Journal of applied logics (Print) ISSN 2631-9810
Journal of applied logics (Online) ISSN 2631-9829

Journal of Applied Logics

The IfCoLog Journal of Logics and their Applications

Volume 8 • Issue 3 • April 2021

Available online at
www.collegepublications.co.uk/journals/ifcolog/
Free open access

Journal of Applied Logics - IfCoLog Journal of Logics and their Applications

Volume 8, Number 3

April 2021

Disclaimer

Statements of fact and opinion in the articles in Journal of Applied Logics - IfCoLog Journal of Logics and their Applications (JALs-FLAP) are those of the respective authors and contributors and not of the JALs-FLAP. Neither College Publications nor the JALs-FLAP make any representation, express or implied, in respect of the accuracy of the material in this journal and cannot accept any legal responsibility or liability for any errors or omissions that may be made. The reader should make his/her own evaluation as to the appropriateness or otherwise of any experimental technique described.

ISBN 978-1-84890-365-4
ISSN (E) 2631-9829
ISSN (P) 2631-9810

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

http://www.collegepublications.co.uk

Editorial Board

Scope and Submissions

This journal considers submission in all areas of pure and applied logic, including:

- pure logical systems
- proof theory
- constructive logic
- categorical logic
- modal and temporal logic
- model theory
- recursion theory
- type theory
- nominal theory
- nonclassical logics
- nonmonotonic logic
- numerical and uncertainty reasoning
- logic and AI
- foundations of logic programming
- belief change/revision
- systems of knowledge and belief
- logics and semantics of programming
- specification and verification
- agent theory
- databases
- dynamic logic
- quantum logic
- algebraic logic
- logic and cognition
- probabilistic logic
- logic and networks
- neuro-logical systems
- complexity
- argumentation theory
- logic and computation
- logic and language
- logic engineering
- knowledge-based systems
- automated reasoning
- knowledge representation
- logic in hardware and VLSI
- natural language
- concurrent computation
- planning

This journal will also consider papers on the application of logic in other subject areas: philosophy, cognitive science, physics etc. provided they have some formal content.

Submissions should be sent to Jane Spurr (jane@janespurr.net) as a pdf file, preferably compiled in LaTeX using the IFCoLog class file.

Contents

ARTICLES

On the Computational Complexity of Model Checking for Dynamic Epistemic Logic with S5 Models

Ronald de Haan
Institute for Logic, Language and Computation (ILLC), University of Amsterdam
`me@ronalddehaan.eu`

Iris van de Pol*
Institute for Logic, Language and Computation (ILLC), University of Amsterdam
`i.p.a.vandepol@uva.nl`

Abstract

Dynamic epistemic logic (DEL) is a logical framework for representing and reasoning about knowledge change for multiple agents. An important computational task in this framework is the model checking problem, which has been shown to be PSPACE-hard even for S5 models and two agents—in the presence of other features, such as multi-pointed models. We answer open questions in the literature about the complexity of this problem in more restricted settings. We provide a detailed complexity analysis of the model checking problem for DEL, where we consider various combinations of restrictions, such as the number of agents, whether the models are single-pointed or multi-pointed, and whether postconditions are allowed in the updates. In particular, we show that the problem is already PSPACE-hard in (1) the case of one agent, multi-pointed S5 models, and no postconditions, and (2) the case of two agents, only single-pointed S5 models, and no postconditions. In addition, we study the setting where only semi-private announcements are allowed as updates. We show that for this case the problem is already PSPACE-hard when restricted to two agents and three propositional variables. The results that we obtain in this paper help outline the exact boundaries of the restricted settings for which the model checking problem for DEL is computationally tractable.

We would like to thank anonymous reviewers for their useful feedback on previous versions of the paper.

*Iris van de Pol was supported by Gravitation Grant 024.001.006 of the Language in Interaction Consortium from the Dutch Research Council (NWO).

1 Introduction

Dynamic epistemic logic (or DEL, for short) is a logical framework for representing and reasoning about knowledge (and belief) change for multiple agents. This framework has applications in philosophy, cognitive science, computer science and artificial intelligence (see, e.g., [7, 10, 13, 18, 26, 31]). For instance, reasoning about information and knowledge change is an important topic for multi-agent and distributed systems [22].

DEL is a very general and expressive framework, but many settings where the framework is used allow strong restrictions. For instance, in the context of reasoning about knowledge, the semantic models for the logic are often restricted to models that contain only equivalence relations (also called S5 models).

For many of the applications of DEL, computational and algorithmic aspects of the framework are highly relevant. It is important to study the complexity of computational problems associated with the logic to determine to what extent it can be used in practical settings, and what algorithmic approaches are best suited to solve these problems. One important computational task is the problem of model checking, where the question is to decide whether a formula is true in a model.

The complexity of the model checking problem for DEL has been a topic of investigation in the literature. For a restricted fragment of DEL, known as public announcement logic [4, 5, 27], the model checking problem is polynomial-time solvable [8, 23]. The problem of DEL model checking, in its general form, has been shown to be PSPACE-complete [2, 16], even in the case of two agents and S5 models. However, these hardness proofs crucially depend on the use of multi-pointed models, and therefore do not apply for the case where the problem is restricted to single-pointed S5 models. This open question was answered with a PSPACE-hardness proof for the restricted case where all models are single-pointed S5 models, but where the number of agents is unbounded [29, 30]. It remained open whether these PSPACE-hardness results extend to more restrictive settings (e.g., only two agents and single-pointed S5 models).

In this paper, we investigate to what extent these PSPACE-hardness results hold for more demanding combinations of restrictions. In other words, we study the exact boundaries between (A) the combinations of restrictions that lead to the model checking problem being polynomial-time solvable, and (B) the combinations of restrictions for which the model checking problem is computationally intractable. Various examples of restrictions have been found that fit in either (A) or (B), but no structural investigation has been done on the exact boundaries between these two areas. Investigating these exact boundaries is useful and relevant, for example, for the development of (implemented) algorithms for the model checking problem for

DEL—we discuss this relevance in more detail in Section 5.

Other related work Various topics related to DEL model checking have been studied in the literature. For (several restricted variants of) a knowledge update framework based on epistemic logic, the computational complexity of the model checking problem has been investigated [6]. Other related work includes implementations of algorithms for DEL model checking [9, 15]. Additionally, research has been done on the complexity of the satisfiability problem for (fragments of) DEL [2, 25].

Results and contributions In this paper we provide a detailed computational complexity analysis of the model checking problem for DEL, restricted to S5 models. We consider various different restricted settings of this problem.

For the case of arbitrary event models, we have the following results.

- We make the following folklore result explicit: that the problem is polynomial-time solvable in the case of a single agent and single-pointed S5 models without postconditions (Proposition 1).

- We show that a similar restriction (single agent and single-pointed S5 models) where postconditions are allowed already leads to Δ_2^{p}-hardness (Theorem 2).

- When multi-pointed event models are allowed, we show that the problem is PSPACE-hard even for the case of a single agent and S5 models without postconditions (Theorem 3).

- For the case where there are two agents, we show that the problem is already PSPACE-hard when restricted to single-pointed S5 models without postconditions and with only three propositional variables (Theorem 4).

An overview of the complexity results for arbitrary event models can be found in Table 1. These results outline the boundaries of the tractable setting of the folklore results pinpointed in Proposition 1—they indicate that relaxing any of the three elements of the condition (i.e., a single agent, single-pointed models, and no postconditions) results in computational hardness.

Additionally, we consider the setting where instead of arbitrary event models, only semi-private announcements can be used—this is a restricted class of event models. In this setting, the problem is known to be PSPACE-hard, when an arbitrary number of agents is allowed (i.e., when the number of agents is part of the problem input) [29, Theorem 4].

- We show that the problem is already PSPACE-hard in the case where there are only two agents and only three propositional variables (Theorem 5).[1]

# agents	single- or multi-pointed	postconditions	complexity	
1	single	no	in P	(Proposition 1)
1	single	yes	Δ_2^{p}-hard	(Theorem 2)
1	multi	no / yes	PSPACE-complete	(Theorem 3)
≥ 2	single / multi	no / yes	PSPACE-complete	(Theorem 4)

Table 1: Complexity results for the model checking problem for DEL with S5 models and S5 event models.

Interpretation of the results The results that we obtain in the paper contribute to our understanding of the computational complexity of the model checking problem for DEL. In particular, our results form a useful step towards a better comprehension of how the various elements of the framework of DEL contribute to the computational costs of the model checking problem. For example, the hardness result of Theorem 4 indicates that introducing a second agent—even when severely restricting several other aspects of the problem—already leads to a problem that in the worst case is as hard as the general, unrestricted problem. This improved insight can—in future work—be used to develop (implemented) algorithms for DEL model checking that work more efficiently in different settings and for different applications. We discuss the relevance and significance of our results in more detail in Section 5.

Roadmap We begin in Section 2 with reviewing basic notions and notation from dynamic epistemic logic and complexity theory. Then, in Section 3, we present the complexity results for the various settings that involve updates with (arbitrary) event models. In Section 4, we present our PSPACE-hardness proof for the setting of semi-private announcements. We discuss the relevance and significance of our results for the computational and algorithmic study of the model checking problem for DEL in Section 5. Finally, we conclude and suggest directions for future research in Section 6.

[1] We would like to point out that Theorem 5 is a stronger result than Theorem 4—Theorem 5 implies the result of Theorem 4. We present it as two separate results because the proof of Theorem 4 acts as a stepping stone for proving Theorem 5—the proof of Theorem 4 is useful for understanding the elaborate proof of Theorem 5.

2 Preliminaries

We briefly review some basic notions from dynamic epistemic logic and complexity theory that are required for the complexity results that we present in this paper.

2.1 Dynamic Epistemic Logic

We begin by reviewing the syntax and semantics of dynamic epistemic logic. We consider a version of this logic that is often considered in the literature (e.g., by Van Ditmarsch, Van der Hoek and Kooi [13]). After describing the logic that we consider in this paper, we briefly relate it to other variants of dynamic epistemic logic that have been considered in the literature.

We fix a countable set $\mathcal{P}$ of propositional variables, and a finite set $\mathcal{A}$ of agents. We begin with introducing the basic language of epistemic logic, and its semantics. The semantics of epistemic logic is based on a type of (Kripke) structures called *epistemic models*. Epistemic models are structures that are used to represent the agents' knowledge about the world and about the other agents' knowledge.

Definition 1 (Epistemic models)**.** *An* epistemic model *is a tuple $\mathcal{M} = (W, R, V)$, where W is a non-empty set of worlds, R maps each agent $a \in \mathcal{A}$ to a relation $R_a \subseteq W \times W$, and $V : \mathcal{P} \to 2^W$ is a function called a valuation. By a slight abuse of notation, we write $w \in \mathcal{M}$ for $w \in W$. We also write $v \in R_a(w)$ for vR_aw. A* single-pointed model *is a pair $(\mathcal{M}, w)$ consisting of an epistemic model $\mathcal{M}$ and a* designated *(or* pointed*)* world *$w \in \mathcal{M}$. A* multi-pointed model *is a pair $(\mathcal{M}, W_d)$ consisting of an epistemic model $\mathcal{M}$ and a subset W_d of designated worlds.*

Definition 2 (Basic epistemic language)**.** *The language $\mathcal{L}_{\text{EL}}$ of epistemic logic is defined as the set of formulas φ defined inductively as follows, where p ranges over $\mathcal{P}$ and a ranges over $\mathcal{A}$:*

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi.$$

The formula $\perp$ is an abbreviation for $p \wedge \neg p$, and the formula $\top$ is an abbreviation for $\neg\perp$. A formula of the form $(\varphi_1 \vee \varphi_2)$ abbreviates $\neg(\neg\varphi_1 \wedge \neg\varphi_2)$, and a formula of the form $(\varphi_1 \to \varphi_2)$ abbreviates $(\neg\varphi_1 \vee \varphi_2)$. Moreover, a formula of the form $\hat{K}_a\varphi$ is an abbreviation for $\neg K_a\neg\varphi$. We call formulas of the form p or $\neg p$ literals. We denote the set of all literals by Lit*.*

Intuitively, the formula $K_a\varphi$ expresses that 'agent a knows that φ holds in the current situation.' Next, we define when a formula in the basic epistemic language is true in a world of an epistemic model.

Definition 3 (Truth conditions for $\mathcal{L}_{\mathrm{EL}}$). *Given an epistemic model $\mathcal{M} = (W, R, V)$, we inductively define the relation $\models \subseteq W \times \mathcal{L}_{\mathrm{EL}}$ as follows. For all $w \in W$:*

$\mathcal{M}, w \models p$	*iff*	$w \in V(p)$
$\mathcal{M}, w \models \neg\varphi$	*iff*	*not* $\mathcal{M}, w \models \varphi$
$\mathcal{M}, w \models \varphi_1 \wedge \varphi_2$	*iff*	*both* $\mathcal{M}, w \models \varphi_1$ *and* $\mathcal{M}, w \models \varphi_2$
$\mathcal{M}, w \models K_a\varphi$	*iff*	*for all* $v \in R_a(w)$*, it holds that* $\mathcal{M}, v \models \varphi$

The statement $\mathcal{M}, w \models \varphi$ expresses that the formula φ is true in world w in the model $\mathcal{M}$.

The framework of dynamic epistemic logic extends the basic epistemic logic with a notion of updates, that are based on another type of structures: *event models*. These are used to represent the effects of an event on the world and the knowledge of the agents. The notion of event models that we use in this paper involves postconditions—to bring about changes in the factual state of the world. Event models with postconditions have been studied and used in the literature on dynamic epistemic logic and epistemic planning (see, e.g., [10, 12]).

Definition 4 (Event models). *An* event model *is a tuple $\mathcal{E} = (E, S, \mathsf{pre}, \mathsf{post})$, where E is a non-empty and finite set of possible events, S maps each agent $a \in \mathcal{A}$ to a relation $S_a \subseteq E \times E$, $\mathsf{pre} : E \to \mathcal{L}_{\mathrm{EL}}$ is a function that maps each event to a precondition expressed in the epistemic language, and $\mathsf{post} : E \to 2^{\mathrm{Lit}}$ is a function that maps each event to a set of literals (not containing complementary literals)[2]. For convenience, we write $\top$ to denote an empty postcondition. By a slight abuse of notation, we write $e \in \mathcal{E}$ for $e \in E$. A* single-pointed event model *is a pair $(\mathcal{E}, e)$ consisting of an event model $\mathcal{E}$ and a* designated *(or* pointed*) event $e \in \mathcal{E}$. A* multi-pointed event model *is a pair $(\mathcal{E}, E_d)$ consisting of an event model $\mathcal{E}$ and a subset $E_d \subseteq E$ of designated events.*

The language of dynamic epistemic logic extends the basic epistemic language with update modalities.

Definition 5 (Dynamic epistemic language). *The language $\mathcal{L}_{\mathrm{DEL}}$ of dynamic epistemic logic is defined as the set of formulas φ defined inductively as follows:*

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_a\varphi \mid [\mathcal{E}, e]\varphi \mid [\mathcal{E}, E_d]\varphi,$$

where p ranges over $\mathcal{P}$ and a ranges over $\mathcal{A}$, and where $(\mathcal{E}, e)$ and $(\mathcal{E}, E_d)$ are single- and multi-pointed event models, respectively. A formula of the form $\langle\mathcal{E}, e\rangle\varphi$ is an

[2] Alternatively, one can define postconditions using a function $\mathsf{post} : E \times \mathcal{P} \to \mathcal{L}_{\mathrm{EL}}$, (see, e.g., [12]). The complexity results in this paper also hold when this alternative definition is used.

abbreviation for $\neg[\mathcal{E}, e]\neg\varphi$*; we use a similar abbreviation* $\langle\mathcal{E}, E_d\rangle\varphi$ *for updates with multi-pointed event models.*

The effect of these event models is defined using the following notion of product update.

Definition 6 (Product update)**.** *Let* $\mathcal{M} = (W, R, V)$ *be an epistemic model and let* $\mathcal{E} = (E, S, \mathsf{pre}, \mathsf{post})$ *be an event model. The* product update of $\mathcal{M}$ by $\mathcal{E}$ *is the epistemic model* $\mathcal{M} \otimes \mathcal{E} = (W', R', V')$ *defined as follows, where* p *ranges over* $\mathcal{P}$ *and* a *ranges over* $\mathcal{A}$*:*

$$\begin{aligned} W' =\ & \{ (w,e) \in W \times E : \mathcal{M}, w \models \mathsf{pre}(e) \} \\ R'_a =\ & \{ ((w,e),(w',e')) \in W' \times W' : wR_aw' \text{ and } eS_ae' \} \\ V'(p) =\ & \{ (w,e) \in W' : w \in V(p) \text{ and } \neg p \notin \mathsf{post}(e) \} \cup \\ & \{ (w,e) \in W' : p \in \mathsf{post}(e) \} \end{aligned}$$

Next, we define when a formula in the dynamic epistemic language is true in a world of an epistemic model.

Definition 7 (Truth conditions for $\mathcal{L}_{\mathrm{DEL}}$)**.** *Given an epistemic model* $\mathcal{M} = (W, R, V)$ *and a formula* $\varphi \in \mathcal{L}_{\mathrm{DEL}}$*, we inductively define the relation* $\models\ \subseteq W \times \mathcal{L}_{\mathrm{DEL}}$ *as follows. For all* $w \in W$*:*

$$\begin{array}{lll} \mathcal{M}, w \models [\mathcal{E}, e]\varphi & \textit{iff} & \mathcal{M}, w \models \mathsf{pre}(e) \textit{ implies } \mathcal{M} \otimes \mathcal{E}, (w,e) \models \varphi \\ \mathcal{M}, w \models [\mathcal{E}, E_d]\varphi & \textit{iff} & \mathcal{M}, w \models [\mathcal{E}, e]\varphi \textit{ for all } e \in E_d \end{array}$$

The other cases are identical to Definition 3. Again, the statement $\mathcal{M}, w \models \varphi$ *expresses that the formula* φ *is true in state* w *in the model* $\mathcal{M}$*.*

(Having defined the language $\mathcal{L}_{\mathrm{DEL}}$, we could now also change the definition of preconditions in event models to be functions $\mathsf{pre} : E \to \mathcal{L}_{\mathrm{DEL}}$ mapping events to formulas in the dynamic epistemic language $\mathcal{L}_{\mathrm{DEL}}$. The definition of product update would work in an entirely similar way. All results in this paper work for either definition of preconditions pre.)

We can then define truth of a formula $\varphi \in \mathcal{L}_{\mathrm{DEL}}$ in epistemic models as follows. A formula φ is true in a single-pointed epistemic model $(\mathcal{M}, w)$ if $\mathcal{M}, w \models \varphi$, and a formula φ is true in a multi-pointed epistemic model $(\mathcal{M}, W_d)$ if $\mathcal{M}, w \models \varphi$ for all $w \in W_d$.

For the purposes of representing knowledge, the relations in epistemic models and event models are often restricted to be equivalence relations, that is, reflexive, transitive and symmetric (see, e.g., [13]). Models that satisfy these requirements are also called *S5 models*, after the axiomatic system that characterizes this type of relations. In the remainder of this paper, we consider only epistemic models and event models that are S5 models. All our hardness results hold for S5 models, as well as for arbitrary models.

For the sake of convenience, we will often depict epistemic models and event models graphically. We will represent worlds with solid dots, events with solid squares, designated worlds and events with a circle or square around them, valuations, preconditions and postconditions with labels next to the dots, and relations with labelled lines between the dots. Since we restrict ourselves to S5 models, and thus to equivalence relations, all relations are symmetric and it suffices to represent relations with undirected lines. Moreover, the reflexive relations are not represented graphically. For a valuation of a world w, we use the literals that the valuation makes true in world w as a label, and for the preconditions and postconditions of an event e, we use the label $\langle \mathsf{pre}(e), \mathsf{post}(e) \rangle$. Moreover, since all epistemic models and event models that we consider in this paper have reflexive relations, in order not to clutter the graphical representation of models, we do not explicitly depict the reflexive relations. For an example of an epistemic model with its graphical representation, see Figure 1, and for an example of an event model with its graphical representation, see Figure 2.

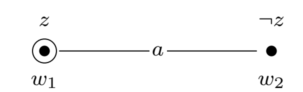


Figure 1: The epistemic model $(\mathcal{M}, w_1)$ for the set $\mathcal{A} = \{a, b\}$ of agents and a single proposition z, where $\mathcal{M} = (W, R, V)$, $W = \{w_1, w_2\}$, $R_a = \{(w_1, w_1), (w_1, w_2), (w_2, w_1), (w_2, w_2)\}$, $R_b = \{(w_1, w_1), (w_2, w_2)\}$, and $V(z) = \{w_1\}$.

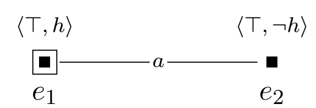


Figure 2: The event model $(\mathcal{E}, e_1)$ for the set $\mathcal{A} = \{a, b\}$ of agents and a single proposition h, where $\mathcal{M} = (E, S, \mathsf{pre}, \mathsf{post})$, $E = \{e_1, e_2\}$, $R_a = \{(e_1, e_1), (e_1, e_2), (e_2, e_1), (e_2, e_2)\}$, $R_b = \{(e_1, e_1), (e_2, e_2)\}$, $\mathsf{pre}(e_1) = \mathsf{pre}(e_2) = \top$, $\mathsf{post}(e_1) = h$, and $\mathsf{post}(e_2) = \neg h$.

Semi-private announcements A particular type of S5 event models that has been considered in the literature are *semi-private* (or *semi-public*) *announcements* [3]. Intuitively, a semi-private announcement publicly announces one of two formulas φ_1, φ_2 to a subset A of agents, and to the remaining agents it publicly announces that one of the two formulas is the case, and that the agents in A learned which one is. A semi-private announcement for formulas φ_1, φ_2 and a subset $A \subseteq \mathcal{A}$ of agents is represented by the event model in Figure 3.

To illustrate the notion of semi-private announcements, consider the following example scenario. There are two agents, Ayla (a) and Blair (b). Ayla flips a coin, which lands either on heads (h) or on tails ($\neg h$), and hides the result of the coin flip from Blair. Blair sees that the coin is flipped and that Ayla knows the result of the coin flip, but Blair herself does not see the result of the coin flip. This semi-private announcement is represented by the event model $\mathcal{E}$ that is depicted in Figure 2 (in the event model depicted in Figure 2, the coin lands on heads).

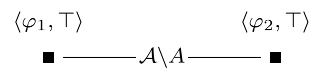


Figure 3: A semi-private announcement for formulas φ_1, φ_2 and the subset $A \subseteq \mathcal{A}$ of agents.

Relations to other variants of Dynamic Epistemic Logic The formalism of dynamic epistemic logic that we consider is based on the one originally introduced by Baltag, Moss, and Solecki [4, 5]. Their language only considers single-pointed event models. A few years later, Baltag and Moss [3] extended this original language to include regular operators (union, composition and 'star') for the update modalities. The language that we consider corresponds to the variant of their language with only the union operator. The language presented by Van Ditmarsch et al. in their textbook [13] resembles the language that we consider, as their framework also allows the union operator for updates, but not the composition or 'star' operators. The union operator for update modalities corresponds to allowing multi-pointed event models. Because it simplifies notation, we use multi-pointed models, following the notation of other existing work [10]. Additionally, the language that we consider also allows events to have postconditions, unlike the language presented by Van Ditmarsch et al. [13].

2.2 Computational Complexity

Next, we review some basic notions from computational complexity that are used in the proofs of the results that we present. We assume the reader to be familiar with the complexity classes P and NP, and with basic notions such as polynomial-time reductions. For more details, we refer to textbooks on computational complexity theory (see, e.g., [1]).

The class PSPACE consists of all decision problems that can be solved by an algorithm that uses a polynomial amount of space (memory). Alternatively, one can characterize the class PSPACE as all decision problems for which there exists a polynomial-time reduction to the problem QSat, that is defined using quantified Boolean formulas as follows. A (fully) quantified Boolean formula (in prenex form) is a formula of the form $Q_1x_1Q_2x_2 \dots Q_nx_n.\psi$, where all x_i are propositional variables, each Q_i is either an existential or a universal quantifier, and ψ is a (quantifier-free) propositional formula over the variables $x_1, \dots, x_n$. Truth for such formulas is defined in the usual way. The problem QSat consists of deciding whether a given quantified Boolean formula is true. Moreover, QSat is PSPACE-hard even when restricted to the case where $Q_i = \exists$ for odd i and $Q_i = \forall$ for even i. (For the proofs of Theorems 3, 4 and 5, we will use reductions from this restricted variant of QSat.)

Additionally, one can restrict the number of quantifier alternations occurring in quantified Boolean formulas, i.e., the number of times where $Q_i \neq Q_{i+1}$. For each constant $k \geq 1$ number of alternations, this leads to a different complexity class. These classes together constitute the Polynomial Hierarchy. We consider the complexity classes Σ^{p}_k, for each $k \geq 1$. The complexity class Σ^{p}_k consists of all decision problems for which there exists a polynomial-time reduction to the problem QSat_k. Instances of QSat_k are quantified Boolean formulas of the form $\exists x_1 \dots \exists x_{\ell_1} \forall x_{\ell_1+1} \dots \forall x_{\ell_2} \dots Q_k x_{\ell_{k-1}+1} \dots Q_k x_{\ell_k}.\ \psi$, where $Q_k = \exists$ if k is odd and $Q_k = \forall$ if k is even, where $1 \leq \ell_1 \leq \dots \leq \ell_k$, and where ψ is quantifier-free. The question is to decide whether the quantified Boolean formula is true.

The last complexity class that we consider is Δ^{p}_2. We give a definition of this class that is based on algorithms with access to an oracle, i.e., a black box that is able to decide certain decision problems in a single operation. Consider the problem Sat of deciding satisfiability of a given propositional formula. The class Δ^{p}_2 consists of all decision problems that can be solved by a polynomial-time algorithm with access to an oracle for Sat. Alternatively, the class Δ^{p}_2 consists of all decision problems for which there exists a polynomial-time reduction to the problem where one is given a satisfiable propositional formula φ over the variables $x_1, \dots, x_n$, and the question is whether the lexicographically maximal assignment that satisfies φ (given the fixed ordering $x_1 \prec \dots \prec x_n$) sets variable x_n to true [24]. An assignment α_1 is lexicographically larger than an assignment α_2 (given the ordering $x_1 \prec \cdots \prec x_n$) if there exists some $1 \leq i \leq n$ such that $\alpha_1(x_i) = 1$, $\alpha_2(x_i) = 0$, and for all $1 \leq j \leq i$ it holds that $\alpha_1(x_j) = \alpha_2(x_j)$.

