibid., 27; *his italics*; also see Patterson, 2010). Indeed, such "disagreement [...] concerns the *necessary properties of law* and, therefore, *the nature of law*" (ibid., 27f.; *his italics*). "[T]he positivist treats the law like custom," but "natural lawyers [...] believe that the nature of law is similar in this regard to the nature of political morality" (ibid., 28).

So, although we could consider the 'backing' as some legal fact, it always needs further backing(s) outside law. In other words, if we want the justification by backing to be more revealing, to consider the backing fundamentally as legal fact is not enough. To think of the justification more essentially, we introduce two legal positions with more detail in the next section.

3.3 Two legal positions and three types of facts

First of all, since many think the natural laws are self-evident, the main difference between legal positivists and natural lawyers embodies in their views on how to justify the legal validity of positive laws. Thus, all justifications involved in the following are of positive laws.

As “one of the two great traditions in legal philosophy,” legal positivism holds “two central beliefs: first, that what counts as law in any particular society is *fundamentally* a matter of social fact or convention (‘the social thesis’); second, that there is no necessary connection between law and morality (‘the separability thesis’)” (Patterson, 2010, p. 228; *italics added*). In other words, the legality (or legal validity) of law comes from its institutional nature, having nothing to do with morality.¹ By distinguishing social and moral facts, in legal positivists’ view, the legal validity is justified merely by social facts.

By contrast, natural lawyers, no matter traditional or modern ones, all deny the separability thesis: they believe morality and law are interrelated (see Patterson, 2010, pp. 211-226). For instance, such belief is well reflected in a famous claim—“an unjust law is no law at all.” However, this statement seems self-contradictory in saying “an apparently valid law is ‘not law’” (Patterson, 2010, p. 214). “A [...] reasonable interpretation [...] is that unjust laws are not laws ‘in the fullest sense’” (ibid.; *italics added*). Namely, “it does not carry the same moral force or offer the same reasons for action as laws consistent with ‘higher law’” (ibid.). Thus, both social and moral facts are necessary for justifying the legality of law from natural lawyer’s view.

Since two legal positions have different necessary components for justifying legal validity in mind, we could build two types of argumentation structures (see Fig. 8 and Fig. 9).

¹ Nevertheless, there are two main approaches of positivism, one is “restrictive” construal, e.g., Joseph Raz, “hold[ing] that it can never be a criterion of legal validity that a norm possess moral value” (Patterson, 2010, p. 230; also see Raz, 1979, pp. 37-52; Raz, 1985, pp. 311-20), the other, which is named as “inclusive” construal (“incorporationism” or “inclusive legal positivism”), “only commit[s] to two weaker claims [...]” (Patterson, 2010, p. 230; also see Hart, 1994). Discussions can be even more complicated, but in the sense that all legal positivists consider social conventions as the fundamental factor of deciding a norm legal, we draw Fig. 9.

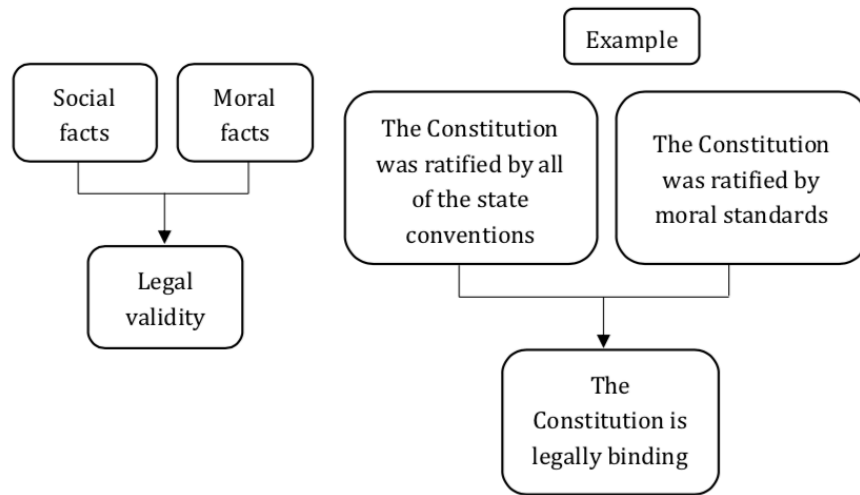


Figure 8—Coordinative justification of legal validity by natural lawyers²

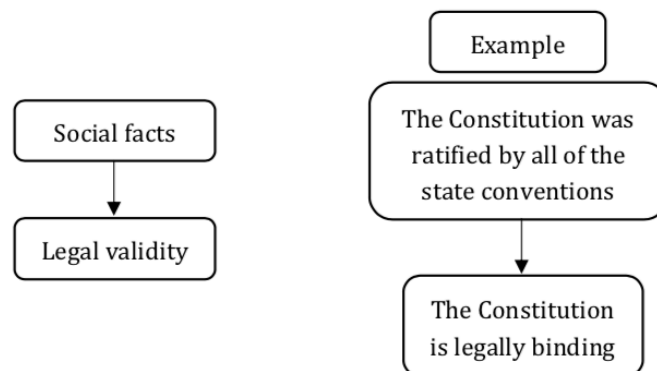


Figure 9—Justificatory of legal authority by legal positivists

We see the justification of legal validity is made of either social facts or social and moral facts. Thus, we need to identify and distinguish three types of facts more clearly. We begin with social facts. Like many others,

² The justification of legal authority by moral fact can be complex argumentation. As we only show a single argumentation, “moral fact” is singular. Pragmadiialecticians distinguish single argumentation from complex argumentation by the number of reasons, saying that “a single argumentation [...] consist[s] of one explicit ‘reason’ [...], [while] ‘complex’ argumentation consist[s] of more reasons” (van Eemeren et al., 2014, p. 23; also see van Eemeren & Grootendorst, 1992).

Durkheim defined social facts as “... consist[ing] of manners of acting, thinking and feeling external to the individual, which are invested with a coercive power by virtue of which they exercise control over him” (Durkheim, 1982). Following this definition, we unfortunately find both legal and moral facts can be considered as social facts. For solving this problem, in this essay, we assume social facts as the facts other than legal and moral facts. We thus need to identify the scopes of legal and moral facts, respectively.

For moral facts, we should not bring all facts that are related with morality into the set of moral facts, since we only care about those which could directly support legal validity. To reach such a limited description, we define moral facts simply as a kind of fact stating something about the consistency between law and moral standards, e.g., a fact stating whether the law is consistent with moral standards, or stating the degree of their consistency. Moreover, we consider statements like “according to moral authority, the Constitution is (morally) good” as moral evaluations, rather than (moral) facts.³

To identify the scope of legal facts, let us see what revisions can be done to the definition—“a fact about either the existence or the content of a particular legal system” (Shapiro, 2011, p. 25). Firstly, we would like to add those statements claiming legal validity of a rule or system, for instance, both conclusions in Fig. 8 and Fig. 9. Secondly, although Shapiro maintains that the authoritative status of the person in question is a legal fact, we would like to preclude such kind of facts which only contain the facts of particular cases. Because they look more like social facts. By contrast, we agree with Shapiro that a fact contains the content of a rule is a legal fact, for instance, “the law in California prohibits driving in excess of 65 miles per hour.”

Based on above considerations, we revise Shapiro’s definition of legal fact as that *a general fact directly relates with the existence, content or validity of a legal system*. “General” is used for precluding those facts only stating any specific legal cases, while “directly” is for limiting the scope of legal facts so as to preclude irrelevant facts. Moreover, similar with Shapiro, we assume the “legal system” as containing all components inside of it.

Although we have offered argumentation structures of two legal positions, they are not final versions. In the next section, we offer a revised version from *natural lawyer’s* view.

³ Mulligan & Correia (2017) holds that “Facts [...] are opposed to theories and to values (cf. Rundle 1993).”

3.4 How to justify a legal rule from natural lawyer's view

Following Shapiro's point of view, we can use legal authority of the Constitution to justify that of a rule (see Fig. 8 and Fig. 9). However, such authority of the Constitution given by social or moral facts is generally but not universally valid. For instance, "the Constitution was ratified by all of the state conventions" does not necessarily mean the *whole* Constitution was consistent with all state conventions. In other words, some other parts of the Constitution may conflict with state conventions, without denying that the Constitution being *generally* ratified. In turn, it is also quite impossible for the ratified parts of Constitution to align with all of the state conventions. Normally, the former only aligns with most conventions, for state conventions may even conflict among themselves.