3 Results for updates with arbitrary S5 models

In this section, we provide complexity results for the model checking problem for DEL when arbitrary event models are allowed for the update modalities in the formulas. For several cases, we prove PSPACE-hardness. Since the problem was recently shown to be in PSPACE for the most general variant of dynamic epistemic logic that we consider in this paper [2, 29, 30], these hardness results suffice to show PSPACE-completeness.

3.1 Polynomial-time solvability

We begin with showing polynomial-time solvability for the strongest restriction that we consider in this paper: a single agent, single-pointed S5 models, and no postconditions. This is a result that is well-known and can be seen as part of the folklore of the DEL literature, but for which—to the best of our knowledge—no detailed proof has been published. The high-level idea of the proof is straightforward: even though updating the model with an event model in this setting might duplicate a lot of words—potentially resulting in an exponential blow-up in the number of worlds—these worlds are copies of only a small number of distinct worlds. We can identify a representative for each of the worlds in polynomial-time, and at each step in the recursive evaluation of the formula, we keep only these representative worlds. To work out an algorithm that implements this proof idea requires some attention to algorithmic details (e.g., using the technique of dynamic programming). We present the proof of Proposition 1 in detail to give insight into the exact algorithmic details involved in the proof—providing a precise recipe that can be used to implement the algorithm.

Proposition 1. *The model checking problem for DEL with S5 models is polynomial-time solvable when restricted to instances with a single agent and only single-pointed event models without postconditions.*

Proof. We describe a polynomial-time algorithm that solves the problem. The main idea behind this algorithm is the following. Even though the updates might cause an exponential blow-up in the number of worlds in the model, in this restricted setting, we only need to remember a small number of these worlds.

Concretely, since there is only a single agent a, and since there is only a single designated world w_0, we only need to remember the set of worlds that are connected with an a-relation to the designated world w_0. Moreover, among these worlds, we can merge those with an identical valuation. Since the event models contain no postconditions, this (contracted) set of worlds can only decrease after updates, i.e., updates can only remove worlds from this set.

Formally, we can describe this argument as follows. Let $(\mathcal{M}, w_0)$ be a single-pointed S5 epistemic model with one agent, and let $(\mathcal{E}, e_0)$ be a single-pointed S5 event model with one agent and no postconditions. Then $\mathcal{M} \otimes \mathcal{E}$ is bisimilar to a submodel $\mathcal{M}'$ of $\mathcal{M}$, that is, to some $\mathcal{M}'$ that can be obtained from $\mathcal{M}$ by removing some worlds. Specifically, let W' be the set of worlds in $\mathcal{M}$ that are a-accessible from w_0, and let E' be the set of events in $\mathcal{E}$ that are a-accessible from e_0. Then, let $W'' \subseteq W'$ be the subset of worlds that satisfy the precondition of at least one $e \in E'$. One can straightforwardly verify that $(\mathcal{M} \otimes \mathcal{E}, (w_0, e_0))$ is bisimilar to the submodel $(\mathcal{M}', w_0)$ of $(\mathcal{M}, w_0)$ induced by W''. Moreover, $\mathcal{M}'$ can be computed in polynomial time.

Using this property, we can construct a recursive algorithm to decide whether $\mathcal{M}, w \models \varphi$. We consider several cases. In the case where $\varphi = p$ for some $p \in \mathcal{P}$, the problem can easily be solved in polynomial time, by simply checking whether $w \in V(p)$. In the case where $\varphi = \neg\varphi_1$, we can recursively call the algorithm to decide whether $\mathcal{M}, w \models \varphi_1$, and return the opposite answer. Similarly, for $\varphi = \varphi_1 \wedge \varphi_2$, we can straightforwardly decide whether $\mathcal{M}, w \models \varphi$ by first recursively determining whether $\mathcal{M}, w \models \varphi_1$ and whether $\mathcal{M}, w \models \varphi_2$. In the case where $\varphi = K_a\varphi_1$, we firstly recursively determine whether $\mathcal{M}, w' \models \varphi_1$ for each $w' \in W$ that is a-accessible from w. This information immediately determines whether $\mathcal{M}, w \models K_a\varphi_1$.

Finally, consider the case where $\varphi = [\mathcal{E}, e]\varphi_1$. In this case, we firstly recursively decide whether $\mathcal{M}, w \models \mathsf{pre}(e)$. If this is not the case, then trivially, $\mathcal{M}, w \models \varphi$. Otherwise, we construct the submodel $\mathcal{M}'$ of $\mathcal{M}$ that is bisimilar to $\mathcal{M} \otimes \mathcal{E}$. This can be done as described above. In order to do this, we need to decide which states in W' satisfy the precondition of some $e' \in E'$, where $W' \subseteq W$ and $E' \subseteq E$ are defined as explained above. This can be done by recursive calls of the algorithm. Having determined W', and having constructed $\mathcal{M}'$, we can now answer the question whether $\mathcal{M}, w \models [\mathcal{E}, e]\varphi_1$ by using only $\mathcal{M}'$, w and φ_1. We know that w is a world in $\mathcal{M}'$, since $\mathcal{M}, w \models \mathsf{pre}(e)$. Since $\mathcal{M}'$ is bisimilar to $\mathcal{M} \otimes \mathcal{E}$, it holds that $\mathcal{M} \otimes \mathcal{E}, (w, e) \models \varphi_1$ if and only if $\mathcal{M}', w \models \varphi_1$. Therefore, by recursively calling the algorithm to decide whether $\mathcal{M}', w \models \varphi_1$, we can decide whether $\mathcal{M}, w \models [\mathcal{E}, e]\varphi_1$.

It is straightforward to verify that this recursive algorithm correctly decides whether $\mathcal{M}, w \models \varphi$. However, naively executing this recursive algorithm will, in the worst case result in an exponential running time. This is because for the case for $\varphi = [\mathcal{E}, e]\varphi_1$, the algorithm makes multiple (say $b \geq 2$) recursive calls for $\mathsf{pre}(e)$, and $\mathsf{pre}(e)$ could contain subformulas of the form $[\mathcal{E}', e']\varphi'$—which in turn triggers multiple recursive calls for $\mathsf{pre}(e')$ for each of the b branches in the recursion tree, and so forth. As the number of these iterations can grow linearly with the input size (say $f(n)$), the recursion tree can be of exponential size (namely, of size $\geq 2^{f(n)}$). We describe how to modify the algorithm to run in polynomial time, using the technique of memoization. Whenever a recursive call is made to decide whether $\mathcal{N}, u \models \psi$, for some submodel $\mathcal{N}$ of $\mathcal{M}$, some world w in $\mathcal{N}$, and some subformula ψ of φ, the result of this recursive call is stored in a lookup table. Moreover, before making a recursive call to decide whether $\mathcal{N}, u \models \psi$, the lookup table is consulted, and if an answer is stored, the algorithm uses this answer instead of executing the recursive call.

The number of submodels $\mathcal{N}$ of $\mathcal{M}$ that need to be considered in the execution of the modified algorithm is upper bounded by the number of occurrences of update operators $[\mathcal{E}, e]$ in the formula φ that is given as input to the problem. Therefore, the size of the lookup table is polynomial in the input size. Moreover, computing the answer for any entry in the lookup table can be done in polynomial time (using the answers for other entries in the lookup table). Therefore, the modified algorithm decides whether $\mathcal{M}, w \models \varphi$ in polynomial time. □

3.2 Hardness results for one agent

Next, we consider the restriction where we have a single agent and single-pointed models, but where postconditions are allowed in the event models.[3] In this case, the problem is Δ_2^{p}-hard. This hardness result is interesting because it helps identify the boundaries of the tractable fragment of Proposition 1. The result of Theorem 2 shows that adding the single element of postconditions to this tractable fragment leads to computational hardness.

Theorem 2. *The model checking problem for DEL with S5 models restricted to instances with a single agent and only single-pointed models, but where event models can contain postconditions, is Δ_2^{p}-hard.*

Proof. To show Δ_2^{p}-hardness, we give a polynomial-time reduction from the problem of deciding whether the lexicographically maximal assignment that satisfies a given

[3] The reader might wonder for what type of situations the DEL setting of Theorem 2 (a single agent, single-pointed S5 models, and postconditions) could be useful. This restricted setting is relevant, for example, when reasoning about the epistemic state of a single agent in the face of undercainty over changes in the world (made by nature). A simple example of a situation where such reasoning plays a role is in the analysis of single-player memory games where the state of (parts of) the game board can be changed randomly by the rules of the game.

propositional formula φ over variables $x_1, \ldots, x_n$ sets the variable x_n to true. Let φ be an instance of this problem, with variables $x_1, \ldots, x_n$. We construct a single-pointed epistemic model $(\mathcal{M}, w_0)$ with a single agent a and a DEL-formula χ whose updates consist of single-pointed event models (that contain postconditions), such that $\mathcal{M}, w_0 \models \chi$ if and only if x_n is true in the lexicographically maximal assignment that satisfies φ.

In addition to the propositional variables $x_1, \ldots, x_n$, we introduce a variable z. Then, we construct the model $(\mathcal{M}, w_0)$ as depicted in Figure 4.

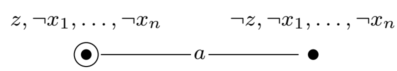


Figure 4: The epistemic model $(\mathcal{M}, w_0)$, used in the proof of Theorem 2.

Then, for each $1 \leq i \leq n$, we introduce the single-pointed event model $(\mathcal{E}_i, e_i)$ as depicted in Figure 5. Intuitively, these updates will serve to generate, for each possible truth assignment α to the variables $x_1, \ldots, x_n$, a world that agrees with α (and that sets z to false), in addition to the designated world (where z is true). We will denote the model resulting from updating $(\mathcal{M}, w_0)$ subsequently with the updates $(\mathcal{E}_1, e_1), \ldots, (\mathcal{E}_n, e_n)$ by $(\mathcal{M}', w')$.

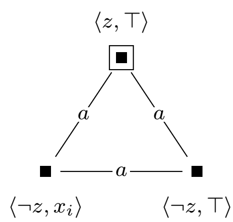


Figure 5: The event model $(\mathcal{E}_i, e_i)$, used in the proof of Theorem 2.

Next, for each $1 \leq i \leq n$, we introduce the single-pointed event model $(\mathcal{E}'_i, e'_i)$ as depicted in Figure 6. Intuitively, we will use the event models $(\mathcal{E}'_i, e'_i)$ to obtain (many copies of) the lexicographically maximal assignment that satisfies φ. Applying the update $(\mathcal{E}'_i, e'_i)$ to $(\mathcal{M}', w')$ will set the variable x_i to true in all worlds (that satisfy $\neg z$) if there is an assignment (among the remaining assignments) that satisfies φ and that sets x_i to true, and it will set the variable x_i to false in all worlds (that satisfy $\neg z$) otherwise. Then, after applying the updates $(\mathcal{E}'_1, e'_1), \ldots, (\mathcal{E}'_n, e'_n)$ to $(\mathcal{M}', w')$, all worlds in the resulting model will have the same valuation—namely, a valuation that agrees with the lexicographically maximal assignment that satisfies φ. In particular, the variable x_n is true in this valuation if and only if x_n is true in the lexicographically maximal assignment that satisfies φ.

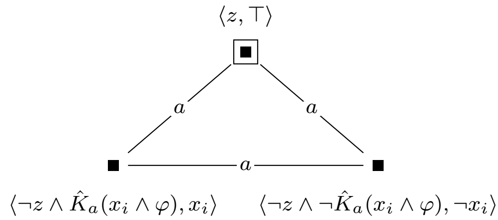


Figure 6: The event model $(\mathcal{E}'_i, e'_i)$, used in the proof of Theorem 2.

We then let $\chi = [\mathcal{E}_1, e_1] \dots [\mathcal{E}_n, e_n][\mathcal{E}'_1, e'_1] \dots [\mathcal{E}'_n, e'_n]\hat{K}_a x_n$. We now formally show that the lexicographically maximal assignment α_0 that satisfies φ sets x_n to true if and only if $\mathcal{M}, w_0 \models \chi$. In order to do so, we will prove the following claim. The model $(\mathcal{M}'', w'') = (\mathcal{M}, w_0) \otimes (\mathcal{E}_1, e_1) \otimes \cdots \otimes (\mathcal{E}_n, e_n) \otimes (\mathcal{E}'_1, e'_1) \otimes \cdots \otimes (\mathcal{E}'_n, e'_n)$ consists of a world w'' that sets z to true and all other variables to false, and of worlds that set z to false and that agree with α_0 on the variables $x_1, \dots, x_n$. Firstly, it is straightforward to verify that $(\mathcal{M}', w') = (\mathcal{M}, w_0) \otimes (\mathcal{E}_1, e_1) \otimes \cdots \otimes (\mathcal{E}_n, e_n)$ consists of the world w' and exactly one world corresponding to each truth assignment α to the variables $x_1, \dots, x_n$.

Then, applying the update $(\mathcal{E}'_1, e'_1)$ to $(\mathcal{M}', w')$ has two possible outcomes: either (1) if there exists a model of φ that sets x_1 to true, then in all worlds (that set z to false) the variable x_1 will be set to true; or (2) if there exists no model of φ that sets x_1 to true, then in all worlds (that set z to false) the variable x_1 will be set to false. For each $1 < i \leq n$, subsequently applying the update $(\mathcal{E}'_i, e'_i)$ has an entirely similar effect. By a straightforward inductive argument, it then follows that all the worlds in $(\mathcal{M}'', w'')$ that set z to false agree with the lexicographically maximal model of φ.

Therefore, $\mathcal{M}, w_0 \models \chi$ if and only if x_n is true in the lexicographically maximal model of φ, and we can conclude that the problem is Δ^{p}_2-hard. □

When we allow multi-pointed models, the problem turns out to be PSPACE-hard, even when restricted to a single agent (Theorem 3). This hardness result adds to our understanding of the boundaries of the algorithmically tractable fragment of Proposition 1. Whereas Theorem 2 showed that adding postconditions leads to intractability, the following result shows that adding multi-pointedness to the models instead (and having no further additions) also leads to intractability. In other words, the following result indicates that leaving the fragment of Proposition 1 by a different route also requires giving up polynomial-time algorithms for model checking.

In the literature, PSPACE-hardness results have been shown for a setting that is similar to the one used in Theorem 3—i.e., [16, Proposition 7.2] and [2, Theorem 2]. The difference is that these proofs in the literature depend on particular features of the DEL setting—the result of [16, Proposition 7.2] depends on there being two agents, and the result of [2, Theorem 2] depends on relations not being serial—whereas the result of Theorem 3 holds also for the case with both a single agent and S5 relations. The proofs from the literature depends crucially on there being two agents or non-serial relations—they are used to encode the quantifiers in a quantified Boolean formula. The main technical hurdle that needs to be overcome to establish Theorem 3 is to encode the quantification of a quantified Boolean formula in DEL using a single agent and using S5 relations. We do so by starting with an S5 epistemic model $\mathcal{M}$ that includes worlds that set different propositional variables $x_1, \dots, x_n$ to true, and using multi-pointed event models to quantify over different possibilities of deleting worlds from $\mathcal{M}$.

Theorem 3. *The model checking problem for DEL with S5 models restricted to instances with a single agent and no postconditions in the event models, but where models can be multi-pointed, is* PSPACE*-hard.*

Proof. In order to show PSPACE-hardness, we give a polynomial-time reduction from the problem of deciding whether a quantified Boolean formula is true. Let $\varphi = \exists x_1 \forall x_2 \dots \exists x_{n-1} \forall x_n . \psi$ be a quantified Boolean formula, where ψ is quantifier-free (we assume without loss of generality that n is even). We construct a single-pointed epistemic model $(\mathcal{M}, w_0)$ with one agent a and a DEL-formula χ (containing updates with multi-pointed event models) such that $\mathcal{M}, w_0 \models \chi$ if and only if φ is true.

The first main idea behind this reduction is that we represent truth assignments to the propositional variables $x_1, \dots, x_n$ with connected groups of worlds. Let α be a truth assignment to the variables $x_1, \dots, x_n$, and let $x_{i_1}, \dots, x_{i_\ell}$ be the variables that α sets to true. We then represent α by means of a group of worlds $w_0, w_1, \dots, w_\ell$, where the world w_0 makes no propositional variable true, and for each $1 \leq j \leq \ell$, world w_j makes exactly one propositional variable true (namely, x_{i_j}). These worlds $w_0, w_1, \dots, w_\ell$ are fully connected. This collection of worlds $w_0, w_1, \dots, w_\ell$ is what we call the *group of worlds corresponding to* α. Moreover, the designated state is w_0. Consider the truth assignment $\alpha = \{x_1 \mapsto 1, x_2 \mapsto 1, x_3 \mapsto 0, x_4 \mapsto 1\}$, for example. In Figure 7 we show the group of worlds that we use to represent this truth assignment α. We let the model $\mathcal{M}$ be the group of worlds corresponding to the truth assignment α_0 that assigns all variables $x_1, \dots, x_n$ to true.

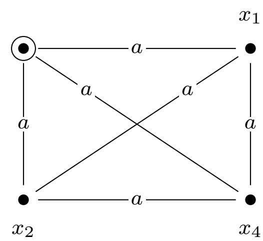


Figure 7: The group of worlds that we use to represent the truth assignment $\alpha = \{x_1 \mapsto 1, x_2 \mapsto 1, x_3 \mapsto 0, x_4 \mapsto 1\}$, in the proof of Theorem 3.

The next main idea is that we represent existential and universal quantification of the propositional variables using the dynamic operators $\langle \mathcal{E}, E \rangle$ and $[\mathcal{E}, E]$, respectively. For each propositional variable x_i in the quantified Boolean formula, we introduce the multi-pointed event model $(\mathcal{E}_i, E_i)$ as depicted in Figure 8. We use the event models $(\mathcal{E}_1, E_1), \dots, (\mathcal{E}_n, E_n)$, to create (disconnected) groups of worlds (that all have a designated world) that correspond to each possible truth assignment α to the variables $x_1, \dots, x_n$.

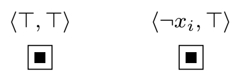


Figure 8: The multi-pointed event model $(\mathcal{E}_i, E_i)$ corresponding to variable x_i, used in the proof of Theorem 3.

Using the alternation of diamond dynamic operators and box dynamic operators, we can simulate existential and universal quantification of variables in the formula φ. We simulate an existentially quantified variable $\exists x_i$ by the dynamic operator $\langle \mathcal{E}_i, E_i \rangle$—a formula of the form $\langle \mathcal{E}_i, E_i \rangle \phi$ is true if and only if $\langle \mathcal{E}_i, e_i \rangle \phi$ is true for some $e_i \in E_i$. Similarly, we simulate a universally quantified variable $\forall x_i$ by the dynamic operator $[\mathcal{E}_i, E_i]$—a formula of the form $[\mathcal{E}_i, E_i]\phi$ is true if and only $[\mathcal{E}_i, e_i]\phi$ is true for all $e_i \in E_i$.

Concretely, we let $\chi = \langle \mathcal{E}_1, E_1 \rangle [\mathcal{E}_2, E_2] \dots \langle \mathcal{E}_{n-1}, E_{n-1} \rangle [\mathcal{E}_n, E_n] \chi'$, where χ' is the formula obtained from ψ by replacing each occurrence of a propositional variable x_i by the formula $\hat{K}_a x_i$.

We show that φ is a true quantified Boolean formula if and only if $\mathcal{M}, w_0 \models \chi$. In order to do so, we prove the following statement, relating truth assignments α to the variables $x_1, \dots, x_n$ to groups of worlds containing a designated world. The statement that we will prove inductively for all $1 \leq i \leq n+1$ is the following.

Statement: Let α be any truth assignment to the variables $x_1, \dots, x_n$ that sets all variables $x_i, \dots, x_n$ to true, and let α' be the restriction of α to the variables $x_1, \dots, x_{i-1}$. Moreover, let M be a group of worlds that corresponds to the truth assignment α, containing a designated world w. Then $Q_i x_i \dots \exists x_{n-1} \forall x_n . \psi$ is true under α' if and only if:

- w makes $\langle \mathcal{E}_i, E_i \rangle \dots [\mathcal{E}_n, E_n]\chi'$ true, if i is odd; and
- w makes $[\mathcal{E}_i, E_i] \dots [\mathcal{E}_n, E_n]\chi'$ true, if i is even.

The statement for $i = 1$ implies that $\mathcal{M}, w_0 \models \chi$ if and only if φ is a true quantified Boolean formula. We show that the statement holds for $i = 1$ by showing that the statement holds for all $1 \leq i \leq n+1$. We begin by showing that the statement holds for $i = n+1$. In this case, we know that $\alpha = \alpha'$ is a truth assignment to the variables $x_1, \dots, x_n$. Moreover, by construction of χ' we know that w makes χ' true if and only if α satisfies ψ. Therefore, the statement holds.

Next, we let $1 \leq i \leq n$ be arbitrary, and we assume that the statement holds for $i+1$. We now distinguish two cases: either (1) $Q_i = \exists$, i.e., the i-th quantifier of φ is existential, or (2) $Q_i = \forall$, i.e., the i-th quantifier of φ is universal.

First, consider case (1). Suppose that $\exists x_i \dots \exists x_{n-1} \forall x_n . \psi$ is true under α'. Then there exists a truth assignment α'' to the variables $x_1, \dots, x_i$ that agrees with α' on the variables $x_1, \dots, x_{i-1}$ and for which $\forall x_{i+1} \dots \exists x_{n-1} \forall x_n . \psi$ is true under α''. Therefore, there exists some event $e \in E_i$ such that the group $M' = \{ (v, e) : v \in M \text{ and } M, v \models \mathsf{pre}(e) \}$ of worlds and the world $w' = (w, e)$, together with the assignment α''' that agrees with α'' on the variables $x_1, \dots, x_i$ and that sets the variables $x_{i+1}, \dots, x_n$ to true, satisfy the requirements for the statement for $i+1$. Then, by the induction hypothesis we know that w' makes $[\mathcal{E}_{i+1}, E_{i+1}] \dots \langle \mathcal{E}_{n-1}, E_{n-1} \rangle [\mathcal{E}_n, E_n]\chi'$ true. Therefore, we can conclude that w makes $\langle \mathcal{E}_i, E_i \rangle \dots \langle \mathcal{E}_{n-1}, E_{n-1} \rangle [\mathcal{E}_n, E_n]\chi'$ true.

Conversely, suppose that w makes $\langle \mathcal{E}_i, E_i \rangle \dots \langle \mathcal{E}_{n-1}, E_{n-1} \rangle [\mathcal{E}_n, E_n]\chi'$ true. This can only be the case if there is some event $e \in E_i$ such that the set M' and w' (defined as above) correspond to a truth assignment α''' (also defined as above). Then, by the induction hypothesis, we know that $\forall x_{i+1} \dots \exists x_{n-1} \forall x_n . \psi$ is true under α'' (obtained from α''' as above). Therefore, since α'' extends α', we can conclude that $\exists x_i \dots \exists x_{n-1} \forall x_n . \psi$ is true under α'.

The argument for case (2) is entirely analogous (yet dual). We omit a detailed treatment of this case. This concludes the inductive proof of the statement for all $1 \leq i \leq n+1$, and thus concludes our proof that $\mathcal{M}, w_0 \models \chi$ if and only if φ is true. $\square$

3.3 Hardness results for two agents

Next, we show that when we consider the case of two agents, the model checking problem for DEL is PSPACE-hard, even when we only allow single-pointed models without postconditions (Theorem 4). This hardness result adds yet another piece of understanding of the boundaries of the algorithmically tractable fragment of Proposition 1. Namely, it shows that leaving the algorithmically tractable fragment of Proposition 1 by another one of the possible different routes—that is, by adding a second agent only—leads to computational hardness. In fact, this is the third of the three most obvious ways of extending the fragment of Proposition 1. Therefore, Theorem 4—together with Theorems 2 and 3—indicates that all individual restrictions in the fragment of Proposition 1 are necessary to obtain polynomial-time solvability.

Additionally, the result of Theorem 4 shows that the inherent hardness in the model checking problem for DEL with two agents holds even when we restrict the setting to include only three propositional variables x_1, x_2, x_3—in addition to having only two agents and single-pointed S5 models only.

Similarly to the case of Theorem 3, the result of Theorem 4 differs from similar results from the literature—i.e., [16, Proposition 7.2] and [2, Theorem 2]—in that it considers different restrictions than these results. In particular, the former result [16, Proposition 7.2] requires the use of multi-pointed models, and the latter result [2, Theorem 2] requires non-serial relations in the models (and these proofs crucially depend on these features). In the literature, there are also PSPACE hardness results for the case of single-pointed S5 models [29, 30]. However, these results hold for the case where an unbounded number of agents are used—and the proofs given in the literature crucially depend on the number of agents not being bounded. The result of Theorem 4 holds for the case with only two agents and where only single-pointed S5 models are allowed.

The main technical hurdle that needs to be overcome to establish Theorem 4 is to encode the differently quantified variables of a quantified Boolean formula using single-pointed S5 models and using only two agents and three propositional variables. We do so—roughly—by (i) representing propositional variables by alternating chains of worlds where the end of the chain is marked by a designated propositional variable z_0,[4] (ii) representing quantification over different variables in the quantified Boolean formula using S5 event models where different alternating chains of relations lead to different copies of the original model that represent different truth assignments to the variables of the quantified Boolean formula, and (iii) by introducing DEL formulas that interact appropriately with the gadgets of (i) and (ii), making use

[4] This technique was previously used by Van de Pol, Van Rooij, and Szymanik in a hardness proof for DEL model checking [30, Proposition 3].

of only two auxiliary variables z_1 and z_2. The main challenge in (i)–(iii) is in the details of the intricate construction of the DEL formulas and in the detailed argument that they interact in exactly the right way with the event models—this is why the proof of Theorem 4 is of considerable length.