Thus, we represent the relation between the Constitution with state conventions in Fig. 10. Both areas S1 and S3 are the parts that the Constitution does not coincide with state conventions. They, respectively, refer to the parts that are written or reflected in the Constitution but not in the state conventions and vice versa.

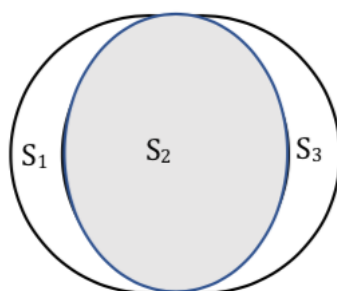


Figure 10 – The possible misalign relation between the Constitution with state conventions

We thus may fail, if we believe the validity of rules can be justified by moral facts on the Constitution. Because such moral facts could only generally justify the validity of the whole legal system rather than every part of it. To improve the strength for justification legal validity of a particular rule, we need another coordinative help from the moral facts directly related with the rule in question (see Fig. 11). We call the branch by appealing to the Constitution as method 1 (M1), the other as method 2 (M2). Nevertheless, the real difference between M1 and M2 is not whether appealing to the Constitution, but whether inside the legal system. When applying M1, there may be some intermediate steps on the way from the Constitution to justify legal validity of Rule A. For instance, the legal validity of Rule A is ratified by Rule B's legality, and so forth, but finally by the Constitution (or from the Constitution and related social

and moral facts of Rule B at the same time). Consequently, we define M2 as a method that directly appeals to social and moral facts. To distinguish M1 and M2 in brief, we call M1 as the method inside of legal system, while M2 as the method outside of legal system.

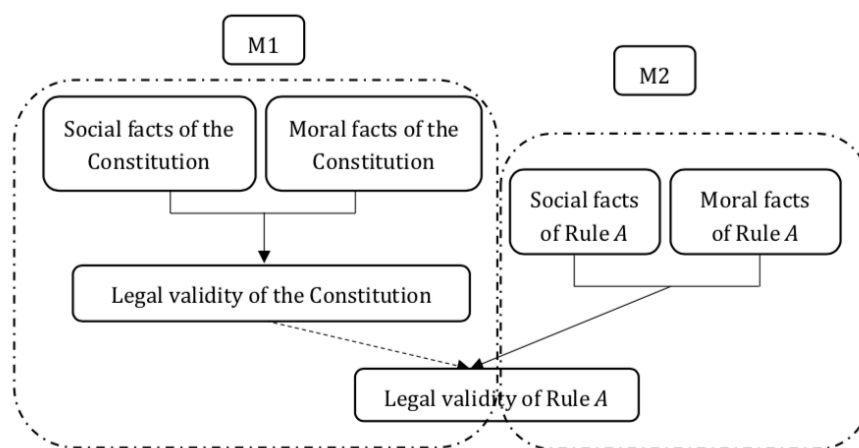


Figure 11 – Coordinative argumentation for justifying legal validity of rule A

Now we are able to answer the question—“what kinds of backing are *legal* backing of justification in hard cases?” Following natural lawyers’ view, both social and moral facts are necessary. But the social fact appealed are not statistical facts, e.g., “90% of people who stole a car are sentenced to two years in prison,” but, for instance, “the parliament has published a relevant legal statue” or “legal officials treat the state conventions as having had the power to ratify the Constitution.” In the sense that these social facts offer legal validity to the rules, we call them validity facts. For the same reason, the moral facts considered as legal backings are also called validity facts. In a word, legal backings can be either social or moral facts, which are collectively called validity backing or validity facts.

4. CONCLUSION

In this essay, for laying the basis for justifying the ‘W’ in LJOHC, we presented how to justify a legal rule by ‘B’. Based on the dichotomous distinction of legal positions, i.e., legal positivism and legal naturalism, we constructed two general argumentation structures, respectively. For a better construction, we also identified three types of facts. Finally, we focused on the justification from natural lawyers’ view, and made some improvements.

REFERENCES

- Durkheim, E. (1982) [1st pub. 1895]. *The Rules of Sociological Method and Selected Texts on Sociology and its Method*. (S. Lukes, Ed., W.D. Halls, Trans.). New York: Free Press.
- Goodnight, G.T. (1993). Legitimation inferences: An additional component for the Toulmin model. *Informal Logic*, 15, 41-52.
- Grennan, W. (1997). *Informal Logic: Issues and Techniques*. Montreal: McGill-Queen's University Press.
- Hart, H. (1994). *The Concept of Law* (Updated ed.). Oxford: Clarendon Press.
- Hitchcock, D. (2006). Good reasoning on the Toulmin model. In D. Hitchcock & B. Verheij (Eds.), *Arguing on the Toulmin Model. New Essays in Argument Analysis and Evaluation* (pp. 203-218). Netherlands: Springer.
- Mulligan, K., & Correia, F. (2017). "Facts," *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), E.N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2017/entries/facts/>.
- Patterson, D. (Ed.) (2010) *A Companion to Philosophy of Law and Legal Theory* (Second ed.). Malden, MA: Blackwell Publishing.
- Prakken, H. (2005). AI & law, logic and argument schemes. *Argumentation*, 19, 303-320.
- Raz, J. (1979). *The Authority of Law: Essays on Law and Morality*. Oxford: Oxford University Press.
- Raz, J. (1985). Authority, Law and Morality. *The Monist* 68:295-324.
- Shapiro, S. (2011). *Legality*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Toulmin, S.E. (2003). *The Uses of Argument* (Updated ed.). Cambridge: Cambridge University Press.
- van Eemeren, F.H., & Grootendorst, R. (1992). *Argumentation, Communication, and Fallacies*. Hillsdale: Lawrence Erlbaum.
- van Eemeren, F.H., Garssen, B., Krabbe, E.C.W., Snoek Henkemans, F., Verheij, B., & Wagemans, J.H.M. (2014). *Handbook of Argumentation Theory*. Dordrecht: Springer Netherlands.
- Yu, S., & Zenker, F. (2019). Schemes, critical questions, and complete argument evaluation. *Under review*.

How Should We Classify Argument Schemes?

DAVID ZAREFSKY
Northwestern University
d-zarefsky@northwestern.edu

Argument schemes serve a purpose akin to validity in formal deduction, by using critical questions to test the soundness of arguments. The reason to classify argument schemes is so that similar types of arguments will raise similar critical questions. Some writers propose large numbers of argument schemes; others propose very few. While describing argumentative reality may require many schemes, critical testing of arguments is best served by few – though not as few as the three van Eemeren and his colleagues propose.

KEYWORDS: analogy, argument scheme, cause, example, form, sign, testimony, validity, warrant

1. INTRODUCTION

Once again at ECA, I find myself taking up a seemingly simple but foundational question that has little direct relationship to the conference theme. This time it is the question, how should we classify argument schemes? My interest in this topic grew out of a practical problem. I was working on a textbook (Zarefsky, 2019) and had to decide how to organize my treatment of non-deductive inference patterns in ordinary argumentation. To deal with this question intelligently, I had to address a number of prior questions along the way.

2. PRELIMINARY QUESTIONS

2.1 What do we mean by an argument scheme?

An argument scheme is a template for arguments of a particular type. It is a pattern, not a specific argument. It can accommodate a variety of arguments on different subjects so long as they fit the pattern.

The defining characteristic of the pattern is the warrant, in Toulmin's (1958) sense. It is responsible for authorizing inferences

from the grounds of an argument to the claim it advances. It is the warrant that establishes, in the given case, that the grounds add to the argument's probative force by counting as support for the claim. So, for example, arguments on different subjects that depend on causal warrants will belong to the same argument scheme, as will all that rely on analogies to establish a claim. There will be as many argument schemes, then, as there are prototypes of warrants.

2.2 What is the purpose of an argument scheme?

Argument schemes are analogous to the concept of validity in formal deduction. Validity, along with the truth of the grounds, is one of the principal determinants of an argument's soundness. Validity refers to correctness of form. It is not itself concerned with an argument's truth, but rather asks: *If* the statement of the grounds is true, must the claim be true? In a valid formal argument, the grounds will entail the claim. For the grounds to be true and the claim false would be a contradiction.

This gives a check on the argument above and beyond the truth of the grounds. Not only is the evidence correct in what it says, but it *counts as* evidence for the claim's being true. This is important even outside the realm of formal deduction, since the claim goes beyond the grounds, taking us from the known to the relatively unknown. The claim cannot be guaranteed even if the grounds are true. It follows with some degree of probability.