Theorem 4. *The model checking problem for DEL is* PSPACE*-hard, even when restricted to the case where the question is whether* $\mathcal{M}, w_0 \models [\mathcal{E}_1, e_1] \dots [\mathcal{E}_n, e_n]\chi$, *where:*

- *the model* $(\mathcal{M}, w_0)$ *is a single-pointed S5 model;*
- *all the* $(\mathcal{E}_i, e_i)$ *are single-pointed S5 event models without postconditions;*
- χ *is an epistemic formula without update modalities that contains (multiple occurrences of) only three propositional variables; and*
- *there are only two agents.*

Proof. In order to show PSPACE-hardness, we give a polynomial-time reduction from the problem of deciding whether a quantified Boolean formula is true. Let $\varphi = \exists x_1 \forall x_2 \dots \exists x_{n-1} \forall x_n . \psi$ be a quantified Boolean formula, where ψ is quantifier-free. We construct an epistemic model $(\mathcal{M}, w_0)$ with two agents a, b and a DEL-formula ξ such that $\mathcal{M}, w_0 \models \xi$ if and only if φ is true.

The first main idea behind the reduction is that we use two propositional variables, say z_0 and z_1, to represent an arbitrary number of propositions, by creating chains of worlds that represent these propositions. Let $x_1, \dots, x_n$ be the propositions that we want to represent. Then we represent a proposition x_j by a chain of worlds of length $j + 1$ that are connected alternatingly by b-relations and a relations. In this chain, the last world is the only world that makes z_0 true. Moreover, the first world in the chain is the only world that makes z_1 true. An example of such a chain that we use to represent proposition x_3 can be found in Figure 9.

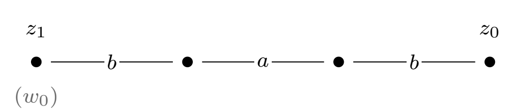


Figure 9: The chain of worlds that we use to represent proposition x_3 in the proof of Theorem 4. The first world in the chain is the world w_0, that is depicted on the left.

The next idea that plays an important role in the reduction is that we will group together (the first worlds) of several such chains to represent a truth assignment to the propositions $x_1, \dots, x_n$. Let α be a truth assignment to the propositions $x_1, \dots, x_n$, and let $x_{i_1}, \dots, x_{i_\ell}$ be the propositions that α sets to true (for $1 \leq i_1 < \dots < i_\ell \leq n$). Then we represent the truth assignment α in the following way. We take the chains corresponding to the propositions $x_{i_1}, \dots, x_{i_\ell}$, and we connect the first world (the world that is labelled with w_0 in Figure 9) of each two of these chains with an a-relation. In other words, we connect all these first worlds together in a fully connected clique of a-relations. Moreover, to this a-clique of worlds, we add a designated world where both z_1 and a third propositional variable z_2 are true. The collection of all worlds in the chains corresponding to the propositions $x_{i_1}, \dots, x_{i_\ell}$ and this additional designated world is what we call the *group of worlds representing* α. For the sake of convenience, we call the world where z_1 and z_2 are true the *central world* of the group of worlds. In Figure 10 we give an example of such a group of worlds that we use to represent the truth assignment $\alpha = \{x_1 \mapsto 1, x_2 \mapsto 1, x_3 \mapsto 0, x_4 \mapsto 1\}$.

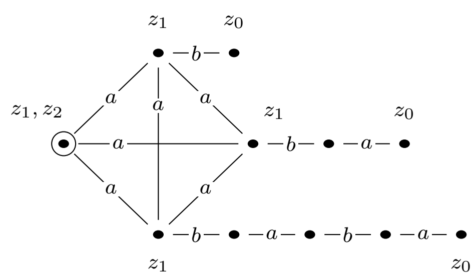


Figure 10: The group of worlds that we use to represent the truth assignment $\alpha = \{y_1 \mapsto 1, y_2 \mapsto 1, y_3 \mapsto 0, y_4 \mapsto 1\}$ in the proof of Theorem 4.

By using the expressivity of epistemic logic, we can construct formulas that extract information from these representations of truth assignments. Intuitively, we can check whether a truth assignment α sets a proposition x_j to true by checking whether the group of worlds representing α contains a chain of worlds of length exactly $j + 1$. Formally, we will define a formula χ_j for each $1 \leq j \leq n$, which is true in the designated world if and only if the group contains a chain representing proposition x_j. We describe how to construct the formulas χ_j. Firstly, we inductively define formulas χ_j^a and χ_j^b, for all $1 \leq j \leq n$ as follows. Intuitively, the formula χ_j^a is true in exactly those worlds from which there is an alternating chain that ends in a z_0-world, that is of length at least j and that starts with an a-relation. Similarly, the formula χ_j^b is true in exactly those worlds from which there is an alternating chain that ends in a z_0-world, that is of length at least j, and that starts with a b-relation. We let $\chi_0^a = \chi_0^b = z_0$, and for each $j > 0$, we let $\chi_j^b = \hat{K}_b(\neg z_1 \wedge \neg z_2 \wedge \chi_{j-1}^a)$

and $\chi_j^a = \hat{K}_a(\neg z_1 \wedge \neg z_2 \wedge \chi_{j-1}^b)$. Now, using the formulas χ_j^b, we can define the formulas χ_j. We let $\chi_j = z_1 \wedge \neg z_2 \wedge \chi_j^b \wedge \neg\chi_{j-1}^b$. As a result of this definition, the formula χ_j is true in exactly those worlds that are the first world of a chain of length exactly j.

For example, consider the formula $\chi_2 = z_1 \wedge \neg z_2 \wedge \hat{K}_b(\neg z_1 \wedge \neg z_2 \wedge \hat{K}_a(\neg z_1 \wedge \neg z_2 \wedge z_0)) \wedge \neg\hat{K}_b(\neg z_1 \wedge \neg z_2 \wedge z_0)$ and consider the group of worlds depicted in Figure 10. This formula is true only in the first world of the chain of length 2.

The epistemic model $(\mathcal{M}, w_0)$ that we use in the reduction is based on the model $\mathcal{M}_{\alpha_0}$ representing the truth assignment $\alpha_0 : \{x_1, \dots, x_n\} \to \{0, 1\}$ that sets all propositions $x_1, \dots, x_n$ to true. To obtain $\mathcal{M}$, we will add a number of additional worlds to the model $\mathcal{M}_{\alpha_0}$, that we will use to simulate the behavior of the existential and universal quantifiers in the DEL-formula χ that we will construct below. Specifically, we will add alternating chains of worlds to the model that are similar to the chains that represent the propositions $x_1, \dots, x_n$. However, the additional chains that we add differ in two aspects from the chains that represent the propositions $x_1, \dots, x_n$: (1) the additional chains start with an a-relation instead of starting with a b-relation, and (2) in the first world of the additional chains, the propositional variable z_2 is true instead of the variable z_1. For each $1 \leq i \leq n$, we add such an additional chain of length i, and we connect the first worlds of these additional chains, together with the designated world, in a clique of b-relations. To illustrate this, the model $(\mathcal{M}, w_0)$ that results from this construction is shown in Figure 11, for the case where $n = 3$. For the sake of convenience, we will denote these additional chains by z_2-*chains*, and the chains that represent the propositions $x_1, \dots, x_n$ by z_1-*chains* (after the propositional variables that are true in the first worlds of these chains).

To check whether an alternating chain of length exactly j, that starts from a z_2-world with an a-relation, is present in the model, we define formulas χ_j' similarly to the way we defined the formulas χ_j. Specifically, we let $\chi_j' = \neg z_1 \wedge z_2 \wedge \chi_j^a \wedge \neg\chi_{j-1}^a$.

We will use the z_2-chains together with the formulas χ_j' to keep track of an additional counter. We will use this counter as a technical trick to implement the simulation of existentially and universally quantified variables in the formula φ.

Next, we describe how we can generate all possible truth assignments over the variables $x_1, \dots, x_n$ from the initial model $\mathcal{M}$. We do this in such a way that we can afterwards express the existential and universal quantifications of the formula φ using modal operators in the epistemic language. In order to generate groups of worlds that represent truth assignments α that differ from the all-ones assignment α_0, we will apply updates that copy the existing worlds but that eliminate (the first worlds of) chains of a certain length. This is the third main idea behind this reduction.

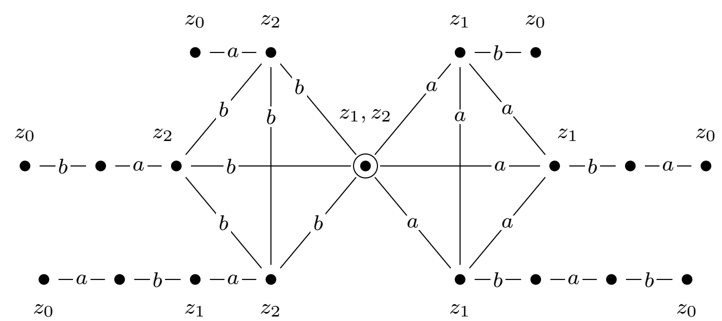


Figure 11: The epistemic model $(\mathcal{M}, w_0)$ for the case where $n = 3$, as used in the proof of Theorem 4.

Specifically, we will introduce a single-pointed event model $(\mathcal{E}_i, e_i)$ for each propositional variable x_i, that is depicted in Figure 12. Intuitively, what happens when the update $(\mathcal{E}_i, e_i)$ is applied is the following. All existing groups of worlds will be duplicated, resulting in five copies—this corresponds to the five events $f_i^1, \dots, f_i^5$ in the event model. The resulting groups of worlds will be connected corresponding to the relations between the events in the event model. That is, for any existing group of worlds, three of its copies (corresponding to the events f_i^1, f_i^2, and f_i^3) will be connected by b-relations. The second and third of these copies (the ones corresponding to the events f_i^2 and f_i^3) will be connected by a-relations to the fourth and fifth copy (corresponding to the events f_i^4 and f_i^5), respectively. Moreover, in the fourth and fifth copy, (the first world of) the z_2-chain of length i is removed, and in the fifth copy, (the first world of) the z_1-chain of length i is removed as well. These effects of removing (the first worlds of) chains is enforced by the preconditions of the events in the event model.

By applying the updates $(\mathcal{E}_1, e_1), \dots, (\mathcal{E}_n, e_n)$, we generate many (in fact, an exponential number of) copies of the model $\mathcal{M}$, in each of which certain chains of worlds are removed, and which are connected to each other by means of a-relations and b-relations in the way described in the previous paragraph. In particular, for each truth assignment α to the propositions $x_1, \dots, x_n$, there is some group of worlds that corresponds to α.

Finally, we construct the DEL-formula ξ. We let $\xi = [\mathcal{E}_1, e_1] \dots [\mathcal{E}_n, e_n]\xi_1$, where we define ξ_1 below. The formula ξ_1 exploits the structure of the epistemic model $\mathcal{M}'$, that results from updating the model $\mathcal{M}$ with the updates $(\mathcal{E}_i, e_i)$, to simulate the semantics of the quantified Boolean formula φ. For each $1 \leq i \leq n + 1$, we define ξ_i

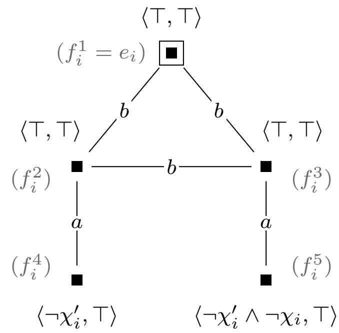


Figure 12: The event model $(\mathcal{E}_i, e_i)$ corresponding to the propositional variable x_i, used in the proof of Theorem 4. The events are labelled $f_i^1, \dots, f_i^5$.

inductively as follows:

$$\xi_i = \begin{cases} \psi' & \text{if } i = n+1, \\ \hat{K}_b \hat{K}_a (z_1 \wedge z_2 \wedge \bigwedge_{1 \le j \le i} \neg \hat{K}_b \chi_j' \wedge \bigwedge_{i < j \le n} \hat{K}_b \chi_j' \wedge \xi_{i+1}) & \text{for odd } i \le n, \\ K_b K_a ((z_1 \wedge z_2 \wedge \bigwedge_{1 \le j \le i} \neg \hat{K}_b \chi_j' \wedge \bigwedge_{i < j \le n} \hat{K}_b \chi_j') \rightarrow \xi_{i+1}) & \text{for even } i \le n. \end{cases}$$

Here, ψ' is the formula obtained from ψ (the quantifier-free part of the quantified Boolean formula φ) by replacing each occurrence of a propositional variable x_i by the formula $\hat{K}_a \chi_i$.

We use the formulas ξ_i to express the formula φ with its existentially and universally quantified variables. Intuitively, the formulas ξ_i navigate through the groups of worlds in the model $\mathcal{M}'$—resulting from updating $\mathcal{M}$ with the event models $(\mathcal{E}_i, e_i)$—as follows. Consider the central world of some group of worlds in the model $\mathcal{M}'$, and consider the formula ξ_i for some odd $i \le n$. For ξ_i to be true in this world, the formula ξ_{i+1} needs to be true in the central world of *some* group of worlds that corresponds to one of the events f_i^4 or f_i^5 from the event model $(\mathcal{E}_i, e_i)$. Similarly, for the formula ξ_i to be true in this world, for even $i \le n$, the formula ξ_{i+1} needs to be true in the central world of *both* groups of worlds that correspond to the events f_i^4 and f_i^5. In this way, for odd $i \le n$, the formula ξ_i together with the event model $(\mathcal{E}_i, e_i)$ serves to simulate an existential choice of a truth value for the variable x_i. Similarly, for even $i \le n$, the formula ξ_i together with the event model $(\mathcal{E}_i, e_i)$ serves to simulate a universal choice of a truth value for the variable x_i.

The model $(\mathcal{M}, w_0)$ and the formula ξ can be constructed in polynomial time in the size of the quantified Boolean formula φ. Furthermore, in the constructed instance, there are only two agents, the epistemic model $(\mathcal{M}, w_0)$ is a single-pointed S5 model, all event models $(\mathcal{E}_i, e_i)$ are single-pointed S5 models without postconditions, and ξ_1 is a formula without update modalities that contains (many occurrences of) only three propositional variables z_0, z_1, z_2.

We show that φ is a true quantified Boolean formula if and only if $\mathcal{M}, w_0 \models \xi$. In order to do so, we prove the following (technical) statement relating truth assignments α to the propositions $x_1, \dots, x_n$ and (particular) worlds w in the epistemic model $(\mathcal{M}', w_0') = (\mathcal{M}, w_0) \otimes (\mathcal{E}_1, e_1) \otimes \cdots \otimes (\mathcal{E}_n, e_n)$. Before we give the statement that we will prove, we observe that every world w that sets both z_1 and z_2 to true is the central world of some group of worlds that represents a truth assignment α to the propositions $x_1, \dots, x_n$. For the sake of convenience, we will say that w corresponds to the truth assignment α. The statement that we will prove for all $1 \leq i \leq n+1$ is the following.

Statement: Let α be any truth assignment to the propositions $x_1, \dots, x_{i-1}$. Moreover, let w be any world in the model $(\mathcal{M}', w_0')$ such that:

1. w makes z_1 and z_2 true,
2. w makes $\hat{K}_b \chi_j'$ false for all $1 \leq j < i$,
3. w makes $\hat{K}_b \chi_j'$ true for all $i \leq j \leq n$, and
4. the truth assignment corresponding to w agrees with α on the propositions $x_1, \dots, x_{i-1}$.

Then the (partially) quantified Boolean formula $Q_i x_i \dots \exists x_{n-1} \forall x_n . \psi$ is true under α if and only if w makes ξ_i true.

Observe that for $i = 1$, the world w_0' satisfies all four conditions. Therefore, the statement for $i = 1$ implies that $\mathcal{M}, w_0 \models \xi$ if and only if φ is a true quantified Boolean formula. Thus, proving this statement for all $1 \leq i \leq n+1$ suffices to show the correctness of our reduction.

We begin by showing that the statement holds for $i = n+1$. In this case, we know that α is a truth assignment to the propositions $x_1, \dots, x_n$. Moreover, $\xi_{n+1} = \psi'$. By construction of ψ', we know that w makes ψ' true if and only if α satisfies ψ. Therefore, the statement holds for $i = n+1$.

Next, we let $1 \leq i \leq n$ be arbitrary, and we assume that the statement holds for $i+1$. That is, the statement holds for every combination of a truth assignment α and a world w that satisfies the conditions. Since w is a world in the model $(\mathcal{M}, w_0) \otimes (\mathcal{E}_1, e_1) \otimes \cdots \otimes (\mathcal{E}_n, e_n)$, and since w makes z_1 and z_2 true, we know that $w =$

$(w_0, e'_1, \dots, e'_n)$, for some $e'_1, \dots, e'_n$, where $e'_j \in \mathcal{E}_j$ for all $1 \leq j \leq n$. We know that w makes $\hat{K}_b \chi'_j$ true for all $i \leq j \leq n$. Therefore, we know that for each $i \leq j \leq n$ it holds that $e'_j \in \{f_j^1, f_j^2, f_j^3\}$.

We now distinguish two cases: either (1) i is odd, or (2) i is even. In case (1), the i-th quantifier of φ is existential, and in case (2), the i-th quantifier of φ is universal. First, consider case (1). Suppose that $\exists x_i \dots \exists x_{n-1} \forall x_n.\psi$ is true under α. Then there exists some truth assignment α' to the propositions $x_1, \dots, x_i$ that agrees with α on the propositions $x_1, \dots, x_{i-1}$ and that ensures that $\forall x_{i+1} \dots \exists x_{n-1} \forall x_n.\psi$ is true under α'. Suppose that $\alpha'(x_i) = 0$; the case for $\alpha'(x_i) = 1$ is entirely similar. Now, consider the worlds $w' = (w_0, e'_1, \dots, e'_{i-1}, f_i^3, e'_{i+1}, \dots, e'_n)$ and $w'' = (w_0, e'_1, \dots, e'_{i-1}, f_i^5, e'_{i+1}, \dots, e'_n)$. By the construction of $(\mathcal{E}_i, e_i)$, by the semantics of product update, and by the fact that $e'_i \in \{f_i^1, f_i^2, f_i^3\}$, it holds that $(w, w') \in R_b$ and $(w', w'') \in R_a$. Moreover, it is straightforward to verify that w'' satisfies conditions (1)–(4), for the truth assignment α'. Also, we know that $\forall x_{i+1} \dots \exists x_{n-1} \forall x_n.\psi$ is true under α'. Therefore, by the induction hypothesis, we know that w'' makes the formula ξ_{i+1} true. It then follows from the definition of ξ_i that w' and w'' witness that w makes ξ_i true.

Conversely, suppose that w makes ξ_i true. Moreover, suppose that w' and w'' (as defined above) witness this. (The only other possible worlds u' and u'' that could witness this are obtained from w by replacing e'_i by f_i^2 and f_i^4, respectively. The case where u' and u'' witness that w makes ξ_i true is entirely similar.) This means that w'' makes ξ_{i+1} true. Then, by the induction hypothesis, it follows that the truth assignment α' to the propositions $x_1, \dots, x_i$ corresponding to the world w'' satisfies $\forall x_{i+1} \dots \exists x_{n-1} \forall x_n.\psi$. Moreover, since α' agrees with α on the propositions $x_1, \dots, x_{i-1}$, it follows that $\exists x_i \dots \exists x_{n-1} \forall x_n.\psi$ is true under α.

Next, consider case (2). Suppose that $\forall x_i \dots \exists x_{n-1} \forall x_n.\psi$ is true under α. Then for both truth assignments α' to the variables $x_1, \dots, x_i$ that agree with α it holds that $\exists x_{i+1} \dots \exists x_{n-1} \forall x_n.\psi$ is true under α'. The only worlds that satisfy $(z_1 \wedge z_2 \wedge \bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j \wedge \bigwedge_{i < j \leq n} \hat{K}_b \chi'_j)$ and that are accessible from w by a b-relation followed by an a-relation are the worlds u_1 and u_2, where $u_1 = (w_0, e'_1, \dots, e'_{i-1}, f_i^4, e'_{i+1}, \dots, e'_n)$ and $u_2 = (w_0, e'_1, \dots, e'_{i-1}, f_i^5, e'_{i+1}, \dots, e'_n)$. Moreover, the truth assignments α_1 and α_2 that correspond to u_1 and u_2, respectively, agree with α on the propositions $x_1, \dots, x_{i-1}$. Because $\forall x_i \dots \exists x_{n-1} \forall x_n.\psi$ is true under α, we know that $\exists x_{i+1} \dots \exists x_{n-1} \forall x_n.\psi$ is true under both α_1 and α_2. Then, by the induction hypothesis it follows that both u_1 and u_2 make ξ_{i+1} true. Therefore, w makes ξ_i true.

Conversely, suppose that w makes ξ_i true. By the definition of ξ_i, we then know that all worlds that are accessible from w by a b-relation followed by an a-relation and that make $(z_1 \wedge z_2 \wedge \bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j \wedge \bigwedge_{i < j \leq n} \hat{K}_b \chi'_j)$ true, also make ξ_{i+1} true. Consider the worlds u_1 and u_2, as defined above. These are both accessible from w by a b-relation followed by an a-relation, and they make $(z_1 \wedge z_2 \wedge \bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j \wedge \bigwedge_{i < j \leq n} \hat{K}_b \chi'_j)$ true. Therefore, both u_1 and u_2 make ξ_{i+1} true. Also, the truth assignments α_1 and α_2 that correspond to u_1 and u_2, respectively, agree with α on the propositions $x_1, \dots, x_{i-1}$. Moreover, the truth assignments α_1 and α_2 are both possible truth assignments to the propositions $x_1, \dots, x_i$ that agree with α. By the induction hypothesis, the formula $\exists x_{i+1} \dots \exists x_{n-1} \forall x_n . \psi$ is true under both α_1 and α_2. Therefore, we can conclude that $\forall x_i \dots \exists x_{n-1} \forall x_n . \psi$ is true under α.

This concludes the inductive proof of the statement for all $1 \leq i \leq n+1$, and thus concludes our correctness proof. Therefore, we can conclude that the problem is PSPACE-hard. □

4 Results for semi-private announcements

Next, we consider the model checking problem for DEL when restricted to updates that are semi-private announcements. In fact, PSPACE-hardness for the setting with semi-private announcements (rather than allowing arbitrary event models) already follows from a recent PSPACE-hardness proof for a restricted variant of the model checking problem— see [29, Theorem 4] and [30, Theorem 1]. In that PSPACE-hardness result, the number of agents is unbounded, i.e., the number of agents is part of the problem input. We show that the problem is already PSPACE-hard when the number of agents is bounded by any constant $k \geq 2$.

The result of Theorem 5 shows that the inherent hardness of the model checking problem for DEL—which we saw in Theorem 4 is already present in a very restricted setting—is even present when we restrict event models to be of a very specific shape (i.e., semi-private announcements).

Theorem 5 is a stronger result than Theorem 4—Theorem 5 implies the result of Theorem 4. We presented the proof of Theorem 4 in full detail, because it allows us to explain the proof of Theorem 5 in a clear way. In fact, the result of Theorem 5 is stronger than all other PSPACE-hardness results for the model checking problem for DEL in the literature—i.e., [16, Proposition 7.2], [2, Theorem 2] and [29, 30]. These results all depend on allowing certain parts of the DEL setting being unrestricted—e.g., allowing multi-pointed models, allowing more than two agents, or allowing non-serial relations—and their proofs cannot easily be modified to work for the more restricted setting of Theorem 5 with (single-pointed) semi-private announcements, S5 relations, only two agents, and only three propositional variables.

The main technical hurdle that needs to be overcome to establish Theorem 5 is to simulate the event models that we used in the proof of Theorem 4 (each consisting of 5 events) by using a sequence of semi-private announcements (which are event models with 2 events). We do this by constructing—for each event model—three semi-private announcements, in such a way that when these three semi-private announcements are composed, they take the role of the event model in the proof. In order to make sure that the semi-private announcements correctly take over the role of the event models in the proof, we also need to adapt the part of the DEL formula that expresses (the unquantified part of) the quantified Boolean formula.

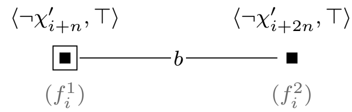
Theorem 5. *The model checking problem for DEL is* PSPACE*-hard, even when restricted to the case where the question is whether* $\mathcal{M}, w_0 \models [\mathcal{E}_1, e_1] \dots [\mathcal{E}_n, e_n]\chi$, *where:*

- *the model* $(\mathcal{M}, w_0)$ *is a single-pointed S5 model;*
- *all the* $(\mathcal{E}_i, e_i)$ *are (single-pointed) semi-private announcements;*
- χ *is an epistemic formula without update modalities that contains (multiple occurrences of) only three propositional variables; and*
- *there are only two agents.*

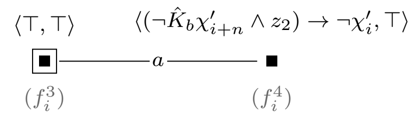
Proof. We modify the proof of Theorem 4 to work also for the case of semi-private announcements. Most prominently, we will replace the event models $(\mathcal{E}_i, e_i)$ that are used in the proof of Theorem 4 (shown in Figure 12) by a number of event models for semi-private announcements. Intuitively, these semi-private announcements will take the role of the event models $(\mathcal{E}_i, e_i)$. In order to make this work, we will also slightly change the initial model $\mathcal{M}$.

As in the proof of Theorem 4, we give a polynomial-time reduction from the problem of deciding whether a quantified Boolean formula is true. Let $\varphi = \exists x_1 \forall x_2 \dots \exists x_{n-1} \forall x_n . \psi$ be a quantified Boolean formula, where ψ is quantifier-free. We construct an epistemic model $(\mathcal{M}, w_0)$ with two agents a, b and a DEL-formula ξ such that $\mathcal{M}, w_0 \models \xi$ if and only if φ is true.

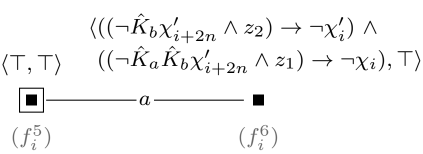
In the proof of Theorem 4, the initial model $\mathcal{M}$ consisted of a central world (where z_1 and z_2 are true), a number of z_1-chains (for each $1 \leq i \leq n$, there is a z_1-chain of length i), and a number of z_2-chains (for each $1 \leq i \leq n$, there is a z_2-chain of length i)—and these worlds are connected by a-relations and b-relations as shown in Figure 11. To obtain the initial model $\mathcal{M}$ that we use in this proof, we add additional z_2-chains. Specifically, for each $1 \leq i \leq 3n$, we will have a z_2-chain of length i. These additional z_2-chains are connected to the central world in exactly the same way as the original z_2-chains (that is, all the first worlds of the z_2-chains are connected in a b-clique to the central world). We will use these additional z_2-chains to simulate the behavior of the event models $(\mathcal{E}_i, e_i)$ from the proof of Theorem 4 with event models corresponding to semi-private announcements. The number of z_1-chains remains the same. The designated world w_0 is the central world (that is, the only world that makes both z_1 and z_2 true), as in the proof of Theorem 4.