It would be most unfortunate, then, if the only predictor we had of the soundness of the claim were the truth of the grounds. When we count on the warrant to license the inference from grounds to claim, we want to know that the inference follows a pattern whose results generally give us confidence.

2.3 How does an argument scheme inspire confidence?

Regardless of the "school" of argumentation theory from which they come, scholars maintain that argument schemes are tested through a series of "critical questions" that, taken together, enable us to know whether the inferential path put forward in an argument is more likely to be right than is its rejection or the search for some alternative path. Satisfactory answers to the critical questions will rule out alternative paths.

The critical questions will vary by the nature of the inference. In the case of analogy, for instance, one of the obvious questions is whether the things being compared are more alike than different. In the case of causal inferences, one of the critical questions is whether there is

a likely alternative cause; another is whether cause has been confused with effect.

Arguments for which the same critical questions are appropriate generally will be instances of the same argument scheme. Any argument properly classified within the scheme will raise the same critical questions.

But this does not take us very far, because there is no standard way in which to frame the critical questions with which to test an argument. They can be framed very broadly or quite narrowly or anything in between. Different writers will pose different numbers of argument schemes, they will define them differently, and they may pose different critical questions. This observation brings us closer to the central concern.

2.4 How many argument schemes are there?

Writers will answer this question differently depending on how narrowly they slice. The next two major sections will consider quite different answers.

3. WALTON'S APPROACH

Douglas Walton imagines there to be lots of argument schemes. In *Argument Schemes for Presumptive Reasoning* (1996), he proposes 25 different moves in a chapter called "The Argumentation Schemes," and in separate chapters he covers varieties of the argument from ignorance, ignoring qualifications, and the argument from consequences. These presumably are additional argument schemes. His total may be larger, and it may be still growing.

Now, so far as I can tell, Walton's primary goal was not to classify for its own sake but to defend the whole category of presumptive inferences against the charge that they are necessarily fallacious. Presumptively reasonable arguments, Walton writes, are "inconclusive and defeasible arguments that nonetheless have a practical function of shifting the burden of proof in a dialogue" (Walton, 1996, p. ix). To illustrate his point, he offers an assemblage of such arguments with brief discussion of each. This is how he derives his argument schemes.

To the degree that his patterns are basically distinct from one another, it makes sense to treat them as separate categories. But to the degree that they are minor variations of the same basic type, not really raising different critical questions from one another, it does not seem necessary to separate them. In fact, proliferating the number of similar argument schemes may make the categories less reliable by reducing

the likelihood that different analysts will assign the same argument to the same scheme.

For example, Walton identifies one scheme as “argument from expert opinion” and another, “argument from position to know.” But in the former scheme, it presumably is one’s expertise that places one in a position to know, thus blurring the distinction. If an alleged expert were not in a position to know something, his or her alleged expertise would not count for much. Conversely, for some purposes (such as eyewitness testimony in court), being in a position to know is what will make one an expert on the specific questions at hand. (I realize that this may be somewhat equivocating the term “expert,” but that is what may happen if there are multiple available schemes in which to place an argument.) In cases like these, we can run the risk of not asking the right critical questions if we are not sure of the argument scheme.

4. VAN EEMEREN, GROOTENDORST, AND SNOECK HENKEMANS’S APPROACH

Another way to classify argument schemes, quite different from the first, is the work of Frans van Eemeren, Rob Grootendorst, and Francisca Snoeck Henkemans. In their textbook, *Argumentation: Analysis, Evaluation, Presentation* (2002), and in other sources, they reduce the number of argument schemes to three: symptomatic argumentation, analogical argumentation, and causal argumentation.

These authors, I should note, use certain terms differently from my usage. What I call the claim, they call a standpoint; and what I refer to as grounds or evidence, they label the argument. We agree, however, in regarding the argument scheme as that which links together grounds and claim (or arguments and standpoint). The argument scheme consists of an inference and the authorization for it.

What van Eemeren and his colleagues call symptomatic argumentation is often understood as *sign*. One characteristic is predictive of another, so the existence of the first will serve as a sign of the (current or future) existence of the second. Predictiveness is the underlying warrant. Similarly, in an analogical argument scheme, the underlying warrant is one of resemblance between the items being compared. And in a causal argument scheme, the underlying warrant is the influence of one factor on another.

These three are widely recognized as different types of argument, each with its own set of critical questions. For example, symptomatic argument is tested by asking, “Aren’t there also other non-Y’s that have the characteristic Z?” and “Aren’t there also other Y’s that do not have the characteristic Z?” (van Eemeren, Grootendorst, & Snoeck Henkemans, 2002, p. 98). Analogical arguments are tested by

asking whether there are significant differences between the items being compared that might outweigh the resemblance (van Eemeren, Grootendorst, & Snoeck Henkemans, 2002, p. 99). And causal arguments are tested by asking whether the alleged cause always leads to the alleged effect (van Eemeren, Grootendorst, & Snoeck Henkemans, 2002, p. 101).

Van Eemeren and his colleagues defend the reduction of argument schemes to a small number by reference to Occam's razor: Other things being equal, simplicity is to be preferred and unnecessary complication avoided. It is easier to remember three categories than 25 or more. And with only three categories, different analysts are far more likely to sort the same argument into the same category. With more arguments needing to satisfy the same critical questions, it is less likely that the critical questions will be fashioned for idiosyncratic arguments rather than the other way around. In short, confining the number of argument schemes can serve as a means of quality control in the evaluation of arguments.

5. ARE THREE ARGUMENT SCHEMES TOO FEW?

What is perhaps unique about the approach of van Eemeren and his colleagues is the belief that the number of argument schemes can be reduced to three. Instinctively it seemed to me that there were more than that. This was the problem I encountered in working on my forthcoming textbook and reflecting in my experiences teaching an undergraduate argumentation course over a period of forty years. Simply put, there were a few argument schemes that seemed to me to fall outside the categories of symptom, analogy, and cause.

5.1 Generalization

To begin with, what about generalizations – argument patterns that relate evidence about the part to claims about the whole, and vice versa? This is a common pattern of reasoning in both its anecdotal and its statistical varieties. The underlying inference is representativeness: what is true of the part is true of the whole, or what is true of the whole is true of the part. This is not the same thing as prediction. It is not that the existence of one thing allows us to predict the existence of another, different thing; it is rather that what we have identified is a representative slice of a larger version of itself. This is a common argument scheme, with different critical questions from those of sign arguments.

5.2 Testimony

Then there are arguments from testimony. To be sure, testimony is a form of evidence. But there also is a form of reasoning based on testimony: p is true because x says so. The underlying inference is one of credibility, that x can be trusted with regard to p . Although fallible, x has credentials and a track record that make x 's word regarding p reliable in the absence of strong reason to the contrary. This again is a common argument scheme with its own critical questions.

5.3 Form

My final example of an argument scheme is a bit more idiosyncratic, and that is the argument from form. The literary and rhetorical critic Kenneth Burke defined form as the arousal and fulfilment of appetites (Burke, 1931/1968, p. 124). Form is a pattern such that if we know the opening elements, we can expect what will come next, and the fulfilment of our expectations gratifies us that we are "in the know" because we figured it out before the arguer told us. This is not the same thing as symptomatic argument. It is not that one thing predicts another; it is that the pattern gratifies one's expectations and that gratification allows us to infer that the pattern is correct.

I have found three subcategories of this pattern: the quasi-mathematical, the quasi-logical, and the narrative. Quasi-mathematical arguments employ what look like mathematical relationships, and this form authorizes inferences that are not really mathematical at all. Consider the property of transitivity: if A is greater than B and B is greater than C , then A is greater than C . But suppose we replace "greater" with "better." Then we get "If A is better than B and B is better than C , then A is better than C ." But "better" is not an objective mathematical notion and is not reducible to quantitative comparisons. Consider the case of American football. If Stanford defeats Notre Dame and Northwestern defeats Stanford, does that mean that Northwestern is better than Notre Dame? What happens the following weekend when Notre Dame beats Northwestern?

Quasi-logical arguments (the term was invented by Perelman and Olbrechts-Tyteca [1958/1969]) look like logical rather than mathematical relationships. Consider the form of a common scientific argument: If a person becomes frustrated, the person will become aggressive; the person becomes aggressive; therefore, he or she is frustrated. Logically this is actually the fallacy of affirming the consequent. One could be aggressive for reasons other than frustration. We use the procedure of control groups and randomized trials, and we ask critical questions about alternate explanations, in order to make

them less plausible. Only then do we say that the form of the argument looks right. The fact that we can rule out alternatives, not the seemingly logical structure, is what allows us to accept a formally invalid inference as a reasonable argument resulting from its logical form.