(a) The semi-private announcement $(\mathcal{E}_i^1, f_i^1)$



(b) The semi-private announcement $(\mathcal{E}_i^2, f_i^3)$



(c) The semi-private announcement $(\mathcal{E}_i^3, f_i^5)$

Figure 13: The semi-private announcements $(\mathcal{E}_i^1, f_i^1)$, $(\mathcal{E}_i^2, f_i^3)$ and $(\mathcal{E}_i^3, f_i^5)$ used in the proof of Theorem 5.

The event models $(\mathcal{E}_i, e_i)$ that are used in the proof of Theorem 4 we replace by the semi-private announcements $(\mathcal{E}_i^1, f_i^1)$, $(\mathcal{E}_i^2, f_i^3)$, and $(\mathcal{E}_i^3, f_i^5)$, as shown in Figure 13. The intuition behind these updates is the following. Firstly, the semi-private announcement $\mathcal{E}_i^1$, shown in Figure 13a, transforms every group of worlds into two copies, and allows a choice between these two copies when following a b-relation. Moreover, in one copy, every (first world of the) z_2-chain of length $i+n$ is removed, and in the other copy, every (first world of the) z_2-chain of length $i+2n$ is removed. In other words, the choice between these two copies determines whether the formula $\hat{K}_b\chi'_{i+n}$ or the formula $\hat{K}_b\chi'_{i+2n}$ is false in the central world. Then for every group of worlds that does not include a z_2-chain of length $i+n$, the semi-private announcement $\mathcal{E}_i^2$, shown in Figure 13b, creates an a-accessible copy where the z_2-chain of length i is removed. Similarly, for every group of worlds that does not include a z_2-chain of length $i + 2n$, the semi-private announcement $\mathcal{E}_i^3$, shown in Figure 13c, creates an a-accessible copy where both the z_2-chain of length i and the z_1-chain of length i are removed.

Next, we construct the DEL-formula ξ. We let $\xi = [\mathcal{E}_1^1, f_1^1][\mathcal{E}_1^2, f_1^3][\mathcal{E}_1^3, f_1^5] \dots [\mathcal{E}_n^1, f_n^1][\mathcal{E}_n^2, f_n^3][\mathcal{E}_n^3, f_n^5]\xi_1$, where ξ_1 is defined as follows, similarly to the definition used in the proof of Theorem 4. For each $1 \leq i \leq n+1$, we define ξ_i inductively as follows:

$$\xi_i = \begin{cases} \psi' & \text{if } i = n+1, \\ \hat{K}_b(\bigwedge\limits_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j \wedge \hat{K}_a(z_1 \wedge z_2 \wedge \bigwedge\limits_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j \wedge \bigwedge\limits_{i < j \leq n} \hat{K}_b \chi'_j \wedge \xi_{i+1})) & \text{for odd } i \leq n, \\ K_b((\bigwedge\limits_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j) \rightarrow K_a((z_1 \wedge z_2 \wedge \bigwedge\limits_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j \wedge \bigwedge\limits_{i < j \leq n} \hat{K}_b \chi'_j) \rightarrow \xi_{i+1})) & \text{for even } i \leq n. \end{cases}$$

Here, ψ' is the formula obtained from ψ (the quantifier-free part of the quantified Boolean formula φ) by replacing each occurrence of a propositional variable x_i by the formula $\hat{K}_a \chi_i$.

The formulas ξ_i that we defined above are very similar to their counterparts in the proof of Theorem 4—and the idea behind their use in the proof is entirely the same as in the proof of Theorem 4. The only difference is the addition of the subformulas $\bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j$ after the first modal operator. These additional subformulas are needed to ensure that some additional worlds—that are a by-product of the combination of the semi-private announcements $(\mathcal{E}_i^1, f_i^1)$, $(\mathcal{E}_i^2, f_i^3)$, and $(\mathcal{E}_i^3, f_i^5)$—do not interfere in the reduction.

We show that φ is a true quantified Boolean formula if and only if $\mathcal{M}, w_0 \models \xi$. In order to do so, as in the proof of Theorem 4, we prove the following (technical) statement relating truth assignments α to the propositions $x_1, \dots, x_n$ and (particular) worlds w in the epistemic model $(\mathcal{M}', w'_0) = (\mathcal{M}, w_0) \otimes (\mathcal{E}_1^1, f_1^1) \otimes \cdots \otimes (\mathcal{E}_n^3, f_n^5)$. Before we give the statement that we will prove, we observe that every world w that sets both z_1 and z_2 to true is the central world of some group of worlds that represents a truth assignment α to the propositions $x_1, \dots, x_n$. For the sake of convenience, we will say that w corresponds to the truth assignment α. The statement that we will prove for all $1 \leq i \leq n+1$ is the following.

Statement: Let α be any truth assignment to the propositions $x_1, \dots, x_{i-1}$. Moreover, let w be any world in the model $(\mathcal{M}', w'_0)$ such that:

1. w makes z_1 and z_2 true,
2. w makes $\hat{K}_b \chi'_j$ false for all $1 \leq j < i$,
3. w makes $\hat{K}_b \chi'_j$ true for all $i \leq j \leq n$, and

4. the truth assignment corresponding to w agrees with α on the propositions $x_1, \dots, x_{i-1}$.

Then the (partially) quantified Boolean formula $Q_i x_i \dots \exists x_{n-1} \forall x_n . \psi$ is true under α if and only if w makes ξ_i true.

Observe that for $i = 1$, the world w'_0 satisfies all four conditions. Therefore, the statement for $i = 1$ implies that $\mathcal{M}, w_0 \models \xi$ if and only if φ is a true quantified Boolean formula. Thus, proving this statement for all $1 \leq i \leq n+1$ suffices to show the correctness of our reduction.

We begin by showing that the statement holds for $i = n+1$. In this case, we know that α is a truth assignment to the propositions $x_1, \dots, x_n$. Moreover, $\xi_{n+1} = \psi'$. By construction of ψ', we know that w makes ψ' true if and only if α satisfies ψ. Therefore, the statement holds for $i = n+1$.

Next, we let $1 \leq i \leq n$ be arbitrary, and we assume that the statement holds for $i+1$. That is, the statement holds for every combination of a truth assignment α and a world w that satisfies the conditions. Since w is a world in the model $(\mathcal{M}, w_0) \otimes (\mathcal{E}^1_1, f^1_1) \otimes \dots \otimes (\mathcal{E}^3_n, f^5_n)$, and since w makes z_1 and z_2 true, we know that $w = (w_0, g_1, g'_1, g''_1, \dots, g_n, g'_n, g''_n)$, for some $g_1, g'_1, g''_1, \dots, g_n, g'_n, g''_n$, where for each $1 \leq j \leq n$, it holds that $g_j \in \{f^1_j, f^2_j\}$, $g'_j \in \{f^3_j, f^4_j\}$, and $g''_j \in \{f^5_j, f^6_j\}$.

We now distinguish two cases: either (1) i is odd, or (2) i is even. In case (1), the i-th quantifier of φ is existential, and in case (2), the i-th quantifier of φ is universal. First, consider case (1). Suppose that $\exists x_i \dots \exists x_{n-1} \forall x_n . \psi$ is true under α. Then there exists some truth assignment α' to the propositions $x_1, \dots, x_i$ that agrees with α on the propositions $x_1, \dots, x_{i-1}$ and that ensures that $\forall x_{i+1} \dots \exists x_{n-1} \forall x_n . \psi$ is true under α'. Suppose that $\alpha'(x_i) = 0$; the case for $\alpha'(x_i) = 1$ is entirely similar. Now, consider the worlds $w' = (w_0, g_1, \dots, g''_{i-1}, f^2_i, g'_i, g''_i, g_{i+1}, \dots, g''_n)$ and $w'' = (w_0, g_1, \dots, g''_{i-1}, f^2_i, g'_i, f^6_i, g_{i+1}, \dots, g''_n)$ By the construction of $\mathcal{E}^1_i$, $\mathcal{E}^2_i$, and $\mathcal{E}^3_i$, and by the semantics of product update, it holds that $(w, w') \in R_b$ and $(w', w'') \in R_a$. Also, we know that w' makes $\bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j$ true. Moreover, it is straightforward to verify that w'' satisfies conditions (1)–(4), for the truth assignment α'. Also, we know that $\forall x_{i+1} \dots \exists x_{n-1} \forall x_n . \psi$ is true under α'. Therefore, by the induction hypothesis, we know that w'' makes the formula ξ_{i+1} true. It then follows from the definition of ξ_i that w' and w'' witness that w makes ξ_i true.

Conversely, suppose that w makes ξ_i true. Moreover, suppose that w' and w'' (as defined above) witness this. It could also be the case that the worlds u' and u'' witness this, which are obtained from w by replacing g_i by f^1_i, and by replacing g_i by f^1_i and g'_i by f^4_i, respectively. The case where u' and u'' witness that w makes ξ_i true is entirely similar. (There are also variants of w' and w'', and of u' and u'', that could witness the fact that w makes ξ_i true. These variants can be obtained by replacing g_j, g'_j, and g''_j—for $i < j \leq n$—ensuring that for all $i < j \leq n$ it holds that neither (1) $g_j = f_j^1$ and $g'_j = f_j^4$ nor (2) $g_j = f_j^2$ and $g''_j = f_j^6$. The following argument is entirely similar for these variants. Therefore, we restrict our attention to the worlds w' and w''.) The assumption that w' and w'' witness that w makes ξ_i true implies that w' makes $\bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j$ true and that w'' makes ξ_{i+1} true. Then, by the induction hypothesis, it follows that the truth assignment α' to the propositions $x_1, \dots, x_i$ corresponding to the world w'' satisfies $\forall x_{i+1} \dots \exists x_{n-1} \forall x_n . \psi$. Moreover, since α' agrees with α on the propositions $x_1, \dots, x_{i-1}$, it follows that $\exists x_i \dots \exists x_{n-1} \forall x_n . \psi$ is true under α.

Next, consider case (2). Suppose that $\forall x_i \dots \exists x_{n-1} \forall x_n . \psi$ is true under α. Then for both truth assignments α' to the variables $x_1, \dots, x_i$ that agree with α it holds that $\exists x_{i+1} \dots \exists x_{n-1} \forall x_n . \psi$ is true under α'. We need to look at those worlds that satisfy $(z_1 \wedge z_2 \wedge \bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j \wedge \bigwedge_{i < j \leq n} \hat{K}_b \chi'_j)$ and that are accessible from w by a b-relation followed by an a-relation (where the intermediate world makes $\bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j$ true). For our argument, it suffices to look at the worlds u_1 and u_2, where $u_1 = (w_0, g_1, \dots, g''_{i-1}, f_i^1, f_i^4, g''_i, g_{i+1}, \dots, g''_n)$ and $w'' = (w_0, g_1, \dots, g''_{i-1}, f_i^2, g'_i, f_i^6, g_{i+1}, \dots, g''_n)$. (As in the argument for case (1) above, there are variants of these worlds that also satisfy the requirements. The argument for these variants is entirely similar, and therefore we restrict our attention to the worlds u_1 and u_2.) The truth assignments α_1 and α_2 that correspond to u_1 and u_2, respectively, agree with α on the propositions $x_1, \dots, x_{i-1}$. Because $\forall x_i \dots \exists x_{n-1} \forall x_n . \psi$ is true under α, we know that $\exists x_{i+1} \dots \exists x_{n-1} \forall x_n . \psi$ is true under both α_1 and α_2. Then, by the induction hypothesis it follows that both u_1 and u_2 make ξ_{i+1} true. Therefore, w makes ξ_i true.

Conversely, suppose that w makes ξ_i true. By the definition of ξ_i, we then know that all worlds that are accessible from w by a b-relation followed by an a-relation and that make $(z_1 \wedge z_2 \wedge \bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j \wedge \bigwedge_{i < j \leq n} \hat{K}_b \chi'_j)$ true (where the intermediate world makes $\bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j$ true), also make ξ_{i+1} true. Consider the worlds u_1 and u_2, as defined above. These are both accessible from w by a b-relation followed by an a-relation (where the intermediate world makes $\bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j$ true), and they make $(z_1 \wedge z_2 \wedge \bigwedge_{1 \leq j \leq i} \neg \hat{K}_b \chi'_j \wedge \bigwedge_{i < j \leq n} \hat{K}_b \chi'_j)$ true. Therefore, both u_1 and u_2 make ξ_{i+1} true. Also, the truth assignments α_1 and α_2 that correspond to u_1 and u_2, respectively, agree with α on the propositions $x_1, \dots, x_{i-1}$. Moreover, the truth assignments α_1 and α_2 are both possible truth assignments to the propositions $x_1, \dots, x_i$ that agree with α. By the induction hypothesis, the formula $\exists x_{i+1} \dots \exists x_{n-1} \forall x_n . \psi$ is true under both α_1 and α_2. Therefore, we can conclude that $\forall x_i \dots \exists x_{n-1} \forall x_n . \psi$ is true under α.

This concludes the inductive proof of the statement for all $1 \leq i \leq n + 1$, and thus concludes our correctness proof. Therefore, we can conclude that the problem is PSPACE-hard. □

5 Discussion

In this section, we reflect on the relevance and significance of our results in the overall endeavor of obtaining a well-informed and useful understanding of the computational properties of the model checking problem for DEL. In particular, we discuss (a) how our results contribute to the undertaking of getting a detailed picture of the computational complexity of DEL model checking, (b) why such a detailed theoretical picture is useful and important for the development and improvement of DEL model checking algorithms, and (c) how we can get an even more detailed picture of the complexity of DEL model checking in future research.

Detailed worst-case complexity analysis In this paper—as in most works in the literature on the study of the computational properties of DEL—we adopt the framework of worst-case computational complexity analysis (see, e.g., [1]). This is a framework that has been hugely influential and successful, but that also has its inherent downsides. One of its main disadvantages is that it is prone to give an overly negative image of the computational difficulty of a problem. It is not uncommon for a problem to be computationally hard in the worst case sense, while instances of this problem that come up in applications can be solved efficiently. Therefore, for a worst-case computational complexity analysis to provide an accurate picture of the inherent complexity of a problem, it needs to be as fine-grained and detailed as possible. This means that it needs to consider many different restricted settings that are relevant to applications that use the problem under study.

The results that we provided in Sections 3 and 4 contribute to the detail and fine-grainedness of the computational complexity study of the model checking problem for DEL. Previous work on the computational complexity of the problem investigated restricted settings—where certain components of the problem are restricted in number or shape. However, this mostly involved studies where restrictions only involve a single component of the problem [2, 8, 16, 23]. There has been some research that considered combinations of restrictions [29, 30], but this work only focuses on the line between polynomial-time solvable (and an extension thereof: fixed-parameter tractability) and computationally intractable. Our results both (i) take into account combinations of restrictions on different components of the problem, and (ii) are aimed at identifying the exact degree of complexity of the problem—distinguishing between different degrees of computational intractability. As such, our results provide an important and useful step in the direction of establishing a detailed and fine-grained picture of the worst-case complexity of the model checking problem for DEL.

Guidance for model checking algorithms Establishing a detailed picture of the exact degree of computational complexity for a wide range of restricted settings provides a good guide for the development of practical model checking algorithms. For example, the results of Theorems 4 and 5 show that the model checking problem can require polynomial space even in restricted settings with a limited number of agents and propositions. This suggests that better performing algorithms could be obtained by using optimized algorithmic approaches for PSPACE-complete problems. For example, it would be interesting to investigate whether encoding the model checking problem for DEL into the satisfiability problem for quantified Boolean formulas (QBFs) and subsequently invoking QBF solvers on the resulting formula could lead to model checking algorithms that are competitive with existing model checking algorithms—even on instances that involve only a limited number of agents and propositions. Such an approach would have the benefit that years of research and engineering effort on developing QBF solvers (see, e.g., [17]) could be leveraged to get efficient algorithms.

The results that we established in this paper indicate that currently implemented model checking algorithms for DEL—that are based on constructing a representation of the epistemic model resulting from the original model and updates applied to it—are likely to run into barriers of combinatorial explosion already in very limited settings. For example, the results of Theorems 2, 3 and 4 show that deviating from the restricted setting of Proposition 1 in one of various minimal ways leads to a setting where any (deterministic) algorithm cannot run in polynomial time in the worst case. Examples of model checking algorithms for which these results are relevant are those of DEMO [15] and SMCDEL [9, 21].

Our results also suggest directions for experimental evaluation of (implemented) model checking algorithms. For example, it would be useful to investigate on which types of instances algorithms such as DEMO and SMCDEL perform well, and on which types of instances they in fact run into barriers of combinatorial explosion. For instances where DEMO and SMCDEL perform poorly, it would be interesting to study whether model checking algorithms based on QBF solvers—and other optimized algorithmic methods for PSPACE-complete problems—perform better. The computational complexity results in this paper provide indications for which properties of inputs could have an impact on the performance of different model checking algorithms.

Parameterized complexity analysis The foundational computational complexity results for DEL model checking that we developed in this paper (and that other papers in the literature developed) enable an interesting direction for future research—namely, to employ the framework of *parameterized complexity theory* (see, e.g., [11, 14, 20]). This would take the undertaking of providing a more fine-grained worst-case complexity analysis even a step further. Parameterized complexity theory provides a complexity-theoretic framework that can be used to identify which parts of the problem input contribute in what way to the running time (or space usage) of algorithms solving the model checking problem for DEL. This framework has already been used to initiate a more detailed investigation of the computational complexity of the model checking problem for DEL [29, 30]. Further pursuing this research direction has the potential of yielding useful and relevant insights into the computational properties of DEL.

The results in this paper provide a constructive foundation for establishing parameterized complexity results for the model checking problem for DEL. For example, the hardness result of Theorem 4 tells us that the model checking problem for DEL is para-PSPACE-complete when parameterized by the number of agents and the number of propositional variables occurring in the formula—and thus is not fixed-parameter tractable for this parameter. (For more details on the relation between traditional computational complexity results and parameterized complexity, we refer to the literature—e.g., [19, 20].)

Studying other restrictions Another way forward in the study of the computational properties of the model checking problem for DEL that is pointed at by the results that we provide in this paper, is to consider restrictions on the problem input that go beyond counting simple quantities in the input (such as the number of agents or the number of propositions) and instead are based on structural properties of the input. An example of this would be to consider the computational properties of settings where models are restricted to those whose underlying graph has certain graph-theoretic properties—such as bounded treewidth.[5] Whereas simple restrictions only lead to positive algorithmic results in only a very limited number of cases—as indicated by the results that we provide in this paper—structural restrictions have the potential of leading to positive results in more general settings. Tractability results based on such structural properties could then of course be used to develop efficient model checking algorithms that are tailored to application settings where

[5]Treewidth is a measure that, intuitively, captures how similar a graph is to a tree (trees have minimum possible treewidth). Restricting problems to graphs of bounded treewidth often yields efficient algorithms for problems that are intractable in general (see, e.g., [11, Chapter 7]).

these structural properties show up in problem inputs. The framework of parameterized complexity theory is particularly well suited to analyze the impact of structural properties of the problem input on the computational complexity of the model checking problem for DEL.

6 Conclusion

We extended the investigation of the computational complexity of the model checking problem for DEL by providing a detailed computational complexity analysis of the model checking problem for various (previously uninvestigated) combinations of restrictions on the DEL model. In particular, we studied various restrictions of the problem where all models are S5, including bounds on the number of agents, allowing only single-pointed models, allowing no postconditions, and allowing semi-private announcements rather than updates with arbitrary event models. We showed that the problem is already PSPACE-hard for very restricted settings.

Future research includes extending the computational complexity analysis to additional restricted settings. For instance, it would be interesting to see whether the polynomial-time algorithm for Proposition 1 can be extended to the setting where the models contain only relations that are transitive, Euclidean and serial (KD45 models). Additionally, it would be interesting to further investigate the contribution of various parameters of the problem input to the computational costs required to solve the problem—continuing an endeavor that was recently initiated [29, 30]. In the setting of KD45 models, it would also be interesting to investigate the complexity of the problem for the case where all updates are private announcements (i.e., a public announcement to a subset of agents, where the remaining agents have no awareness that any action has taken place). Moreover, future research includes obtaining upper bounds for the case where we only found lower bounds (i.e., for the case of one agent, a single-pointed models, and single-pointed event models with postconditions, where we showed Δ_2^{p}-hardness).

References

[1] Sanjeev Arora and Boaz Barak. *Computational Complexity – A Modern Approach.* Cambridge University Press, 2009.

[2] Guillaume Aucher and François Schwarzentruber. On the complexity of dynamic epistemic logic. In *Proceedings of the 14th conference on Theoretical Aspects of Rationality and Knowledge (TARK 2013)*, Chennai, India, 2013.

[3] Alexandru Baltag and Lawrence Moss. Logics for epistemic programs. *Synthese*, 139(2):165–224, 2004.

[4] Alexandru Baltag, Lawrence Moss, and Slawomir Solecki. The logic of public announcements, common knowledge, and private suspicions. In *Proceedings of the 7th conference on Theoretical Aspects of Rationality and Knowledge (TARK 1998)*, pages 43–56. Morgan Kaufmann Publishers Inc., 1998.

[5] Alexandru Baltag, Lawrence Moss, and Slawomir Solecki. The logic of public announcements, common knowledge, and private suspicions. Technical report, Indiana University, 1999.

[6] Chitta Baral and Yan Zhang. Knowledge updates: Semantics and complexity issues. *Artificial Intelligence*, 164(1-2):209–243, 2005.

[7] Johan van Benthem. Epistemic logic and epistemology: The state of their affairs. *Philosophical Studies*, 128(1):49–76, 2006.

[8] Johan van Benthem, Jan van Eijck, and Barteld Kooi. Logics of communication and change. *Information and computation*, 204(11):1620–1662, 2006.

[9] Johan van Benthem, Jan van Eijck, Malvin Gattinger, and Kaile Su. Symbolic model checking for dynamic epistemic logic – S5 and beyond. *Journal of Logic and Computation*, 28(2):367–402, 2018.

[10] Thomas Bolander and Mikkel Birkegaard Andersen. Epistemic planning for single and multi-agent systems. *Journal of Applied Non-Classical Logics*, 21(1):9–34, 2011.

[11] Marek Cygan, Fedor V Fomin, Lukasz Kowalik, Daniel Lokshtanov, Dániel Marx, Marcin Pilipczuk, Michal Pilipczuk, and Saket Saurabh. *Parameterized Algorithms*. Springer, 2015.

[12] Hans van Ditmarsch and Barteld Kooi. Semantic results for ontic and epistemic change. In Giacomo Bonanno, Wiebe van der Hoek, and Michael Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, pages 87–118. 2006.

[13] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Pieter Kooi. *Dynamic epistemic logic*, volume 337 of *Synthese library*. Springer, 2008.

[14] Rodney G. Downey and Michael R. Fellows. *Fundamentals of Parameterized Complexity*. Texts in Computer Science. Springer Verlag, 2013.

[15] Jan van Eijck. DEMO – A demo of epistemic modelling. In Johan van Benthem, Dov Gabbay, and Benedikt Löwe, editors, *Interactive Logic – Proceedings of the 7th Augustus de Morgan Workshop*, volume 1, pages 303–362, 2007.

[16] Jan van Eijck and François Schwarzentruber. Epistemic probability logic simplified. In *Advances in Modal Logic*, pages 158–177, 2014.

[17] Massimo Narizzano Enrico Giunchiglia, Paolo Marin. Reasoning with quantified Boolean formulas. In Armin Biere, Marijn Heule, Hans van Maaren, and Toby Walsh, editors, *Handbook of Satisfiability*, pages 761–780. IOS Press, 2009.

[18] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, 1995.

[19] Jörg Flum and Martin Grohe. Describing parameterized complexity classes. *Information and Computation*, 187(2):291–319, 2003.

[20] Jörg Flum and Martin Grohe. *Parameterized Complexity Theory*, volume XIV of *Texts*

in Theoretical Computer Science. Springer Verlag, Berlin, 2006.

[21] Malvin Gattinger. SMCDEL — an implementation of symbolic model checking for dynamic epistemic logic with binary decision diagrams, 2018. Version 1.0.0.

[22] Wiebe van der Hoek and Michael Wooldridge. Multi-agent systems. In *Handbook of Knowledge Representation*, volume 3 of *Foundations of Artificial Intelligence*, pages 887–928. Elsevier, 2008.

[23] Barteld Kooi and Johan van Benthem. Reduction axioms for epistemic actions. In R. Schmidt, I. Pratt-Hartmann, M. Reynolds, and H. Wansing, editors, *AiML-2004: Advances in Modal Logic, Department of Computer Science, University of Manchester, Technical report series, UMCS-04-9-1*, pages 197–211, 2004.

[24] Mark W. Krentel. Generalizations of OptP to the Polynomial Hierarchy. *Theoretical Computer Science*, 97(2):183–198, 1992.

[25] Carsten Lutz. Complexity and succinctness of public announcement logic. In Hideyuki Nakashima, Michael P. Wellman, Gerhard Weiss, and Peter Stone, editors, *Proceedings of the fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 137–143. IFAAMAS, 2006.

[26] John-Jules Meyer and Wiebe van der Hoek. Epistemic logic for computer science and artificial intelligence, 1995.

[27] Jan Plaza. Logics of public communications. In M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, and Z.W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems: poster session program*, pages 201–216. Oak Ridge National Laboratory, 1989. (Reprinted as [28]).

[28] Jan Plaza. Logics of public communications. *Synthese*, 158(2):165–179, 2007.

[29] Iris van de Pol. How Difficult is it to Think that you Think that I Think that ...? *A DEL-based Computational-level Model of Theory of Mind and its Complexity.* Master's thesis, University of Amsterdam, the Netherlands, 2015.

[30] Iris van de Pol, Iris van Rooij, and Jakub Szymanik. Parameterized complexity of theory of mind reasoning in dynamic epistemic logic. *Journal of Logic, Language and Information*, 2018.