Lastly, narrative form has its own conventions and policies. The existence of characters, a plot, a conflict, momentum toward resolution, and the denouement are all elements in the pattern. As Walter Fisher (1987) has pointed out, coherence and narrative fidelity become the critical questions by which we determine whether the story embodied in the narrative holds together and makes sense as a story. If it does, we will count those facts as warrants for taking the story seriously and being influenced by it.

6. CONCLUSION

In sum, then, I set out to offer a minimal number of argument schemes and ended up with six: example, based on the warrant of representativeness; analogy, based on the warrant of similarity; sign, based on the warrant of predictiveness; cause, based on the warrant of influence; testimony, based on the warrant of credibility; and form, based on the warrant of expectation.

Are a few argument schemes better than many? It depends on one's purpose. If the goal is to describe argumentative reality which adheres inexactly to patterns, or to show that presumptive inferences are reasonable, more probably is better. It gives the analyst more tools to use and permits more precise description. But if the goal is to learn how to build and test arguments, fewer is probably better. They are easier to remember and to sort, and it is more likely that different analysts will sort them the same way. Having thus sorted them, one should have a clear choice of critical questions to ask. Since this was my goal in writing a textbook, I opted for a small number of argument schemes. But I could not get the number down to three.

As an aside, I hope this brief example of practical problems in textbook writing illustrates that this sort of writing is far from simple. It uncovers basic issues, confusions in usage, and gaps in theory – all in the process of trying to make the subject matter accessible.

REFERENCES

- Burke, K. (1931/1968). *Permanence and change*. Berkeley and Los Angeles: University of California Press.
- Eemeren, F. H. van, Grootendorst, R., & Snoeck Henkemans, F. (2002). *Argumentation: Analysis, evaluation, presentation*. Mahwah, N.J.: Lawrence Erlbaum.

- Fisher, W. (1987). *Human communication as narration: Toward a philosophy of reason, value, and action*. Columbia: University of South Carolina Press.
- Perelman, Ch., & Olbrechts-Tyteca, L. (1958/1969). *The new rhetoric: A treatise on argumentation* (J. Wilkinson & P. Weaver, trans.). Notre Dame, Ind.: University of Notre Dame Press.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Walton, D. N. (1996). *Argumentation schemes for presumptive reasoning*. Mahwah, N.J.: Lawrence Erlbaum.
- Zarefsky, D. (2019). *The practice of argumentation: Effective reasoning in communication*. Cambridge: Cambridge University Press.

Index of Authors

- Atkinson, David, 225
- Chen, Xi, 463
- Franco, Dominic De, 369
- Jackson, Sally, 325
- Johnson, Steven, 3
- Khomenko, Iryna, 23
- Kišiček, Gabrijela, 35
- Kline, Susan L., 153
- Krabbe, Erik C. W., 49
- Laar, Jan Albert van, 49
- Lagewaard, Thirza, 63
- Lanius, David, 77
- Lawrence, John, 97
- Liefgreen, Alice, 385
- Linqiong, Yan, 115
- Looij, Matthijs, 127
- López, Federico E., 143
- Manohar, Uttara, 153
- Marraud, Hubert, 167
- Mazzi, Davide, 181
- Meter, Christian, 311
- Minghui, Xiong, 115
- Mora-Márquez, Ana María, 421
- Olmos, Paula, 195
- Peach, Harmony, 213
- Pease, Alison, 369
- Peijnenburg, Jeanne, 225
- Phillips, Kathryn, 239
- Pilgram, Roosmaryn, 251
- Plug, H. José, 267
- Plumer, Gilbert, 279
- Reed, Chris, 97
- Ritola, Juho, 291
- Santibáñez, Cristián, 305
- Schneider, Alexander, 311
- Schneider, Jodi, 325
- Sedlaczek, Andrea Sabine, 339
- Sleeuw, Stefan, 355
- Snaith, Mark, 369
- Steenhoven, Bertijn van der, 127
- Tešić, Marko, 385
- Toniolo, Alice, 405
- Visser, Jacky, 97
- Walker, Gustavo Fernández, 421
- Walton, Douglas, 97, 405
- Wohlrapp, Harald R., 435
- Yong-Set, Michael A., 449
- Yu, Shiyang, 463
- Zarefsky, David, 477

BACK