[31] Rineke Verbrugge. Logic and social cognition: The facts matters, and so do computational models. *Journal of Philosophical Logic*, 38(6):649–680, 2009.

 Received 6 June 2018

On the Role of Aristotle's Connexive Axioms in Non-connexive Logics

Nissim Francez
Computer Science dept., the Technion-IIT, Haifa, Israel.
francez@cs.technion.ac.il

Abstract

In this paper, I investigate the effect of Aristotle's connexive axioms in non-connexive logics. I define two properties of formula schemes $\Phi(\varphi)$:

1. *pos*-reflection: $\Phi(\varphi) \equiv \varphi$
2. *neg*-reflection: $\Phi(\varphi) \equiv \neg\varphi$

I then investigate several logics for the reflectivity of $\Phi_{A_1}(\varphi) = \neg(\varphi\rightarrow\neg\varphi)$, and $\Phi_{A_2}(\varphi) = \neg(\neg\varphi\rightarrow\varphi)$.

1 Introduction

The characteristic axioms[1] of *connexive logics* (see [17] for a general survey) are:

Aristotle's axioms:

$$A_1: \ \neg(\varphi\rightarrow\neg\varphi) \tag{1.1}$$

$$A_2: \ \neg(\neg\varphi\rightarrow\varphi) \tag{1.2}$$

Boethius' axioms:

$$B_1: \ (\varphi\rightarrow\psi)\rightarrow\neg(\varphi\rightarrow\neg\psi) \tag{1.3}$$

$$B_2: \ (\varphi\rightarrow\neg\psi)\rightarrow\neg(\varphi\rightarrow\psi) \tag{1.4}$$

Those axioms are *not* classical logic validities, but have a strong intuitive appeal, reflecting properties of a "natural" conditional, closer to the indicative conditional used in natural language.

There are several indications for the appeal of the connexive axioms.

[1] Actually, those are axiom scemes. I use 'axioms' for brevity.

- The growing number of recent publications on various aspects of connexive logics. See the connexive logic site[2].
- The growing number of conferences/workshops dedicated to connexive logics.
- Empirical studies show favourable results for the connexive axioms serving as part of the interpretation of conditionals, enhancing their view as "natural".
    1. In [8] (Section 2), McCall reports the results of an empirical study about truth of conditionals formulated in natural language. The questionnaire included randomly spread instances of connexive conditionals, popularly judged as true.
    2. In [13], Pfeifer reports an empirical study connecting the connexive axioms to probabilistic reasoning.
- In a recent manuscript [3], Crupi and Iacona present a conditional, referred to as 'evidential conditional', that also provides a natural reading of the connexive axioms as indicating a support of the consequent by the antecedent.

Alas, not all logics are connexive ...

In this paper, I am interested in studying an effect that the A-schemes have when they are not logical validities, but are holding (for *some* φs), in non-connexive logics. I refer to this effect as *reflectivity* (defined in Section 2).

Note that reflectivity (in its various guises) is not presented necessarily as a *desirable* feature of a non-connexive logic. Uncovering the reflectivity effect may contribute to the study of connexive logics, for example by relating the latter to some recent trends of *restricting the scope* of those axioms, such as to formulas having some modal strength (cf. [6]). Understanding the reflectivity effect may help in understanding what are we "really" imposing on a connexive logic by admitting the connexive characteristics as axioms.

2 Reflecting formula schemes

Let $\mathcal{L}$ be a logic with an object-language L, with $\{`\neg', `\rightarrow'\} \subseteq L$. A *unary formula scheme* $\Phi(\varphi)$ is a formula schematically constructed over a base schematic formula φ; for example, $\varphi\rightarrow\neg\varphi$.

Suppose there is some model-theory for $\mathcal{L}$, defining satisfaction of an L-formula φ in a model $\mathcal{M}$, denoted by $\mathcal{M}\models\varphi$.

[2]https://sites.google.com/site/connexivelogic/

Definition 2.1 (equivalence)**.** *Two L-formulas φ and ψ are* $\mathcal{L}$-equivalent*, denoted by $\varphi \equiv_{\mathcal{L}} \psi$, iff for every model $\mathcal{M}$ for $\mathcal{L}$:*

$$\mathcal{M} \models \varphi \text{ iff } \mathcal{M} \models \psi$$

I assume that $\mathcal{L}$-equivalence is a congruence over the other connectives in the object-language. This assumption is used implicitly in all proofs. For example, without this assumption, Theorem 4.1 does not hold.

Definition 2.2 (reflective formula schemes)**.** *A unary formula scheme $\Phi(\varphi) \in L$, is:*

- positively-reflective (*pos*-reflective) for $\mathcal{L}$ *iff for every $\varphi \in L$*

$$\Phi(\varphi) \equiv_{\mathcal{L}} \varphi \tag{2.5}$$

- negatively-reflective (*neg*-reflective) for $\mathcal{L}$ *iff for every $\varphi \in L$*

$$\Phi(\varphi) \equiv_{\mathcal{L}} \neg\varphi \tag{2.6}$$

Definition 2.3 (involution)**.** *$\mathcal{L}$ is* involutive *iff the unary scheme $\Phi_{\neg\neg}(\varphi) =^{df.} \neg\neg\varphi$ is pos-reflective (i.e., $\neg\neg\varphi \equiv_{\mathcal{L}} \varphi$).*

The question I am interested in can now be more precisely formulated: what are the reflective capabilities of the unary formula schemes

$$\Phi_{A_1}(\varphi) =^{df.} \neg(\varphi \to \neg\varphi) \qquad \Phi_{A_2}(\varphi) =^{df.} \neg(\neg\varphi \to \varphi) \tag{2.7}$$

based on the A-axioms, respectively, in non-connexive logics?

I start with some case studies, and then move to a general statement.

3 Case studies

3.1 Reflection in classical propositional logic

In this section I consider reflectivity of the connexive A-schemes in (the obviously non-connexive) propositional classical logic. Here '$\to$' is taken as the material implication '$\supset$' in both schemes.

Classical logic is, of course, involutive.

An easy inspection of the standard classical truth-tables for negation and material implication establishes the following immediate consequence.

Proposition 3.1 (classical reflection).

- $\Phi_{A_1}(\varphi)$ *is pos-reflective for classical logic.*
- $\Phi_{A_2}(\varphi)$ *is neg-reflective for classical logic.*

That is, a classical valuation satisfies A_1 iff it satisfies φ itself, and it satisfies A_2 iff it does not satisfy φ itself.

3.2 FDE

In this section I consider reflectivity of the connexive A-schemes in the (non-connexive) Belnap-Dunn 4-valued logic of *first-degree entailment (FDE)* [1, 2, 4], over the signature $\{\neg, \wedge, \vee\}$, with implication defined by

$$\varphi{\rightarrow}\psi =^{df.} \neg\varphi\vee\psi \tag{3.8}$$

while this implication is notoriously "bad[3]" (see, for example, [10], Section 5.1), I include it because, as shown below, the reflectivity induced by it resembles that of classical logic. However, as argued in Section 3.2.3, this similarity in reflectivity to classical logic is an indication for the "badness" of FDE's conditional.

3.2.1 Defining FDE

The logic is based on four truth-values, that can be (or often are) represented as the power-set of the classical truth-values $\{t, f\}$:

- t ($\{t\}$: true only)
- b ($\{t, f\}$: both true and false)
- n ($\emptyset$: neither true nor false)
- f ($\{f\}$: false only)

The truth-tables of the basic connective (cf. [15], p.144 or [5], p. 60) are presented in Figure 1. The truth-table of the defined '$\rightarrow$' is presented in Figure 2.

Models are again valuations, mapping atomic propositions to truth-values, extended to arbitrary formulas by respecting the truth-tables.

[3]For example, invalidation of modus ponens.

$$
\begin{array}{c|c}
\neg & \\
-- & -- \\
t & f \\
b & b \\
n & n \\
f & t
\end{array}
\qquad
\begin{array}{c|c|c|c|c}
\wedge & t & b & n & f \\
-- & -- & -- & -- & -- \\
t & t & b & n & f \\
b & b & b & f & f \\
n & n & f & n & f \\
f & f & f & f & f
\end{array}
\qquad
\begin{array}{c|c|c|c|c}
\vee & t & b & n & f \\
-- & -- & -- & -- & -- \\
t & t & t & t & t \\
b & t & b & t & b \\
n & t & t & n & n \\
f & t & b & n & f
\end{array}
\tag{3.9}
$$

Figure 1: The four-valued truth-tables for FDE

$$
\begin{array}{c|c|c|c|c}
\to & t & b & n & f \\
-- & -- & -- & -- & -- \\
t & t & b & n & f \\
b & t & b & t & b \\
n & t & t & n & n \\
f & t & t & t & t
\end{array}
\tag{3.10}
$$

Figure 2: The four-valued truth-table for FDE's '$\to$'

3.2.2 Reflection in FDE

Again, an easy inspection of the truth-tables establishes that FDE is involutive, and:

Proposition 3.2 (FDE-reflection)**.**

- $\Phi_{A_1}(\varphi)$ *is pos-reflective for* FDE.
- $\Phi_{A_2}(\varphi)$ *is neg-reflective for* FDE.

Proof. The reasoning concerning t and f is like that in the case of classical logic, and the reasoning regarding b is similar to that regarding n. So, I only show the latter.

Φ_{A_1}: Consider any valuation v.

1. Suppose $v[\![\varphi]\!] = n$. Then, by the truth-table for negation, $v[\![\neg\varphi]\!] = n$, and by the truth-table for implication $v[\![\varphi\to\neg\varphi]\!] = n$. Finally, again by the truth-table for negation, $v[\![\neg(\varphi\to\neg\varphi)]\!] = n$.

2. Suppose $v[\![\neg(\varphi \to \neg\varphi)]\!] = n$. Then, by the truth-table for negation, $v[\![\varphi \to \neg\varphi]\!] = n$. By the truth-table of implication, the value n can result for three pairs of truth-values: $(t, n), (n, f), (n, n)$. However, by the truth-table of negation, the first two cannot be attributed to $(\varphi, \neg\varphi)$, hence $v[\![\varphi]\!] = n$.

Thus, *pos*-reflectivity of Φ_{A_1} for FDE is established.

Φ_{A_2}: Consider any valuation v.

1. Suppose $v[\![\varphi]\!] = n$. Then, by the truth-table for negation, $v[\![\neg\varphi]\!] = n$, and by the truth-table for implication $v[\![\neg\varphi \to \varphi]\!] = n$. Finally, again by the truth-table for negation, $v[\![\neg(\neg\varphi \to \varphi)]\!] = n$.
2. Suppose $v[\![\neg(\neg\varphi \to \varphi)]\!] = n$. Then, by the truth-table for negation, $v[\![\neg\varphi \to \varphi]\!] = n$. Again, by the truth-table of implication, the value n can result for three pairs of truth-values: $(t, n), (n, f), (n, n)$. However, by the truth-table of negation, the first two cannot be attributed to $(\varphi, \neg\varphi)$, hence $v[\![\neg\varphi]\!] = n$.

Thus, *neg*-reflectivity of Φ_{A_2} for FDE is established.

$\square$

Consequently, the A-schemes play in FDE (with implication defined "classically") a role similar to the role they play in propositional classical logic.

3.2.3 A "useful" conditional for FDE

As mentioned above, it is well-known that the conditional '$\to$' defined above is a "bad" conditional for FDE. In [16], some guiding lines for a "good" FDE-conditional are suggested. The "good" conditional chosen, also preferred in several other studies of implication in FDE, is known as the '$\to_{cmi}$' ('cmi' stands for 'classical material implication'), with the truth-table below.

$\to_{cmi}$	t	b	n	f
t	t	b	n	f
b	t	b	n	f
n	t	t	t	t
f	t	t	t	t

(3.11)

$$\begin{array}{c|c} \neg & \\ -- & -- \\ t & f \\ n & n \\ f & t \end{array} \qquad \begin{array}{c|c|c|c} \rightarrow & t & n & f \\ -- & -- & -- & -- \\ t & t & n & f \\ n & t & t & n \\ f & t & t & t \end{array} \tag{3.12}$$

Figure 3: Truth-tables for Łukasiewicz's three-valued logic Ł_3

Consequently, the A_1-scheme, when based on the "useful" conditional '$\rightarrow_{cmi}$' *is not pos-reflective*. For a valuation v with $v[\![\varphi]\!] = n$, the truth-value of A_1 has to be t, contradicting[4] the *pos*-reflectivity imposed truth-value n.

A similar lack of *pos*-reflectivity of A_1 holds also for several other conditionals suggested for FDE (cf. [11], [12]).

3.3 Łukasiewicz's three-valued logic Ł_3

In this section, I inspect Łukasiewicz's three-valued logic Ł_3 [7] for reflectivity of the connexive A-schemes. I consider the negation-implication fragment of Ł_3, defined by the following truth-tables (Figure 3) over the truth-value $\{t,\ n,\ f\}$. Clearly, Ł_3 is involutive.

Proposition 3.3 (Ł_3-reflection)**.**

- $\Phi_{A_1}(\varphi)$ *is not pos-reflective for* Ł_3*.*
- $\Phi_{A_2}(\varphi)$ *is not neg-reflective for* Ł_3*.*

Proof. Consider a valuation v and a formula φ s.t. $v[\![\varphi]\!] = n$.

1. By the truth-tables for Ł_3,
   $v[\![\neg(\varphi\rightarrow\neg\varphi)]\!] = f \neq n$. Thus, Φ_{A_1} is not *pos*-reflective.
2. By the truth-tables for Ł_3,
   $v[\![\neg(\neg\varphi\rightarrow\varphi)]\!] = f \neq n(= v[\![\neg\varphi]\!])$. Thus, Φ_{A_2} is not *neg*-reflective.

□

In contrast to propositional classical logic and FDE, where the connexive A-schemes are reflective, both A-schemes *are not* reflective for Ł_3.

[4] Unless the definition of negation is varied too ...

3.4 Intuitionistic logic

I now turn to the question of reflectiveness of the connexive A-schemes in the notoriously *non-involutive* propositional intuitionistic logic IL (see, for example, [9]).

Proposition 3.4 (IL-reflection)**.**

- $\Phi_{A_1}(\varphi)$ *is* not *pos-reflective for* IL*.*
- $\Phi_{A_2}(\varphi)$ is *neg-reflective for* IL*.*

Proof. The proof of both clauses relies on the well-known consequence of Glivenko's theorem on double-negation translation, the consequence stating that a negated formula $\neg\varphi$ is intuitionistically equivalent to another negated formula $\neg\psi$ iff the two negated formulas are classically equivalent.

- The negated formula $\Phi_{A_1}(\varphi)$ is classically, and hence intuitionistically, equivalent to $\neg\neg\varphi$. However, by non-involutiveness, $\neg\neg\varphi$ is not intuitionistically equivalent to φ.

  Hence, *pos*-reflectiveness of Φ_{A_1} fails for IL.

- The negated formula $\Phi_{A_2}(\varphi)$ is classically, and hence intuitionistically, equivalent to $\neg\varphi$, thereby establishing *neg*-reflectiveness of Φ_{A_2} holds for IL.

$\square$

3.5 Post logics with cyclic negation

Post logics, denoted $\mathbf{P_n}$, were introduced by Post [14]. The object-language contains '$\wedge$' (conjunction[5]), '$\vee$' (disjunction) and '$\neg_p$' (negation). It is convenient to take here $\mathcal{V} = \{v_0, \cdots, v_{n-1}\}$, ordered by $v_i \leq v_j$ whenever $i \leq j$. The main interest in these logics is due to the definition of negation, which is *cyclical*. The truth-table of negation is the following.

$$\begin{array}{c|c} \varphi & \neg_p\varphi \\ --- & --- \\ v_0 & v_1 \\ v_1 & v_2 \\ \cdots & \cdots \\ v_{n-1} & v_0 \end{array} \tag{3.13}$$

[5]Post's original definition in [14] considered conjunction as defined by disjunction and negation. Taking it as primitive is a matter of convenience only.

Conjunction and disjunction are defined as, respectively, minimum and maximum.

$$v[\![\varphi\wedge\psi]\!] = min\{v[\![\varphi]\!], v[\![\psi]\!]\} \qquad v[\![\varphi\vee\psi]\!] = max\{v[\![\varphi]\!], v[\![\psi]\!]\} \tag{3.14}$$

Here, too, define the conditional as $\varphi\rightarrow\psi =^{df.} \neg_p\varphi\vee\psi$.

Clearly, $\mathbf{P}_n$ is *not* involutive for $n > 2$.

Proposition 3.5 ($\mathbf{P}_n$-non-reflectiveness)**.** *For $n > 2$:*

1. *$\Phi_{A_1}(\varphi)$ is not pos-reflective for $\mathbf{P}_n$.*
2. *$\Phi_{A_2}(\varphi)$ is not neg-reflective for $\mathbf{P}_n$.*

Proof. 1. It is easily verified that $\neg_p(\varphi\rightarrow\neg_p\varphi) \equiv_{P_n} \neg_p\neg_p\varphi$. However, because of the cyclicality of '$\neg_p$', $\neg_p\neg_p\varphi \not\equiv_{P_n} \varphi$ for $n > 2$.

2. Similarly, $\neg(\neg\varphi\rightarrow\varphi) \equiv_{P_n} \neg(\neg\neg\varphi\vee\varphi)$. But, because of the cyclicality of '$\neg_p$', $\neg_p(\neg_p\neg_p\varphi\vee\varphi) \not\equiv_{P_n} \neg_p\varphi$ for $n > 2$.

□

4 Connexive reflection

I now turn to generalising the reflectivity of the connexive A-schemes in the above special cases.

Theorem 4.1 (reflection)**.** *For any involutive $\mathcal{L}$, $\Phi_{A_1}(\varphi)$ is pos-reflective for $\mathcal{L}$ iff $\Phi_{A_2}(\varphi)$ is neg-reflective for $\mathcal{L}$.*

Proof: let $\mathcal{L}$ be involutive.

1. Assume $\Phi_{A_1}(\varphi)$ is *pos*-reflective for $\mathcal{L}$. Substitute $\neg\varphi$ for φ in Φ_{A_1}. We get

$$\neg\varphi \equiv_{\mathcal{L}}^{\text{ass.}} \Phi_{A_1}(\neg\varphi) = \neg(\neg\varphi\rightarrow\neg\neg\varphi) \equiv_{\mathcal{L}}^{involutiveness} \neg(\neg\varphi\rightarrow\varphi) = \Phi_{A_2}(\varphi)$$

Thus, Φ_{A_2} is *neg*-reflective for $\mathcal{L}$.

2. Assume $\Phi_{A_2}(\varphi)$ is *neg*-reflective for $\mathcal{L}$. Substitute $\neg\varphi$ for φ in Φ_{A_2}. We get

$$\neg\neg\varphi \equiv_{\mathcal{L}}^{\text{ass.}} \neg(\neg\neg\varphi\rightarrow\neg\varphi)$$

and by involutiveness

$$\varphi \equiv_{\mathcal{L}} \neg(\varphi\rightarrow\neg\varphi) = \Phi_{A_1}(\varphi)$$

Thus, Φ_{A_1} is *pos*-reflective for $\mathcal{L}$.

Indeed, the special cases studied above all conform to Theorem 4.1:

- Classical propositional logic and FDE are both involutive, and for both the two connexive A-schemes are reflective.
- Łukasiewicz's three-valued logic Ł$_3$ is involutive, and both of the connexive A-schemes are not reflective for it.
- Intuitionistic propositional logic is not involutive. The scheme Φ_{A_1} fails to be *pos*-reflective for IL, while the scheme Φ_{A_2} *is neg*-reflective for IL.
- Post logics $\mathbf{P}_n$, $n > 2$, are not involutive, nor are the connexive A-schemes reflective for them.

5 Conclusion

The question posed, and partially answered, in this paper is: what is the effect of the formula schemes based on Aristotle's axioms of connexive logics *when not imposed as axioms*?

As far as I am aware of, this question was not raised before in the literature.

A partial characteristic obtained relates to the property of reflectivity: being equivalent to the schematic argument of the scheme (positive reflection) or to its negation (negative reflection).

Theorem 4.1 reveals that when a logic is involutive (i.e., respects double-negation equivalence), either both schemes are reflective (one positively, the other negatively), or none of them is.

Some natural extensions of this work include:

- Find some regularity about the reflectivity for non-involutive logics.
- Find some general conditions (on negation and conditionals) that guarantee reflectivity (or the lack thereof).
- Investigate analogous effects of binary connexive schemes based on B_1 and B_2. This task requires some possible modification of the definition of reflectivity, accounting for the presence of 'ψ'-the second metavariable in the B-schemes. I still have no clear opinion on this.
- Extend the question to schemes based on axioms of other contra-classical logics.

References

[1] Nuel D. Belnap. How a computer should think. In Gilbert Ryle, editor, *Contemporary aspects of philosophy*, pages 30–56. Stocksfield:Oriel Press, 1976.

[2] Nuel D. Belnap. A useful four-valued logic. In J. Michael Dunn and George Epstein, editors, *Modern uses of multiple-valued logic*, pages 8–37. Dordrecht:Reidl, 1977.

[3] Vincenzo Crupi and Andrea Iacona. The evidential conditional. 2019. Unpublished, available at: http://philsci-archive.pitt.edu/16757/13/TEC-archNEW.pdf.

[4] J. Michael Dunn. Intuitive semantics for first-degree entailments and Ôcoupled treesÕ. *Philosophical Studies*, 29:149–168, 1976.

[5] Andreas Kapsner. *Logics and Falsification: a new perspective on constructive semantics.* Springer, Cham Heidelberg New York Dordrecht London, 20014. Trends in Logic 40.

[6] Andreas Kapsner. Humble connexivity. *Logic and logical philosophy*, 28:513–536, 2019. Special Issue: Advances in Connexive Logic. Guest Editors: Hitoshi Omori and Heinrich Wansing, DOI: http://dx.doi.org/10.12775/LLP.2019.001.

[7] Jan Łukasiewicz. On three-valued logic. *Ruch Filozoficzny*, 5:170–171, 1920. In Polish. English translation in: Jan Ł ukasiewicz. Selected Works, edited by L. Borkowski (North- Holland, Amsterdam, 1970).

[8] Storrs McCall. A history of connexivity. In Dov M Gabbay, Francis J. Pelletier, and John Woods, editors, *Handbook of the history of logic, Volume 11: Logic: a history of its central concepts*, pages 415–449. Elsevier, Amsterdam, 2012.

[9] Joan Moschovakis. Intuitionistic logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, winter 2018 edition, 2018.

[10] Hitoshi Omori and Heinrich Wansing. 40 years of FDE: An introductory overview. *Studia Logica*, 105:1021–1049, 2017. Special issue.

[11] Francis J. Pelletier and Allen P. Hazen. K3, Ł3, LP, RM3, A3, FDE, M: How to make many-valued logics work for you. In Hitoshi Omori and Heinrich Wansing, editors, *New Essays on Belnap–Dunn Logic.* Springer, 2020 (forthcoming). Collects philosophical essays on Belnap-Dunn logic by leading scholars.

[12] Yaroslav Petrukhin. Correspondence analysis for first degree entailment. *Logical Investigations*, 22(1):108–124, 2016.

[13] Niki Pfeifer. Experiments on aristotle's thesis: Towards an experimental philosophy of conditionals. *The Monist*, 95(2):223–240, 2012.

[14] Emil Post. Introduction to a general theory of elementary propositions. *American Journal of Mathematics*, 43(3):163–185, 1921.

[15] Graham Priest. *Non-Classical Logic.* Cambridge University Press, 2001.

[16] Geoff Sutcliffe, Francis J. Pelletier, and Allen P. Hazen. Making belnapÕs "useful four-valued logic" useful. In Keith Brawner and Vasile Rus, editors, *Proceedings of the 31st International FLAIRS Conference, Palo Alto*, pages 116–121. AAAI Press, 2018.

[17] Heinrich Wansing. Connexive logic. In Edward N. Zalta, edi-

tor, *The Stanford Encyclopedia of Philosophy*. Fall 2020 edition, 2020. https://plato.stanford.edu/archives/spr2020/entries/logic-connexive/.

Received 4 December 2019

About the Unification Type of Fusions of Modal Logics

Philippe Balbiani*
Toulouse Institute of Computer Science Research
CNRS — Toulouse University, Toulouse, France

Çiğdem Gencer†
Faculty of Arts and Sciences
Istanbul Aydın University, Istanbul, Turkey

Maryam Rostamigiv‡
Toulouse Institute of Computer Science Research
CNRS — Toulouse University, Toulouse, France

Abstract

In a modal logic $\mathbf{L}$, a unifier of a formula φ is a substitution σ such that $\sigma(\varphi)$ is in $\mathbf{L}$. When unifiable formulas have no minimal complete sets of unifiers, they are nullary. Otherwise, they are either infinitary, or finitary, or unitary depending on the cardinality of their minimal complete sets of unifiers. The fusion $\mathbf{L}_1 \otimes \mathbf{L}_2$ of modal logics $\mathbf{L}_1$ and $\mathbf{L}_2$ respectively based on the modal connectives $\Box_1$ and $\Box_2$ is the least modal logic based on these modal connectives and containing both $\mathbf{L}_1$ and $\mathbf{L}_2$. In this paper, we prove that if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is unitary then $\mathbf{L}_1$ and $\mathbf{L}_2$ are unitary and if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is finitary then $\mathbf{L}_1$ and $\mathbf{L}_2$ are either unitary, or finitary. We also prove that the fusion of arbitrary consistent extensions of $\mathbf{S}5$ is nullary when these extensions are different from $\mathbf{Triv}$[1].

The preparation of this paper has been launched on the occasion of a visit of Çiğdem Gencer during the Fall 2018 in Toulouse supported by *Université Paul Sabatier* (Programme *Professeurs invités 2018*). Special acknowledgement is heartily granted to the colleagues of the *Toulouse Institute of Computer Science Research* for their valuable remarks. We are also greatly indebted to Christophe Ringeissen (Lorraine Research Laboratory in Computer Science and its Applications, Nancy, France) and Tinko Tinchev (Sofia University St. Kliment Ohridski, Sofia, Bulgaria) for their helpful suggestions.

*Email address: philippe.balbiani@irit.fr.

†Email addresses: cigdemgencer@aydin.edu.tr and cigdem.gencer@irit.fr, Çiğdem Gencer being also associate researcher at Toulouse Institute of Computer Science Research.

‡Email address: maryam.rostamigiv@irit.fr.

[1]Dzik conjectured that the fusion $\mathbf{S}5 \otimes \mathbf{S}5$ of $\mathbf{S}5$ with itself is either nullary, or infinitary [11, Chapter 6].

1 Introduction

The unification problem in a modal logic $\mathbf{L}$ is to determine, given a formula φ, whether there exists a substitution σ such that $\sigma(\varphi)$ is in $\mathbf{L}$ [1][2]. In that case, σ is a unifier of φ. We shall say that a set of unifiers of a unifiable formula φ is complete if for all unifiers σ of φ, there exists a unifier τ of φ in that set such that τ is more general than σ. When unifiable formulas have no minimal complete sets of unifiers, they are nullary. Otherwise, they are either infinitary, or finitary, or unitary depending on the cardinality of their minimal complete sets of unifiers [11]. To be nullary is considered to be the worst situation for a unifiable formula whereas to be unitary is considered to be better than to be finitary which is itself considered to be better than to be infinitary. The unification type of a modal logic is the worst unification type of its unifiable formulas[3].

The fusion $\mathbf{L}_1 \otimes \mathbf{L}_2$ of modal logics $\mathbf{L}_1$ and $\mathbf{L}_2$ respectively based on the modal connectives $\Box_1$ and $\Box_2$ is the least modal logic based on these modal connectives and containing both $\mathbf{L}_1$ and $\mathbf{L}_2$. A first immediate result is that $\mathbf{L}_1 \otimes \mathbf{L}_2$ is a conservative extension of the modal logics $\mathbf{L}_1$ and $\mathbf{L}_2$ when $\mathbf{L}_1$ and $\mathbf{L}_2$ are consistent. A number of other results — transfer results — have been obtained as well. They concern properties preserved under the operation of forming fusions: the fusion of decidable modal logics is decidable, the fusion of modal logics having uniform interpolation property has uniform interpolation property, etc. See [15, Chapter 4] and [23, 24, 31]. To the best of our knowledge, the preservation of properties related to the unification problem has not been studied yet.

Owing to its strong connections with the admissibility problem [29], the unification problem is an important problem in Applied Non-Classical Logics [1], a domain of investigations where fusions of modal logics are omnipresent [24]. It is therefore nat-

[2]We assume that the reader is at home with tools and techniques in modal logics. In particular, we follow the standard conventions for talking about modal logics: $\mathbf{S}5$ is the least modal logic containing the formulas usually denoted ($\mathbf{T}$), (4) and ($\mathbf{B}$), $\mathbf{KT}$ is the least modal logic containing the formula usually denoted ($\mathbf{T}$), etc. For more on this, see [7, 8, 22]. As a result, in the main body of the paper, we will present neither the algebraic semantics of modal logics, nor the relational semantics of modal logics, prefering to introduce semantic tools and techniques when they are needed.

[3]About the unification type of modal logics, it is known that $\mathbf{KB}$, $\mathbf{KD}$, $\mathbf{KDB}$, $\mathbf{KT}$ and $\mathbf{KTB}$ are nullary [3, 4, 5], $\mathbf{S}5$ and $\mathbf{S}4.3$ are unitary [10, 12], some transitive modal logics like $\mathbf{K}4$ and $\mathbf{S}4$ are finitary [18, 19], $\mathbf{K}$ is nullary [20] and $\mathbf{K}4\mathbf{D}1$ is unitary [21], the nullariness of $\mathbf{KB}$, $\mathbf{KD}$, $\mathbf{KDB}$, $\mathbf{KT}$ and $\mathbf{KTB}$ having only been obtained within the context of unification with parameters. No modal logic is known to be infinitary.

ural to ask how the unification types of modal logics are related to the unification type of their fusion. In this paper, we prove that if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is unitary then $\mathbf{L}_1$ and $\mathbf{L}_2$ are unitary and if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is finitary then $\mathbf{L}_1$ and $\mathbf{L}_2$ are either unitary, or finitary. In other respects, Dzik conjectured that the fusion $\mathbf{S}5 \otimes \mathbf{S}5$ of $\mathbf{S}5$ with itself is either nullary, or infinitary [11, Chapter 6]. Clarifying Dzik's conjecture, we prove that the fusion of arbitrary consistent extensions of $\mathbf{S}5$ is nullary when these extensions are different from **Triv**. An Appendix includes the proofs of some of our results.

Jeřábek has proved that $\mathbf{K}$ is nullary by showing that the $\mathbf{K}$-unifiable formula $x \to \Box x$ has no minimal complete sets of unifiers [20]. In Jeřábek's line of reasoning, the fact that for all $d \geq 0$, $\Box^{d+1}\bot \to \Box^{d}\bot \notin \mathbf{K}$ plays an important role. Unfortunately, for all $d \geq 0$, $\Box^{d}\bot$ is either equivalent to $\bot$, or equivalent to $\Box\bot$ in **KB**, **KD**, **KDB**, **KT** and **KTB**. It follows that Jeřábek's line of reasoning has to be seriously adapted if one wants to apply it to **KB**, **KD**, **KDB**, **KT** and **KTB**. This has been done in [3, 4, 5] by using parameters and by considering much more complicated formulas than $x \to \Box x$. For the fusion of arbitrary consistent extensions of $\mathbf{S}5$ different from **Triv**, a new adaptation of Jeřábek's line of reasoning is described in the course of Lemmas 14–29 and Propositions 7, 8 and 9.

2 Syntax

2.1 Formulas and substitutions

Let VAR be a countably infinite set of *propositional variables* (with typical members denoted x, y, etc). Let PAR be a countably infinite set of *propositional parameters* (with typical members denoted p, q, etc). *Atoms* (denoted α, β, etc) are either variables, or parameters. Let I be a non-empty subset of $\{1, 2\}$. The set FOR_I of *I-formulas* (with typical members denoted φ, ψ, etc) is inductively defined as follows:

- $\varphi, \psi ::= \alpha \mid \bot \mid \neg\varphi \mid (\varphi \vee \psi) \mid \Box_i\varphi$

where i ranges over I. We adopt the standard rules for omission of the parentheses. For all I-formulas φ, we write "φ^0" to mean "$\neg\varphi$" and we write "φ^1" to mean "φ". For all I-formulas φ, let $\mathtt{var}(\varphi)$ be the set of all variables occurring in φ. For all I-formulas φ, the *degree* of φ (denoted $\mathtt{deg}(\varphi)$) is defined as usual. An *I-substitution* is a function σ associating to each variable x an I-formula $\sigma(x)$[4]. We shall say that

[4]Occasionally, we will slightly abuse notation by considering that $\{1\}$-substitutions and $\{2\}$-substitutions are also $\{1, 2\}$-substitutions.

an I-substitution σ *moves* a variable x if $\sigma(x) \neq x$. Following the standard assumption considered in the literature [1], we will always assume that I-substitutions move at most finitely many variables. For all $\{1,2\}$-formulas $\varphi(x_1,\ldots,x_m)$, let $\sigma(\varphi(x_1,\ldots,x_m))$ be the $\{1,2\}$-formula $\varphi(\sigma(x_1),\ldots,\sigma(x_m))$. The *composition* $\sigma \circ \tau$ of the I-substitutions σ and τ is the I-substitution associating to each variable x the I-formula $\tau(\sigma(x))$. Obviously, for all $\{1,2\}$-formulas $\varphi(x_1,\ldots,x_m)$, $(\sigma\circ\tau)(\varphi(x_1,\ldots,x_m))$ is the $\{1,2\}$-formula $\varphi(\tau(\sigma(x_1)),\ldots,\tau(\sigma(x_m)))$.

2.2 Abbreviations and translation functions

The Boolean connectives $\top$, $\wedge$, $\rightarrow$ and $\leftrightarrow$ are defined by the usual abbreviations. For all finite sets X of variables, we will use $\top_X$ as a shorthand for $\bigwedge\{x \vee \top :\ x \in X\}$. As it is traditionally done, in the extreme case when the finite set X of variables is empty, $\top_X$ will be a shorthand for $\top$. The role of the finite set X of variables in the definition of $\top_X$ will become clear in Propositions 3 and 6. Nevertheless, we can already mention that this role is connected to the fact that for all finite sets X of variables, $\top_X$ is a tautology such that $\mathtt{var}(\top_X) = X$. The modal connectives $\Diamond_1$ and $\Diamond_2$ are defined as follows:

- $\Diamond_1\varphi ::= \neg\Box_1\neg\varphi$,
- $\Diamond_2\varphi ::= \neg\Box_2\neg\varphi$.

From now on in this paper,

$$\boxed{\textbf{let } p,q,r \textbf{ be fixed distinct parameters.}}$$

Now, let us define modal connectives that will be useful in Section 6 for proving Proposition 9 saying that the fusion of arbitrary consistent extensions of **S**5 is nullary when these extensions are different from **Triv**. The modal connectives $\boxplus$ and $\boxminus$ are defined as follows:

- $\boxplus\varphi ::= p^1 \wedge q^0 \wedge r^1 \rightarrow \Box_1(p^0 \wedge q^0 \wedge r^0 \rightarrow \Box_2(p^0 \wedge q^0 \wedge r^1 \rightarrow \Box_1(p^0 \wedge q^1 \wedge r^0 \rightarrow \Box_2(p^0 \wedge q^1 \wedge r^1 \rightarrow \Box_1(p^1 \wedge q^0 \wedge r^0 \rightarrow \Box_2(p^1 \wedge q^0 \wedge r^1 \rightarrow \varphi))))))$,
- $\boxminus\varphi ::= p^1 \wedge q^0 \wedge r^1 \rightarrow \Box_2(p^1 \wedge q^0 \wedge r^0 \rightarrow \Box_1(p^0 \wedge q^1 \wedge r^1 \rightarrow \Box_2(p^0 \wedge q^1 \wedge r^0 \rightarrow \Box_1(p^0 \wedge q^0 \wedge r^1 \rightarrow \Box_2(p^0 \wedge q^0 \wedge r^0 \rightarrow \Box_1(p^1 \wedge q^0 \wedge r^1 \rightarrow \varphi))))))$.

For all $k \geq 0$, the modal connectives $\boxplus^k$ and $\boxminus^k$ are inductively defined as follows:

- $\boxplus^0\varphi ::= \varphi$,
- $\boxminus^0\varphi ::= \varphi$,

- $\boxplus^{k+1}\varphi ::= \boxplus \boxplus^{k} \varphi$,
- $\boxminus^{k+1}\varphi ::= \boxminus \boxminus^{k} \varphi$.

For all $k \geq 0$, the modal connectives $\boxplus^{<k}$ and $\boxminus^{<k}$ are inductively defined as follows:

- $\boxplus^{<0}\varphi ::= \top$,
- $\boxminus^{<0}\varphi ::= \top$,
- $\boxplus^{<k+1}\varphi ::= \boxplus^{<k}\varphi \wedge \boxplus^{k}\varphi$,
- $\boxminus^{<k+1}\varphi ::= \boxminus^{<k}\varphi \wedge \boxminus^{k}\varphi$.

Now, let us define translation functions that will be useful in Section 5 for proving Propositions 4 and 5 saying that if the fusion of arbitrary consistent modal logics is unitary then both of them are unitary and if the fusion of arbitrary consistent modal logics is finitary then both of them are either unitary, or finitary. We inductively define for all finite sets X of variables and for all $i \in \{1, 2\}$, the translation functions $\mathtt{tr}_i^T : FOR_{\{1,2\}} \longrightarrow FOR_{\{i\}}$ and $\mathtt{tr}_{X,i}^V : FOR_{\{1,2\}} \longrightarrow FOR_{\{i\}}$ as follows:

- $\mathtt{tr}_i^T(\alpha) = \alpha$,
- $\mathtt{tr}_{X,i}^V(\alpha) = \alpha$,
- $\mathtt{tr}_i^T(\bot) = \bot$,
- $\mathtt{tr}_{X,i}^V(\bot) = \bot$,
- $\mathtt{tr}_i^T(\neg\varphi) = \neg\mathtt{tr}_i^T(\varphi)$,
- $\mathtt{tr}_{X,i}^V(\neg\varphi) = \neg\mathtt{tr}_{X,i}^V(\varphi)$,
- $\mathtt{tr}_i^T(\varphi \vee \psi) = \mathtt{tr}_i^T(\varphi) \vee \mathtt{tr}_i^T(\psi)$,
- $\mathtt{tr}_{X,i}^V(\varphi \vee \psi) = \mathtt{tr}_{X,i}^V(\varphi) \vee \mathtt{tr}_{X,i}^V(\psi)$,
- $\mathtt{tr}_i^T(\Box_j\varphi) = \Box_j\mathtt{tr}_i^T(\varphi)$ when $i = j$,
- $\mathtt{tr}_{X,i}^V(\Box_j\varphi) = \Box_j\mathtt{tr}_{X,i}^V(\varphi)$ when $i = j$,
- $\mathtt{tr}_i^T(\Box_j\varphi) = \mathtt{tr}_i^T(\varphi)$ when $i \neq j$,
- $\mathtt{tr}_{X,i}^V(\Box_j\varphi) = \top_X$ when $i \neq j$.

As the reader can see from the above definition, the translation function $\mathtt{tr}_i^T$ does not depend on X. The reader is invited to appreciate the use of the abbreviation $\top_X$ in the definition of the translation function $\mathtt{tr}_{X,i}^V$.

Lemma 1. *Let X be a finite set of variables and $i \in \{1,2\}$. For all $\{i\}$-formulas φ,*

- $\mathtt{tr}_i^T(\varphi) = \varphi$,
- $\mathtt{tr}_{X,i}^V(\varphi) = \varphi$.

Lemma 2. *Let $i \in \{1,2\}$. For all $\{1,2\}$-formulas φ,*

- $\mathtt{var}(\mathtt{tr}_i^T(\varphi)) = \mathtt{var}(\varphi)$,
- $\mathtt{var}(\mathtt{tr}_{\mathtt{var}(\varphi),i}^V(\varphi)) = \mathtt{var}(\varphi)$.

For all finite sets X of variables, for all $i \in \{1,2\}$ and for all $\{1,2\}$-substitutions σ, let σ_i^T and $\sigma_{X,i}^V$ be the $\{i\}$-substitutions defined as follows:

- for all variables x, $\sigma_i^T(x) = \mathtt{tr}_i^T(\sigma(x))$,
- for all variables x, $\sigma_{X,i}^V(x) = \mathtt{tr}_{X,i}^V(\sigma(x))$.

3 Fusions of modal logics

From now on in this paper,

we write "$\bar{1}$" to mean "2" and we write "$\bar{2}$" to mean "1".

3.1 Modal logics

Let I be a non-empty subset of $\{1,2\}$. An *I-logic* is a set $\mathbf{L}$ of I-formulas such that

- $\mathbf{L}$ contains all I-tautologies
- for all $i \in I$, $\mathbf{L}$ contains all I-formulas of the form $\Box_i(\varphi \to \psi) \to (\Box_i\varphi \to \Box_i\psi)$,
- $\mathbf{L}$ is closed under modus ponens (if $\varphi \in \mathbf{L}$ and $\varphi \to \psi \in \mathbf{L}$ then $\psi \in \mathbf{L}$),
- $\mathbf{L}$ is closed under I-generalization (if $\varphi \in \mathbf{L}$ then for all $i \in I$, $\Box_i\varphi \in \mathbf{L}$).

As is well-known, the intersection of I-logics is an I-logic. Hence, for every set of I-formulas, there exists a least I-logic containing it. We shall say that an I-logic $\mathbf{L}$ is *consistent* if $\mathbf{L} \neq FOR_I$. In this paper, it will be useful to remember that for all $i \in \{1,2\}$,

- if an $\{i\}$-logic $\mathbf{L}$ is consistent then either $\mathbf{L}$ is contained in the least $\{i\}$-logic $\mathbf{Triv}_i$ containing all $\{i\}$-formulas of the form $\Box_i\varphi \leftrightarrow \varphi$, or $\mathbf{L}$ is contained in the least $\{i\}$-logic $\mathbf{Verum}_i$ containing all $\{i\}$-formulas of the form $\Box_i\varphi$.

See [25]. For all $i \in \{1,2\}$, we shall say that an $\{i\}$-logic $\mathbf{L}$ is a *non-trivial extension of* $\mathbf{S}5$ if $\mathbf{L}$ contains the least $\{i\}$-logic $\mathbf{S}5_i$ containing all $\{i\}$-formulas of the form $\Box_i\chi \to \chi$, $\Box_i\chi \to \Box_i\Box_i\chi$ and $\chi \to \Box_i\Diamond_i\chi$ and $\mathbf{L}$ is strictly contained in $\mathbf{Triv}_i$. In this paper, it will be useful to remember that for all $i \in \{1,2\}$, if an $\{i\}$-logic $\mathbf{L}$ is a non-trivial extension of $\mathbf{S}5$ then one of the following conditions holds:

- there exists $\mathbf{kk} \geq 2$ such that $\mathbf{L}$ is equal to the least extension $\mathbf{S}5_i^{\mathbf{kk}}$ of $\mathbf{S}5_i$ containing all $\{i\}$-formulas of the form $\bigwedge\{\Diamond_i\varphi_m :\ 0 \leq m \leq \mathbf{kk}\} \to \bigvee\{\Diamond_i(\varphi_m \wedge \varphi_n) :\ 0 \leq m < n \leq \mathbf{kk}\}$,
- $\mathbf{L} = \mathbf{S}5_i$.

See [26, 27]. It will also be useful to remember that for all $i \in \{1,2\}$,

- for all $\mathbf{kk} \geq 2$, $\mathbf{S}5_i^{\mathbf{kk}}$ is a Kripke complete modal logic characterized by the class of all Kripke frames (W, R_i) where R_i is an equivalence relation on W for which each equivalence class is a finite set of exactly $\mathbf{kk}$ possible worlds,
- for all $\mathbf{kk} \geq 2$, $\mathbf{S}5_i^{\mathbf{kk}}$ is a Kripke complete modal logic characterized by the class of all Kripke frames (W, R_i) where R_i is an equivalence relation on W for which each equivalence class is a finite set of at most $\mathbf{kk}$ possible worlds,
- $\mathbf{S}5_i$ is a Kripke complete modal logic characterized by the class of all Kripke frames (W, R_i) where R_i is an equivalence relation on W for which each equivalence class is a countably infinite set of possible worlds.

3.2 Fusions

Let $\mathbf{L}_1$ be a $\{1\}$-logic and $\mathbf{L}_2$ be a $\{2\}$-logic. The *fusion* of $\mathbf{L}_1$ and $\mathbf{L}_2$ is the least $\{1,2\}$-logic (denoted $\mathbf{L}_1 \otimes \mathbf{L}_2$) containing $\mathbf{L}_1$ and $\mathbf{L}_2$[5]. As is well-known, if $\mathbf{L}_1$ is consistent and $\mathbf{L}_2$ is consistent then $\mathbf{L}_1 \otimes \mathbf{L}_2$ is a conservative extension of $\mathbf{L}_1$ and

[5]Generalized to logics formulated in languages with an arbitrary number of modal connectives, the operation of forming fusions is associative. Therefore it makes sense to define the fusion of an arbitrary number of logics $\mathbf{L}_1, \ldots, \mathbf{L}_n$ respectively formulated in languages with the modal connectives $\Box_1, \ldots, \Box_n$ as being the least logic formulated in the language with the modal connectives $\Box_1, \ldots, \Box_n$ and containing $\mathbf{L}_1, \ldots, \mathbf{L}_n$. For instance, the multimodal logics considered in [9, 13] are fusions of finitely many logics of knowledge. In this paper, we will only consider the operation of forming fusions of two unimodal logics.

$\mathbf{L}_2$ [14, 30][6]. A number of other results — transfer results — have been obtained as well. They concern properties preserved under the operation of forming fusions [15, Chapter 4]: the fusion of decidable modal logics is decidable, the fusion of modal logics having uniform interpolation property has uniform interpolation property, etc. We shall say that $\mathbf{L}_1 \otimes \mathbf{L}_2$ is *tensed* if $\mathbf{L}_1 \otimes \mathbf{L}_2$ contains all $\{1,2\}$-formulas of the form $\varphi \to \Box_1\Diamond_1\varphi$ and $\varphi \to \Box_2\Diamond_2\varphi$. We shall say that $\mathbf{L}_1 \otimes \mathbf{L}_2$ is *smooth* if for all $k,l \geq 0$, if $k > l$ then $\boxplus^k\bot \to \boxplus^l\bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^k\bot \to \boxminus^l\bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Lemma 3. *If $\mathbf{L}_1$ and $\mathbf{L}_2$ are non-trivial extensions of $\mathbf{S}5$ then $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed and smooth.*

Within the context of this paper, it is relevant to investigate the properties of the translation functions $\mathtt{tr}_i^T : FOR_{\{1,2\}} \longrightarrow FOR_{\{i\}}$ and $\mathtt{tr}_{X,i}^V : FOR_{\{1,2\}} \longrightarrow FOR_{\{i\}}$ in $\mathbf{L}_1 \otimes \mathbf{L}_2$ for each finite set X of variables and for each $i \in \{1,2\}$. The following results will be used in Section 5.

Lemma 4. *Let X, Y be finite sets of variables and $i \in \{1,2\}$. For all $\{1,2\}$-formulas φ, $\mathtt{tr}_{X,i}^V(\varphi) \leftrightarrow \mathtt{tr}_{Y,i}^V(\varphi) \in \mathbf{L}_i$.*

Lemma 5. *Let X be a finite set of variables, $i \in \{1,2\}$ and σ be a $\{1,2\}$-substitution. For all $\{1,2\}$-formulas φ,*

- $\sigma_i^T(\mathtt{tr}_i^T(\varphi)) \leftrightarrow \mathtt{tr}_i^T(\sigma(\varphi)) \in \mathbf{L}_i$,
- $\sigma_{X,i}^V(\mathtt{tr}_{X,i}^V(\varphi)) \leftrightarrow \mathtt{tr}_{X,i}^V(\sigma(\varphi)) \in \mathbf{L}_i$.

Lemma 6. *Let X be a finite set of variables, $i \in \{1,2\}$ and σ be an $\{i\}$-substitution. For all $\{1,2\}$-formulas φ,*

- $\sigma(\mathtt{tr}_i^T(\varphi)) \leftrightarrow \mathtt{tr}_i^T(\sigma(\varphi)) \in \mathbf{L}_i$,
- $\sigma(\mathtt{tr}_{X,i}^V(\varphi)) \leftrightarrow \mathtt{tr}_{X,i}^V(\sigma(\varphi)) \in \mathbf{L}_i$.

Lemma 7. *Let X be a finite set of variables. For all $\{1,2\}$-formulas φ,*

- $\mathtt{tr}_1^T(\varphi) \leftrightarrow \varphi \in \mathbf{L}_1 \otimes \mathbf{Triv}_2$,
- $\mathtt{tr}_{X,1}^V(\varphi) \leftrightarrow \varphi \in \mathbf{L}_1 \otimes \mathbf{Verum}_2$,

[6]That is to say, when $\mathbf{L}_1$ and $\mathbf{L}_2$ are consistent, for all $i \in \{1,2\}$ and for all $\{i\}$-formulas φ, if $\varphi \in \mathbf{L}_1 \otimes \mathbf{L}_2$ then $\varphi \in \mathbf{L}_i$. Obviously, if either $\mathbf{L}_1$ is inconsistent, or $\mathbf{L}_2$ is inconsistent then $\mathbf{L}_1 \otimes \mathbf{L}_2$ is inconsistent. In actual fact, as noticed by Kracht and Wolter [23], $\mathbf{L}_1 \otimes \mathbf{L}_2$ is a conservative extension of $\mathbf{L}_1$ and $\mathbf{L}_2$ if and only if either $\mathbf{L}_1$ is consistent and $\mathbf{L}_2$ is consistent, or $\mathbf{L}_1$ is inconsistent and $\mathbf{L}_2$ is inconsistent.

- $\mathtt{tr}_2^T(\varphi) \leftrightarrow \varphi \in \mathbf{Triv}_1 \otimes \mathbf{L}_2$,
- $\mathtt{tr}_{X,2}^V(\varphi) \leftrightarrow \varphi \in \mathbf{Verum}_1 \otimes \mathbf{L}_2$.

Lemma 8. *Let X be a finite set of variables, $i \in \{1,2\}$ and φ be a $\{1,2\}$-formula. If $\varphi \in \mathbf{L}_1 \otimes \mathbf{L}_2$ then*

1. *if $\mathbf{L}_{\bar{i}} \subseteq \mathbf{Triv}_{\bar{i}}$ then $\mathtt{tr}_i^T(\varphi) \in \mathbf{L}_i$,*
2. *if $\mathbf{L}_{\bar{i}} \subseteq \mathbf{Verum}_{\bar{i}}$ then $\mathtt{tr}_{X,i}^V(\varphi) \in \mathbf{L}_i$.*

Within the context of this paper, it is relevant to investigate the properties of the modal connectives $\boxplus$ and $\boxminus$ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. The following results will be used in Section 6.

Lemma 9.
1. *$\mathbf{L}_1 \otimes \mathbf{L}_2$ contains all $\{1,2\}$-formulas of the form $\boxplus(\varphi \to \psi) \to (\boxplus\varphi \to \boxplus\psi)$,*
2. *$\mathbf{L}_1 \otimes \mathbf{L}_2$ contains all $\{1,2\}$-formulas of the form $\boxminus(\varphi \to \psi) \to (\boxminus\varphi \to \boxminus\psi)$,*
3. *$\mathbf{L}_1 \otimes \mathbf{L}_2$ is closed under the rule $\frac{\varphi}{\boxplus\varphi}$,*
4. *$\mathbf{L}_1 \otimes \mathbf{L}_2$ is closed under the rule $\frac{\varphi}{\boxminus\varphi}$,*
5. *if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed then $\mathbf{L}_1 \otimes \mathbf{L}_2$ is closed under the rule $\frac{\neg\varphi\to\boxplus\psi}{\neg\psi\to\boxminus\varphi}$,*
6. *if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed then $\mathbf{L}_1 \otimes \mathbf{L}_2$ is closed under the rule $\frac{\neg\varphi\to\boxminus\psi}{\neg\psi\to\boxplus\varphi}$.*

Lemma 10. *For all $k \geq 0$,*

1. *$\boxplus^k\top \in \mathbf{L}_1 \otimes \mathbf{L}_2$,*
2. *$\boxminus^k\top \in \mathbf{L}_1 \otimes \mathbf{L}_2$,*
3. *$\boxplus^{<k}\top \in \mathbf{L}_1 \otimes \mathbf{L}_2$,*
4. *$\boxminus^{<k}\top \in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 11. *Let $k \geq 0$. For all $\{1,2\}$-formulas φ,*

1. *$\boxplus^{<k+1}\varphi \leftrightarrow \varphi \wedge \boxplus\boxplus^{<k}\varphi \in \mathbf{L}_1 \otimes \mathbf{L}_2$,*
2. *$\boxminus^{<k+1}\varphi \leftrightarrow \varphi \wedge \boxminus\boxminus^{<k}\varphi \in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 12. *Let $k \geq 0$. If $\mathbf{L}_1 \otimes \mathbf{L}_2$ is smooth then*

1. $\boxplus^k\bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$,

2. $\boxminus^k\bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Lemma 13. *Let $k \geq 0$. If $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed and smooth then for all $l \geq 0$, $\boxplus^k\bot \vee \boxminus^l\bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

In anticipation of our results about the unification type of fusions in Sections 5 and 6, we complete this section by defining the families $(\sigma_k)_{k\geq 0}$, $(\tau_k)_{k\geq 0}$, $(\lambda_k)_{k\geq 0}$ and $(\mu_k)_{k\geq 0}$ of $\{1,2\}$-substitutions and by proving some of their properties. From now on in this paper,

$$\boxed{\textbf{let } x \textbf{ be a fixed variable.}}$$

For all $k \geq 0$, let σ_k and τ_k be the $\{1,2\}$-substitutions inductively defined as follows:

- $\sigma_0(x) = \bot$,
- for all variables y distinct from x, $\sigma_0(y) = y$,
- $\tau_0(x) = \top$,
- for all variables y distinct from x, $\tau_0(y) = y$,
- $\sigma_{k+1}(x) = x \wedge \boxplus\sigma_k(x)$,
- for all variables y distinct from x, $\sigma_{k+1}(y) = y$,
- $\tau_{k+1}(x) = \neg(\neg x \wedge \boxminus\neg\tau_k(x))$,
- for all variables y distinct from x, $\tau_{k+1}(y) = y$.

For all $k \geq 0$, let λ_k and μ_k be the $\{1,2\}$-substitutions defined as follows:

- $\lambda_k(x) = x \wedge \boxplus^k\bot$,
- for all variables y distinct from x, $\lambda_k(y) = y$,
- $\mu_k(x) = \neg(\neg x \wedge \boxminus^k\bot)$,
- for all variables y distinct from x, $\mu_k(y) = y$.

Lemma 14. *Let $k \geq 0$. We have $\boxplus^{<k}x \wedge \boxplus^k\bot \to \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^{<k}\neg x \wedge \boxminus^k\bot \to \neg\tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 15. *Let $k \geq 0$. We have $\sigma_k(x) \to x \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg\tau_k(x) \to \neg x \in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 16. *Let $k \geq 0$. We have $\sigma_k(x) \to \boxplus\sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg\tau_k(x) \to \boxminus\neg\tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 17. *Let $k \geq 0$. For all $l \geq 0$, if $k \leq l$ then $\sigma_k(x) \to \boxplus^l\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg\tau_k(x) \to \boxminus^l\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 18. *Let $k \geq 0$. For all $l \geq 0$, if $k \leq l$ then $\boxplus^k\bot \wedge \sigma_l(x) \leftrightarrow \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^k\bot \wedge \neg\tau_l(x) \leftrightarrow \neg\tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 19. *Let $k \geq 0$. For all $l \geq 0$, if $k \leq l$ then $\lambda_l(\sigma_k(x)) \leftrightarrow \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\mu_l(\tau_k(x)) \leftrightarrow \tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 20. *Let $k \geq 0$. For all $l \geq 0$, if $k \geq l$ then $\lambda_l(\sigma_k(x)) \leftrightarrow \sigma_l(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\mu_l(\tau_k(x)) \leftrightarrow \tau_l(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 21. *Let $k \geq 0$. If $\mathbf{L}_1 \otimes \mathbf{L}_2$ is smooth then for all $l \geq 0$, if $k > l$ then $\sigma_k(x) \to \boxplus^l\bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg\tau_k(x) \to \boxminus^l\bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 22. *Let $k \geq 0$. If $\mathbf{L}_1 \otimes \mathbf{L}_2$ is smooth then for all $l \geq 0$, $\boxplus^k\bot \vee \neg\tau_l(x) \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^k\bot \vee \sigma_l(x) \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

4 Unification

4.1 Unifiable formulas and unification types

Let I be a non-empty subset of $\{1, 2\}$. Let $\mathbf{L}$ be an I-logic. We shall say that an I-substitution σ is *equivalent* in $\mathbf{L}$ to an I-substitution τ with respect to a set $\mathcal{X}$ of variables (in symbols $\sigma \simeq^{\mathcal{X}}_{\mathbf{L}} \tau$) if for all variables $y \in \mathcal{X}$, $\sigma(y) \leftrightarrow \tau(y) \in \mathbf{L}$. We shall say that an I-substitution σ is more *general* in $\mathbf{L}$ than an I-substitution τ with respect to a set $\mathcal{X}$ of variables (in symbols $\sigma \preceq^{\mathcal{X}}_{\mathbf{L}} \tau$) if there exists an I-substitution υ such that $\sigma \circ \upsilon \simeq^{\mathcal{X}}_{\mathbf{L}} \tau$. Obviously, for all sets $\mathcal{X}$ of variables and for all I-substitutions σ, τ, if $\sigma \simeq^{\mathcal{X}}_{\mathbf{L}} \tau$ then $\sigma \preceq^{\mathcal{X}}_{\mathbf{L}} \tau$. Moreover, for all sets $\mathcal{X}$ of variables, on the set of all I-substitutions, the binary relation $\simeq^{\mathcal{X}}_{\mathbf{L}}$ is reflexive, symmetric and transitive and the binary relation $\preceq^{\mathcal{X}}_{\mathbf{L}}$ is reflexive and transitive. We shall say that an I-formula φ is $\mathbf{L}$-*unifiable* if there exists an I-substitution σ such that $\sigma(\varphi) \in \mathbf{L}$. In that case, σ is an $\mathbf{L}$-*unifier* of φ. We shall say that a set Σ of $\mathbf{L}$-unifiers of an $\mathbf{L}$-unifiable I-formula φ is $\mathbf{L}$-*complete* if for all $\mathbf{L}$-unifiers σ of φ, there exists $\tau \in \Sigma$ such that $\tau \preceq^{\mathtt{var}(\varphi)}_{\mathbf{L}} \sigma$. As is well-known, if an $\mathbf{L}$-unifiable I-formula has minimal $\mathbf{L}$-complete sets of $\mathbf{L}$-unifiers then these sets have the same cardinality[7]. About the type of

[7]Suppose Σ and Δ are minimal $\mathbf{L}$-complete sets of $\mathbf{L}$-unifiers of the same $\mathbf{L}$-unifiable I-formula φ. By the $\mathbf{L}$-completeness of Σ and Δ, one can readily define functions $f : \Sigma \longrightarrow \Delta$ and $g : \Delta \longrightarrow \Sigma$ such that $f(\sigma) \preceq^{\mathtt{var}(\varphi)}_{\mathbf{L}} \sigma$ for each $\sigma \in \Sigma$ and $g(\delta) \preceq^{\mathtt{var}(\varphi)}_{\mathbf{L}} \delta$ for each $\delta \in \Delta$. By the minimality of Σ and Δ, it follows that f and g are injective. Hence, Σ and Δ have the same cardinality.

L-unifiable I-formulas, we shall say that an **L**-unifiable I-formula

- φ is **L**-*nullary (or of type* 0*)* if there exists no minimal **L**-complete set of **L**-unifiers of φ,
- φ is **L**-*infinitary (or of type* ∞*)* if there exists a minimal **L**-complete set of **L**-unifiers of φ but there exists no finite one,
- φ is **L**-*finitary (or of type* ω*)* if there exists a finite minimal **L**-complete set of **L**-unifiers of φ but there exists no with cardinality 1,
- φ is **L**-*unitary (or of type* 1*)* if there exists a minimal **L**-complete set of **L**-unifiers of φ with cardinality 1.

Obviously, the types "**L**-nullary", "**L**-infinitary", "**L**-finitary" and "**L**-unitary" constitute a set of jointly exhaustive and pairwise distinct situations. To be of type 0 is considered to be the worst situation whereas to be of type 1 is considered to be better than to be of type ω which is itself considered to be better than to be of type ∞. As for the type of **L**, we traditionally distinguish between *elementary unification* and *unification with parameters*:

- elementary unification in **L** is the problem of asking whether a given parameter-free I-formula is **L**-unifiable,
- unification with parameters in **L** is the problem of asking whether a given I-formula is **L**-unifiable.

We shall say that

- **L** is *nullary (or of type* 0*) for elementary unification* if there exists an **L**-nullary **L**-unifiable parameter-free I-formula,
- **L** is *infinitary (or of type* ∞*) for elementary unification* if every **L**-unifiable parameter-free I-formula is either **L**-unitary, or **L**-finitary, or **L**-infinitary and there exists an **L**-infinitary **L**-unifiable parameter-free I-formula,
- **L** is *finitary (or of type* ω*) for elementary unification* if every **L**-unifiable parameter-free I-formula is either **L**-unitary, or **L**-finitary and there exists an **L**-finitary **L**-unifiable parameter-free I-formula,
- **L** is *unitary (or of type* 1*) for elementary unification* if every **L**-unifiable parameter-free I-formula is **L**-unitary.

We shall say that

- **L** is *nullary (or of type* 0*) for unification with parameters* if there exists an **L**-nullary **L**-unifiable I-formula,
- **L** is *infinitary (or of type* ∞*) for unification with parameters* if every **L**-unifiable I-formula is either **L**-unitary, or **L**-finitary, or **L**-infinitary and there exists an **L**-infinitary **L**-unifiable I-formula,
- **L** is *finitary (or of type* ω*) for unification with parameters* if every **L**-unifiable I-formula is either **L**-unitary, or **L**-finitary and there exists an **L**-finitary **L**-unifiable I-formula,
- **L** is *unitary (or of type* 1*) for unification with parameters* if every **L**-unifiable I-formula is **L**-unitary.

Obviously, both for elementary unification and for unification with parameters, the types "nullary", "infinitary", "finitary" and "unitary" constitute a set of jointly exhaustive and pairwise distinct situations. In other respects, the unification type of **L** for elementary unification is at least better than the unification type of **L** for unification with parameters and there is a priori no guarantee that the unification type for elementary unification and the unification type for unification with parameters are equal. For instance, the implication fragment of Classical Propositional Logic is unitary for elementary unification and finitary for unification with parameters [6][8]. To the extent that in cases such as **KB**, **KD**, **KDB**, **KT** and **KTB**, the unification type for unification with parameters is known whereas the unification type for elementary unification is still a mystery[9]. Of course, seeing that the unification type of an equational theory depends not only on the equational theory itself but also on the set of symbols that can occur in the considered unification problems, this phenomenon is already well-known from the theory of unification [2]. Finally, as already noticed by several authors within the context of unimodal logics, there is no I-logic **L** that is known to be infinitary either for elementary unification, or for unification with parameters. See [11].

4.2 Playing with formulas and substitutions

Let $\mathbf{L}_1$ be a consistent $\{1\}$-logic and $\mathbf{L}_2$ be a consistent $\{2\}$-logic.

[8]For all parameter-free formulas φ with $\rightarrow$ as its sole connective, φ is unifiable in Classical Propositional Logic and the so-called Löwenheim substitution ϵ defined by $\epsilon(y) = \varphi \rightarrow y$ for each $y \in \mathtt{var}(\varphi)$ constitutes a minimal complete set of unifiers of it.

[9]**KB**, **KD**, **KDB**, **KT** and **KTB** are nullary for unification with parameters [3, 4, 5].

Lemma 23. *Let $k \geq 0$. For all $l \geq 0$, if $k \leq l$ then $\sigma_l \circ \lambda_k \simeq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma_k$ and $\tau_l \circ \mu_k \simeq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \tau_k$.*

Lemma 24. *Let $k \geq 0$. For all $l \geq 0$, if $k \leq l$ then $\sigma_l \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma_k$ and $\tau_l \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \tau_k$.*

Lemma 25. *Let $k \geq 0$. If $\mathbf{L}_1 \otimes \mathbf{L}_2$ is smooth then for all $l \geq 0$, if $k < l$ then $\sigma_k \not\preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma_l$ and $\tau_k \not\preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \tau_l$.*

Lemma 26. *Let $k \geq 0$. If $\mathbf{L}_1 \otimes \mathbf{L}_2$ is smooth then for all $l \geq 0$, $\sigma_k \not\preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \tau_l$ and $\tau_k \not\preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma_l$.*

From now on in this paper,

$$\boxed{\textbf{let } \varphi \textbf{ be the } \{1,2\}\textbf{-formula } x \to \boxplus x \textbf{ and } \psi \textbf{ be the } \{1,2\}\textbf{-formula } \neg x \to \boxminus \neg x.}$$

The $\{1,2\}$-formulas φ and ψ will be the keys in Section 6 to the determination of the unification type of the fusion of arbitrary consistent extensions of $\mathbf{S}5$. In the meantime, by Lemma 9, if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed then φ and ψ have the same unifiers in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, in that case, as long as we only consider φ and ψ through their unifiers in $\mathbf{L}_1 \otimes \mathbf{L}_2$, it does not matter if we are talking about either φ, or ψ.

Lemma 27. *Let $k \geq 0$. For all unifiers σ of φ in $\mathbf{L}_1 \otimes \mathbf{L}_2$, $\sigma(x) \to \boxplus^{<k} \sigma(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and for all unifiers τ of ψ in $\mathbf{L}_1 \otimes \mathbf{L}_2$, $\neg\tau(x) \to \boxminus^{<k} \neg\tau(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 28. *For all $k \geq 0$, σ_k is a unifier of φ in $\mathbf{L}_1 \otimes \mathbf{L}_2$ and τ_k is a unifier of ψ in $\mathbf{L}_1 \otimes \mathbf{L}_2$.*

Lemma 29. *Let υ be a $\{1,2\}$-substitution. If υ is a unifier of φ in $\mathbf{L}_1 \otimes \mathbf{L}_2$ then for all $k \geq 0$, the following conditions are equivalent:*

(a) $\sigma_k \circ \upsilon \simeq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$,

(b) $\sigma_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$,

(c) $\upsilon(x) \to \boxplus^k \bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$

and if υ is a unifier of ψ in $\mathbf{L}_1 \otimes \mathbf{L}_2$ then for all $k \geq 0$, the following conditions are equivalent:

(d) $\tau_k \circ \upsilon \simeq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$,

(e) $\tau_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$,

(f) $\neg\upsilon(x) \to \boxminus^k \bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

5 General results about the unification type of fusions

Let $\mathbf{L}_1$ be a consistent $\{1\}$-logic and $\mathbf{L}_2$ be a consistent $\{2\}$-logic.

Proposition 1. *Let $i \in \{1,2\}$ and χ be an $\{i\}$-formula. If χ is unifiable in $\mathbf{L}_1 \otimes \mathbf{L}_2$ then χ is $\mathbf{L}_i$-unifiable.*

Proof. Suppose χ is unifiable in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, there exists a $\{1,2\}$-substitution σ such that $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Without loss of generality, suppose $i = 1$. Since $\mathbf{L}_2$ is consistent, either $\mathbf{L}_2 \subseteq \mathbf{Triv}_2$, or $\mathbf{L}_2 \subseteq \mathbf{Verum}_2$. In the former case, $\mathbf{L}_1 \otimes \mathbf{L}_2 \subseteq \mathbf{L}_1 \otimes \mathbf{Triv}_2$[10]. Since $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{Triv}_2$. Thus, by Lemma 7, $\mathtt{tr}_1^T(\sigma(\chi)) \in \mathbf{L}_1 \otimes \mathbf{Triv}_2$. Since $\mathbf{L}_1 \otimes \mathbf{Triv}_2$ is a conservative extension of $\mathbf{L}_1$, it follows that $\mathtt{tr}_1^T(\sigma(\chi)) \in \mathbf{L}_1$. Consequently, by Lemma 5, $\sigma_1^T(\mathtt{tr}_1^T(\chi)) \in \mathbf{L}_1$. Hence, by Lemma 1, $\sigma_1^T(\chi) \in \mathbf{L}_1$. Thus, χ is $\mathbf{L}_1$-unifiable. $\square$

Proposition 2. *Let $i \in \{1,2\}$ and χ be an $\{i\}$-formula. For all complete sets Σ of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$,*

1. *if $L_{\bar{i}} \subseteq \mathbf{Triv}_{\bar{i}}$ then $\{\sigma_i^T : \sigma \in \Sigma\}$ is an $\mathbf{L}_i$-complete set of $\mathbf{L}_i$-unifiers of χ,*
2. *if $L_{\bar{i}} \subseteq \mathbf{Verum}_{\bar{i}}$ then $\{\sigma_{\emptyset,i}^V : \sigma \in \Sigma\}$ is an $\mathbf{L}_i$-complete set of $\mathbf{L}_i$-unifiers of χ.*

Proof. Let Σ be a complete set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. The proof of Item (1) can be done as follows[11].

Suppose $i = 1$ and $L_2 \subseteq \mathbf{Triv}_2$.

Claim $\{\sigma_1^T : \sigma \in \Sigma\}$ is a set of $\mathbf{L}_1$-unifiers of χ.

Proof: It suffices to prove that for all $\sigma \in \Sigma$, $\sigma_1^T(\chi) \in \mathbf{L}_1$. Let $\sigma \in \Sigma$. The proof that $\sigma_1^T(\chi) \in \mathbf{L}_1$ is essentially the one described in the body of the proof of Proposition 1. We include it here for the sake of the completeness. Since Σ is a set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$, we obtain $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_2 \subseteq \mathbf{Triv}_2$, $\mathbf{L}_1 \otimes \mathbf{L}_2 \subseteq \mathbf{L}_1 \otimes \mathbf{Triv}_2$. Since $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{Triv}_2$. Hence, by Lemma 7, $\mathtt{tr}_1^T(\sigma(\chi)) \in \mathbf{L}_1 \otimes \mathbf{Triv}_2$. Since $\mathbf{L}_1 \otimes \mathbf{Triv}_2$ is a conservative extension of $\mathbf{L}_1$, $\mathtt{tr}_1^T(\sigma(\chi)) \in \mathbf{L}_1$. Thus, by Lemma 5, $\sigma_1^T(\mathtt{tr}_1^T(\chi)) \in \mathbf{L}_1$. Hence, by Lemma 1, $\sigma_1^T(\chi) \in \mathbf{L}_1$.

Claim $\{\sigma_1^T : \sigma \in \Sigma\}$ is an $\mathbf{L}_1$-complete set of $\mathbf{L}_1$-unifiers of χ.

[10] In the latter case, the proof can be similarly done.

[11] The proof of Items (2)–(4) can be similarly done.

Proof: By the previous Claim, it suffices to prove that for all $\{1\}$-substitutions σ, if $\sigma(\chi) \in \mathbf{L}_1$ then there exists $\tau \in \Sigma$ such that $\tau_1^T \preceq_{\mathbf{L}_1}^{\mathtt{var}(\chi)} \sigma$. Let σ be a $\{1\}$-substitution. Suppose $\sigma(\chi) \in \mathbf{L}_1$. Hence, $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since Σ is a complete set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$, there exists $\tau \in \Sigma$ such that $\tau \preceq_{\mathbf{L}_1 \otimes \mathbf{L}_2}^{\mathtt{var}(\chi)} \sigma$. Thus, there exists a $\{1,2\}$-substitution υ such that $\tau \circ \upsilon \simeq_{\mathbf{L}_1 \otimes \mathbf{L}_2}^{\mathtt{var}(\chi)} \sigma$. Hence, for all variables $y \in \mathtt{var}(\chi)$, $\upsilon(\tau(y)) \leftrightarrow \sigma(y) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_2 \subseteq \mathbf{Triv}_2$, $\mathbf{L}_1 \otimes \mathbf{L}_2 \subseteq \mathbf{L}_1 \otimes \mathbf{Triv}_2$. Since for all variables $y \in \mathtt{var}(\chi)$, $\upsilon(\tau(y)) \leftrightarrow \sigma(y) \in \mathbf{L}_1 \otimes \mathbf{L}_2$, for all variables $y \in \mathtt{var}(\chi)$, $\upsilon(\tau(y)) \leftrightarrow \sigma(y) \in \mathbf{L}_1 \otimes \mathbf{Triv}_2$. Thus, by Lemma 7, for all variables $y \in \mathtt{var}(\chi)$, $\mathtt{tr}_1^T(\upsilon(\tau(y)) \leftrightarrow \sigma(y)) \in \mathbf{L}_1 \otimes \mathbf{Triv}_2$. Since $\mathbf{L}_1 \otimes \mathbf{Triv}_2$ is a conservative extension of $\mathbf{L}_1$, for all variables $y \in \mathtt{var}(\chi)$, $\mathtt{tr}_1^T(\upsilon(\tau(y)) \leftrightarrow \sigma(y)) \in \mathbf{L}_1$. Thus, for all variables $y \in \mathtt{var}(\chi)$, $\mathtt{tr}_1^T(\upsilon(\tau(y))) \leftrightarrow \mathtt{tr}_1^T(\sigma(y)) \in \mathbf{L}_1$. Consequently, by Lemma 5, for all variables $y \in \mathtt{var}(\chi)$, $\upsilon_1^T(\tau_1^T(y)) \leftrightarrow \mathtt{tr}_1^T(\sigma(y)) \in \mathbf{L}_1$. Hence, by Lemma 1, for all variables $y \in \mathtt{var}(\chi)$, $\upsilon_1^T(\tau_1^T(y)) \leftrightarrow \sigma(y) \in \mathbf{L}_1$. Thus, $\tau_1^T \circ \upsilon_1^T \simeq_{\mathbf{L}_1}^{\mathtt{var}(\chi)} \sigma$. Consequently, $\tau_1^T \preceq_{\mathbf{L}_1}^{\mathtt{var}(\chi)} \sigma$.

This ends the proof of Proposition 2. □

Proposition 3. *Let $i \in \{1,2\}$ and χ be a $\{1,2\}$-formula.*

1. *For all minimal $\mathbf{L}_i$-complete sets Σ of $\mathbf{L}_i$-unifiers of $\mathtt{tr}_i^T(\chi)$, if $L_{\bar{i}} = \mathbf{Triv}_{\bar{i}}$ then Σ is a minimal complete set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$,*
2. *for all minimal $\mathbf{L}_i$-complete sets Σ of $\mathbf{L}_i$-unifiers of $\mathtt{tr}_{\mathtt{var}(\chi),i}^V(\chi)$, if $L_{\bar{i}} = \mathbf{Verum}_{\bar{i}}$ then Σ is a minimal complete set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$.*

Proof. The proof of Item (1) can be done as follows[12].

Let Σ be a minimal $\mathbf{L}_i$-complete set of $\mathbf{L}_i$-unifiers of $\mathtt{tr}_i^T(\chi)$. Suppose $i = 1$ and $L_2 = \mathbf{Triv}_2$.

Claim Σ is a set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof: It suffices to prove that for all $\sigma \in \Sigma$, $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Let $\sigma \in \Sigma$. Since Σ is a set of $\mathbf{L}_i$-unifiers of $\mathtt{tr}_i^T(\chi)$, $\sigma(\mathtt{tr}_i^T(\chi)) \in \mathbf{L}_i$. Hence, by Lemma 6, $\mathtt{tr}_i^T(\sigma(\chi)) \in \mathbf{L}_i$. Since $i = 1$ and $L_2 = \mathbf{Triv}_2$, then by Lemma 7, $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Claim Σ is a complete set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$.

[12]The proof of Item (2) can be similarly done.

Proof: By the previous Claim, it suffices to prove that for all $\{1,2\}$-substitutions σ, if $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ then there exists $\tau \in \Sigma$ such that $\tau \preceq^{\mathtt{var}(\chi)}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma$. Let σ be a $\{1,2\}$-substitution. Suppose $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $i = 1$ and $L_2 = \mathbf{Triv}_2$, then by Lemma 7, $\mathtt{tr}_i^T(\sigma(\chi)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is a conservative extension of $\mathbf{L}_i$, $\mathtt{tr}_i^T(\sigma(\chi)) \in \mathbf{L}_i$. Hence, by Lemma 5, $\sigma_i^T(\mathtt{tr}_i^T(\chi)) \in \mathbf{L}_i$. Thus, σ_i^T is an $\mathbf{L}_i$-unifier of $\mathtt{tr}_i^T(\chi)$. Since Σ is an $\mathbf{L}_i$-complete set of $\mathbf{L}_i$-unifiers of $\mathtt{tr}_i^T(\chi)$, there exists $\tau \in \Sigma$ such that $\tau \preceq^{\mathtt{var}(\mathtt{tr}_i^T(\chi))}_{\mathbf{L}_i} \sigma_i^T$. Consequently, there exists an $\{i\}$-substitution υ such that $\tau \circ \upsilon \simeq^{\mathtt{var}(\mathtt{tr}_i^T(\chi))}_{\mathbf{L}_i} \sigma_i^T$. Hence, for all variables $y \in \mathtt{var}(\mathtt{tr}_i^T(\chi))$, $\upsilon(\tau(y)) \leftrightarrow \sigma_i^T(y) \in \mathbf{L}_i$. Thus, by Lemma 1, for all variables $y \in \mathtt{var}(\mathtt{tr}_i^T(\chi))$, $\mathtt{tr}_i^T(\upsilon(\tau(y))) \leftrightarrow \mathtt{tr}_i^T(\sigma(y)) \in \mathbf{L}_i$. Consequently, for all variables $y \in \mathtt{var}(\mathtt{tr}_i^T(\chi))$, $\mathtt{tr}_i^T(\upsilon(\tau(y))) \leftrightarrow \mathtt{tr}_i^T(\sigma(y)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $i = 1$ and $L_2 = \mathbf{Triv}_2$, then by Lemma 7, for all variables $y \in \mathtt{var}(\mathtt{tr}_i^T(\chi))$, $\upsilon(\tau(y)) \leftrightarrow \sigma(y) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, by Lemma 2, for all variables $y \in \mathtt{var}(\chi)$, $\upsilon(\tau(y)) \leftrightarrow \sigma(y) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $\tau \circ \upsilon \simeq^{\mathtt{var}(\chi)}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma$. Consequently, $\tau \preceq^{\mathtt{var}(\chi)}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma$.

Claim Σ is a minimal complete set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof: By the previous Claim, it suffices to prove that for all $\sigma, \tau \in \Sigma$, if $\sigma \preceq^{\mathtt{var}(\chi)}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \tau$ then $\sigma = \tau$. Let $\sigma, \tau \in \Sigma$. Suppose $\sigma \preceq^{\mathtt{var}(\chi)}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \tau$. Hence, there exists an $\{1,2\}$-substitution υ such that $\sigma \circ \upsilon \simeq^{\mathtt{var}(\chi)}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \tau$. Thus, for all variables $y \in \mathtt{var}(\chi)$, $\upsilon(\sigma(y)) \leftrightarrow \tau(y) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, by Lemma 2, for all variables $y \in \mathtt{var}(\mathtt{tr}_i^T(\chi))$, $\upsilon(\sigma(y)) \leftrightarrow \tau(y) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $i = 1$ and $L_2 = \mathbf{Triv}_2$, then by Lemma 7, for all variables $y \in \mathtt{var}(\mathtt{tr}_i^T(\chi))$, $\mathtt{tr}_i^T(\upsilon(\sigma(y))) \leftrightarrow \mathtt{tr}_i^T(\tau(y)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is a conservative extension of $\mathbf{L}_i$, for all variables $y \in \mathtt{var}(\mathtt{tr}_i^T(\chi))$, $\mathtt{tr}_i^T(\upsilon(\sigma(y))) \leftrightarrow \mathtt{tr}_i^T(\tau(y)) \in \mathbf{L}_i$. Consequently, by Lemmas 1 and 5, for all variables $y \in \mathtt{var}(\mathtt{tr}_i^T(\chi))$, $\upsilon_i^T(\sigma(y)) \leftrightarrow \tau(y) \in \mathbf{L}_i$. Thus, $\sigma \preceq^{\mathtt{var}(\mathtt{tr}_i^T(\chi))}_{\mathbf{L}_i} \tau$. Since Σ is a minimal $\mathbf{L}_i$-complete set of $\mathbf{L}_i$-unifiers of $\mathtt{tr}_i^T(\chi)$, $\sigma = \tau$.

This ends the proof of Proposition 3. $\square$

In the above proof, the reader is invited to appreciate the uses of Lemma 2.

Proposition 4. *Both for elementary unification and for unification with parameters, if* $\mathbf{L}_1 \otimes \mathbf{L}_2$ *is of type* 1 *then for all* $i \in \{1,2\}$, $\mathbf{L}_i$ *is of type* 1.

Proof. Suppose $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type 1. Suppose $i = 1$.

It suffices to prove that for all $\mathbf{L}_1$-unifiable $\{1\}$-formulas χ, χ is $\mathbf{L}_1$-unitary. Let χ be an $\mathbf{L}_1$-unifiable $\{1\}$-formula. Hence, there exists a $\{1\}$-substitution σ such that $\sigma(\chi) \in \mathbf{L}_1$. Thus, $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, χ is unifiable in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type 1, there exists a minimal complete set Σ of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$ with cardinality 1. Since $\mathbf{L}_2$ is consistent, either $\mathbf{L}_2 \subseteq \mathbf{Triv}_2$, or $\mathbf{L}_2 \subseteq \mathbf{Verum}_2$. In the former case, by Proposition 2, $\{\sigma_1^T : \sigma \in \Sigma\}$ is an $\mathbf{L}_1$-complete set of $\mathbf{L}_1$-unifiers of χ[13]. Since the cardinality of Σ is 1, the cardinality of $\{\sigma_1^T : \sigma \in \Sigma\}$ is 1. Consequently, χ is $\mathbf{L}_1$-unitary. □

Notice that the converse of the statement established in Proposition 4 is not always true. For instance, as proved in Section 6, $\mathbf{S}5 \otimes \mathbf{S}5$ is of type 0.

Proposition 5. *Both for elementary unification and for unification with parameters, if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type ω then for all $i \in \{1,2\}$, $\mathbf{L}_i$ is either of type 1, or of type ω.*

Proof. Suppose $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type ω. Suppose $i = 1$.

It suffices to prove that for all $\mathbf{L}_1$-unifiable $\{1\}$-formulas χ, χ is either $\mathbf{L}_1$-unitary, or $\mathbf{L}_1$-finitary. Let χ be an $\mathbf{L}_1$-unifiable $\{1\}$-formula. The proof that χ is either $\mathbf{L}_1$-unitary, or $\mathbf{L}_1$-finitary is essentially the one described in the body of the proof of Proposition 4. We include it here for the sake of the completeness. Since χ is an $\mathbf{L}_1$-unifiable $\{1\}$-formula, there exists a $\{1\}$-substitution σ such that $\sigma(\chi) \in \mathbf{L}_1$. Thus, $\sigma(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, χ is unifiable in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type ω, there exists a finite minimal complete set Σ of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_2$ is consistent, either $\mathbf{L}_2 \subseteq \mathbf{Triv}_2$, or $\mathbf{L}_2 \subseteq \mathbf{Verum}_2$. In the former case, by Proposition 2, $\{\sigma_1^T : \sigma \in \Sigma\}$ is an $\mathbf{L}_1$-complete set of $\mathbf{L}_1$-unifiers of χ[14]. Since Σ is finite, $\{\sigma_1^T : \sigma \in \Sigma\}$ is finite. Consequently, χ is either $\mathbf{L}_1$-unitary, or $\mathbf{L}_1$-finitary. □

Notice that the converse of the statement established in Proposition 5 is not always true. For instance, as proved in [28, Chapter 6], $\mathbf{K}4 \otimes \mathbf{K}4$ and $\mathbf{S}4 \otimes \mathbf{S}4$ are of type 0. After Propositions 4 and 5, it is natural to ask whether both for elementary unification and for unification with parameters, if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type ∞ then for all $i \in \{1,2\}$, $\mathbf{L}_i$ is either of type 1, or of type ω, or of type ∞. Unfortunately, we have not been able to answer this question, seeing that in Proposition 2, it is not clear that if the set Σ considered there is an infinite minimal complete set of unifiers then, when either $L_{\bar{i}} \subseteq \mathbf{Triv}_{\bar{i}}$, or $L_{\bar{i}} \subseteq \mathbf{Verum}_{\bar{i}}$, the corresponding set among $\{\sigma_i^T : \sigma \in \Sigma\}$ and $\{\sigma_{\emptyset,i}^V : \sigma \in \Sigma\}$ is minimal complete too.

[13] In the latter case, the proof can be similarly done.

[14] In the latter case, the proof can be similarly done.

Proposition 6. *Both for elementary unification and for unification with parameters, for all $i \in \{1,2\}$, if either $\mathbf{L}_i = \mathbf{Triv}_i$, or $\mathbf{L}_i = \mathbf{Verum}_i$ then the type of $\mathbf{L}_1 \otimes \mathbf{L}_2$ and the type of $\mathbf{L}_{\bar{i}}$ are equal.*

Proof. Suppose $i = 1$.

Suppose either $\mathbf{L}_1 = \mathbf{Triv}_1$, or $\mathbf{L}_1 = \mathbf{Verum}_1$. In the former case, for the sake of the contradiction, suppose that the type of $\mathbf{L}_1 \otimes \mathbf{L}_2$ and the type of $\mathbf{L}_2$ are not equal[15]. We consider the following cases.

Case $\mathbf{L}_2$ is of type 0: Hence, there exists an $\mathbf{L}_2$-unifiable $\{2\}$-formula χ of type 0. Thus, there exists an $\mathbf{L}_2$-unifier υ of χ. Hence, υ is a unifier of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is not of type 0, there exists a minimal complete set Σ of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 = \mathbf{Triv}_1$, then by Proposition 2, $\{\sigma_2^T : \sigma \in \Sigma\}$ is an $\mathbf{L}_2$-complete set of $\mathbf{L}_2$-unifiers of χ. Since χ is of type 0, $\{\sigma_2^T : \sigma \in \Sigma\}$ is not a minimal $\mathbf{L}_2$-complete set of $\mathbf{L}_2$-unifiers of χ. Consequently, there exists $\sigma, \tau \in \Sigma$ such that $\sigma_2^T \preceq_{\mathbf{L}_2}^{\mathtt{var}(\chi)} \tau_2^T$ and $\sigma_2^T \neq \tau_2^T$. Thus, $\sigma \neq \tau$. Since $\sigma_2^T \preceq_{\mathbf{L}_2}^{\mathtt{var}(\chi)} \tau_2^T$, it follows that there exists a $\{2\}$-substitution λ such that for all variables $y \in \mathtt{var}(\chi)$, $\lambda(\sigma_2^T(y)) \leftrightarrow \tau_2^T(y) \in \mathbf{L}_2$. Consequently, for all variables $y \in \mathtt{var}(\chi)$, $\lambda(\mathtt{tr}_2^T(\sigma(y))) \leftrightarrow \mathtt{tr}_2^T(\tau(y)) \in \mathbf{L}_2$. Thus, by Lemma 6, for all variables $y \in \mathtt{var}(\chi)$, $\mathtt{tr}_2^T(\lambda(\sigma(y))) \leftrightarrow \mathtt{tr}_2^T(\tau(y)) \in \mathbf{L}_2$. Hence, for all variables $y \in \mathtt{var}(\chi)$, $\mathtt{tr}_2^T(\lambda(\sigma(y))) \leftrightarrow \mathtt{tr}_2^T(\tau(y)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 = \mathbf{Triv}_1$, then by Lemma 7, for all variables $y \in \mathtt{var}(\chi)$, $\lambda(\sigma(y)) \leftrightarrow \tau(y) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, $\sigma \circ \lambda \simeq_{\mathbf{L}_1 \otimes \mathbf{L}_2}^{\mathtt{var}(\chi)} \tau$. Thus, $\sigma \preceq_{\mathbf{L}_1 \otimes \mathbf{L}_2}^{\mathtt{var}(\chi)} \tau$. Since Σ is a minimal complete set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$, $\sigma = \tau$: a contradiction.

Case $\mathbf{L}_2$ is of type ∞: Hence, by Propositions 4 and 5, neither $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type 1, nor $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type ω. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is not of type ∞, $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type 0. Thus, there exists a unifiable $\{1,2\}$-formula χ of type 0 in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, there exists a unifier υ of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, $\upsilon(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 = \mathbf{Triv}_1$, then by Lemma 7, $\mathtt{tr}_2^T(\upsilon(\chi)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, by Lemma 5, $\upsilon_2^T(\mathtt{tr}_2^T(\chi)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is a conservative extension of $\mathbf{L}_2$, it follows that $\upsilon_2^T(\mathtt{tr}_2^T(\chi)) \in \mathbf{L}_2$. Consequently, υ_2^T is an $\mathbf{L}_2$-unifier of $\mathtt{tr}_2^T(\chi)$. Since $\mathbf{L}_2$ is of type ∞, there exists a minimal $\mathbf{L}_2$-complete set Σ of $\mathbf{L}_2$-unifiers of $\mathtt{tr}_2^T(\chi)$. Since $L_1 = \mathbf{Triv}_1$, then by Proposition 3, Σ is a minimal complete set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, χ is not of type 0 in $\mathbf{L}_1 \otimes \mathbf{L}_2$: a contradiction.

[15] In the latter case, the proof can be similarly done.

Case $\mathbf{L}_2$ is of type ω: Thus, by Proposition 4, $\mathbf{L}_1 \otimes \mathbf{L}_2$ is not of type 1. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is not of type ω, either $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type 0, or $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type ∞. Consequently, there exists a unifiable $\{1,2\}$-formula χ either of type 0, or of type ∞ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, there exists a unifier υ of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, $\upsilon(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 = \mathbf{Triv}_1$, then by Lemma 7, $\mathtt{tr}_2^T(\upsilon(\chi)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, by Lemma 5, $\upsilon_2^T(\mathtt{tr}_2^T(\chi)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is a conservative extension of $\mathbf{L}_2$, we obtain $\upsilon_2^T(\mathtt{tr}_2^T(\chi)) \in \mathbf{L}_2$. Hence, υ_2^T is an $\mathbf{L}_2$-unifier of $\mathtt{tr}_2^T(\chi)$. Since $\mathbf{L}_2$ is of type ω, there exists a finite minimal $\mathbf{L}_2$-complete set Σ of $\mathbf{L}_2$-unifiers of $\mathtt{tr}_2^T(\chi)$. Since $L_1 = \mathbf{Triv}_1$, then by Proposition 3, Σ is a minimal complete set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, neither χ is of type 0 in $\mathbf{L}_1 \otimes \mathbf{L}_2$, nor χ is of type ∞ in $\mathbf{L}_1 \otimes \mathbf{L}_2$: a contradiction.

Case $\mathbf{L}_2$ is of type 1: Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is not of type 1, there exists a unifiable $\{1,2\}$-formula χ either of type 0, or of type ∞, or of type ω in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, there exists a unifier υ of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, $\upsilon(\chi) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 = \mathbf{Triv}_1$, then by Lemma 7, $\mathtt{tr}_2^T(\upsilon(\chi)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, by Lemma 5, $\upsilon_2^T(\mathtt{tr}_2^T(\chi)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is a conservative extension of $\mathbf{L}_2$, $\upsilon_2^T(\mathtt{tr}_2^T(\chi)) \in \mathbf{L}_2$. Thus, υ_2^T is an $\mathbf{L}_2$-unifier of $\mathtt{tr}_2^T(\chi)$. Since $\mathbf{L}_2$ is of type 1, there exists a finite minimal $\mathbf{L}_2$-complete set Σ of $\mathbf{L}_2$-unifiers of $\mathtt{tr}_2^T(\chi)$ with cardinality 1. Since $L_1 = \mathbf{Triv}_1$, then by Proposition 3, Σ is a minimal complete set of unifiers of χ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, neither χ is of type 0 in $\mathbf{L}_1 \otimes \mathbf{L}_2$, nor χ is of type ∞ in $\mathbf{L}_1 \otimes \mathbf{L}_2$, nor χ is of type ω in $\mathbf{L}_1 \otimes \mathbf{L}_2$: a contradiction. $\square$

In the above proof, the reader is invited to appreciate the uses of Proposition 3.

6 Specific results about the unification type of fusions

Let $\mathbf{L}_1$ be a consistent $\{1\}$-logic and $\mathbf{L}_2$ be a consistent $\{2\}$-logic. In Propositions 7 and 8, for all $m \geq 0$, σ_m and τ_m are the $\{1,2\}$-substitutions defined in Section 3 and φ and ψ are the $\{1,2\}$-formulas defined in Section 4.

Proposition 7. *If $\mathbf{L}_1$ and $\mathbf{L}_2$ are non-trivial extensions of $\mathbf{S}5$ then for all unifiers υ of $\varphi \wedge \psi$ in $\mathbf{L}_1 \otimes \mathbf{L}_2$, there exists $m \geq 0$ such that either $\sigma_m \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$, or $\tau_m \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$.*

Proof. Suppose $\mathbf{L}_1$ and $\mathbf{L}_2$ are non-trivial extensions of $\mathbf{S}5$. Let υ be a unifier of $\varphi \wedge \psi$ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, υ is a unifier of φ in $\mathbf{L}_1 \otimes \mathbf{L}_2$ and υ is a unifier of ψ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Let $m \geq 0$ be such that $\mathtt{deg}(\upsilon(x)) \leq 6m$. Suppose $\sigma_m \not\preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$ and $\tau_m \not\preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$. Since υ is a unifier of φ in $\mathbf{L}_1 \otimes \mathbf{L}_2$ and υ is a unifier of ψ in $\mathbf{L}_1 \otimes \mathbf{L}_2$, then by Lemma 29, $\upsilon(x) \to \boxplus^m \bot \notin \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg \upsilon(x) \to \boxminus^m \bot \notin \mathbf{L}_1 \otimes \mathbf{L}_2$. Since

$\mathbf{L}_1$ and $\mathbf{L}_2$ are non-trivial extensions of $\mathbf{S}5$, either there exists $\mathbf{kk} \geq 2$ such that $\mathbf{L}_1 = \mathbf{S}5_1^{\mathbf{kk}}$, or $\mathbf{L}_1 = \mathbf{S}5_1$ and either there exists $\mathbf{ll} \geq 2$ such that $\mathbf{L}_2 = \mathbf{S}5_2^{\mathbf{ll}}$, or $\mathbf{L}_2 = \mathbf{S}5_2$. Consequently, we have to consider the following cases:

1. there exists $\mathbf{kk} \geq 2$ such that $\mathbf{L}_1 = \mathbf{S}5_1^{\mathbf{kk}}$ and there exists $\mathbf{ll} \geq 2$ such that $\mathbf{L}_2 = \mathbf{S}5_2^{\mathbf{ll}}$,
2. there exists $\mathbf{kk} \geq 2$ such that $\mathbf{L}_1 = \mathbf{S}5_1^{\mathbf{kk}}$ and $\mathbf{L}_2 = \mathbf{S}5_2$,
3. $\mathbf{L}_1 = \mathbf{S}5_1$ and there exists $\mathbf{ll} \geq 2$ such that $\mathbf{L}_2 = \mathbf{S}5_2^{\mathbf{ll}}$,
4. $\mathbf{L}_1 = \mathbf{S}5_1$ and $\mathbf{L}_2 = \mathbf{S}5_2$.

The proof in Case (1) can be done as follows[16].

Remind that

- $\mathbf{S}5_1^{\mathbf{kk}}$ is a Kripke complete $\{1\}$-logic characterized by the class of all Kripke frames (W, R_1) where R_1 is an equivalence relation on W for which each equivalence class is a finite set of exactly $\mathbf{kk}$ possible worlds,
- $\mathbf{S}5_2^{\mathbf{ll}}$ is a Kripke complete $\{2\}$-logic characterized by the class of all Kripke frames (W, R_2) where R_2 is an equivalence relation on W for which each equivalence class is a finite set of exactly $\mathbf{ll}$ possible worlds,
- $\mathbf{S}5_1^{\mathbf{kk}}$ is characterized by the class of all Kripke frames (W, R_1) where R_1 is an equivalence relation on W for which each equivalence class is a finite set of at most $\mathbf{kk}$ possible worlds,
- $\mathbf{S}5_2^{\mathbf{ll}}$ is also characterized by the class of all Kripke frames (W, R_2) where R_2 is an equivalence relation on W for which each equivalence class is a finite set of at most $\mathbf{ll}$ possible worlds.

Since these classes of Kripke frames are closed under the formation of disjoint unions and isomorphic copies, then by [15, Theorem 4.1],

- $\mathbf{L}_1 \otimes \mathbf{L}_2$ is a Kripke complete $\{1, 2\}$-logic characterized both by the class $\mathcal{C}^{=}$ of all Kripke frames (W, R_1, R_2) where R_1 is an equivalence relation on W for which each equivalence class is a finite set of exactly $\mathbf{kk}$ possible worlds and R_2 is an equivalence relation on W for which each equivalence class is a finite set of exactly $\mathbf{ll}$ possible worlds,

[16] The proof of Cases (2)–(4) can be similarly done.

- $\mathbf{L}_1 \otimes \mathbf{L}_2$ is a Kripke complete $\{1,2\}$-logic characterized both by the class $\mathcal{C}^{\leq}$ of all Kripke frames (W, R_1, R_2) where R_1 is an equivalence relation on W for which each equivalence class is a finite set of at most **kk** possible worlds and R_2 is an equivalence relation on W for which each equivalence class is a finite set of at most **ll** possible worlds.

Since $v(x) \rightarrow \boxplus^m \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg v(x) \rightarrow \boxminus^m \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$,

- there exists a Kripke frame $\mathcal{F} = (W, R_1, R_2)$ in $\mathcal{C}^{=}$, there exists a model $\mathcal{M} = (W, R_1, R_2, V)$ based on $\mathcal{F}$ and there exists $t_0 \in W$ such that $\mathcal{M}, t_0 \models v(x) \wedge \neg \boxplus^m \bot$,
- there exists a Kripke frame $\mathcal{F}' = (W', R'_1, R'_2)$ in $\mathcal{C}^{=}$, there exists a model $\mathcal{M}' = (W', R'_1, R'_2, V')$ based on $\mathcal{F}'$ and there exists $t'_0 \in W'$ such that $\mathcal{M}', t'_0 \models \neg v(x) \wedge \neg \boxminus^m \bot$.

Now, let us transform $\mathcal{M}$ and $\mathcal{M}'$ into kinds of tree-like models without affecting satisfiability. An adaptation of the transformation called unravelling [7, Definition 4.51] will enable us to do this. We describe the transformation of $\mathcal{M}$ as follows[17]. A *t_0-tip in $\mathcal{M}$* is a tuple of the form $(u_0, a_1, u_1, \ldots, a_k, u_k)$ where $u_0 = t_0$, $k \geq 0$, $a_1, \ldots, a_k \in \{1,2\}$ and $u_1, \ldots, u_k \in W$ are such that

- for all $i \in \{1, \ldots, k\}$, $u_{i-1} R_{a_i} u_i$,
- for all $i \in \{1, \ldots, k\}$, $u_{i-1} \neq u_i$,
- for all $i \in \{2, \ldots, k\}$, $a_{i-1} \neq a_i$.

Let W'' be the set of all t_0-tips in $\mathcal{M}$. Notice that $(t_0) \in W''$. For all $i \in \{1,2\}$, let R''_i be the equivalence relation on W'' such that for all $(u_0, a_1, u_1, \ldots, a_k, u_k), (v_0, b_1, v_1, \ldots, b_l, v_l) \in W''$, $(u_0, a_1, u_1, \ldots, a_k, u_k) R''_i (v_0, b_1, v_1, \ldots, b_l, v_l)$ iff one of the following conditions holds:

- $(u_0, a_1, u_1, \ldots, a_k, u_k) = (v_0, b_1, v_1, \ldots, b_l, v_l)$,
- $k \geq 1$, $(u_0, a_1, u_1, \ldots, a_{k-1}, u_{k-1}) = (v_0, b_1, v_1, \ldots, b_l, v_l)$ and $a_k = i$,
- $l \geq 1$, $(u_0, a_1, u_1, \ldots, a_k, u_k) = (v_0, b_1, v_1, \ldots, b_{l-1}, v_{l-1})$ and $b_l = i$,
- $k \geq 1$, $l \geq 1$, $(u_0, a_1, u_1, \ldots, a_{k-1}, u_{k-1}) = (v_0, b_1, v_1, \ldots, b_{l-1}, v_{l-1})$, $a_k = i$, and $b_l = i$.

[17] The description of the transformation of $\mathcal{M}'$ can be similarly done.

Notice that for all $(u_0, a_1, u_1, \ldots, a_k, u_k) \in W''$, the equivalence class of $(u_0, a_1, u_1, \ldots, a_k, u_k)$ modulo R''_1 contains exactly **kk** elements and the equivalence class of $(u_0, a_1, u_1, \ldots, a_k, u_k)$ modulo R''_2 contains exactly **ll** elements. Moreover, the intersection of these equivalence classes is the singleton $\{(u_0, a_1, u_1, \ldots, a_k, u_k)\}$. Let V'' be the valuation on W'' such that for all atoms α, $V''(\alpha) = \{(u_0, a_1, u_1, \ldots, a_k, u_k) \in W'' : u_k \in V(\alpha)\}$. Let the *unravelling of* $\mathcal{M}$ *around* t_0 be the structure

- $\mathcal{M}'' = (W'', R''_1, R''_2, V'')$.

Similarly, let the *unravelling of* $\mathcal{M}'$ *around* t'_0 be the structure

- $\mathcal{M}''' = (W''', R'''_1, R'''_2, V''')$.

Let f'' be the function defined from W'' to W and associating to each t_0-tip $(u_0, a_1, u_1, \ldots, a_k, u_k)$ in W'' the possible world u_k in W. Similarly, let f''' be the function defined from W''' to W' and associating to each t'_0-tip $(u'_0, a'_1, u'_1, \ldots, a'_k, u'_k)$ in W''' the possible world u'_k in W'. Obviously, f'' is a bounded morphism from $\mathcal{M}''$ to $\mathcal{M}$ such that $f''((t_0)) = t_0$[18]. Similarly, obviously, f''' is a bounded morphism from $\mathcal{M}'''$ to $\mathcal{M}'$ such that $f'''((t'_0)) = t'_0$. Since $\mathcal{M}, t_0 \models v(x) \wedge \neg \boxplus^m \bot$, then by [7, Proposition 2.14], $\mathcal{M}'', (t_0) \models v(x)$ and $\mathcal{M}'', (t_0) \not\models \boxplus^m\bot$. Similarly, since $\mathcal{M}', t'_0 \models \neg v(x) \wedge \neg \boxminus^m \bot$, then by [7, Proposition 2.14], $\mathcal{M}''', (t'_0) \not\models v(x)$ and $\mathcal{M}''', (t'_0) \not\models \boxminus^m\bot$. Since $\mathcal{M}'', (t_0) \not\models \boxplus^m\bot$, there exists $t_{1,1}, t_{1,2}, t_{1,3}, t_{1,4}, t_{1,5}, t_{1,6}, \ldots, t_{m,1}, t_{m,2}, t_{m,3}, t_{m,4}, t_{m,5}, t_{m,6} \in W$ such that

- $t_0 R_1 t_{1,1} R_2 t_{1,2} R_1 t_{1,3} R_2 t_{1,4} R_1 t_{1,5} R_2 t_{1,6} \ldots R_1 t_{m,1} R_2 t_{m,2} R_1 t_{m,3} R_2 t_{m,4} R_1 t_{m,5} R_2 t_{m,6}$,
- $\mathcal{M}, t_0 \models p^1 \wedge q^0 \wedge r^1$, $\mathcal{M}, t_{1,1} \models p^0 \wedge q^0 \wedge r^0$, $\mathcal{M}, t_{1,2} \models p^0 \wedge q^0 \wedge r^1$, $\mathcal{M}, t_{1,3} \models p^0 \wedge q^1 \wedge r^0$, $\mathcal{M}, t_{1,4} \models p^0 \wedge q^1 \wedge r^1$, $\mathcal{M}, t_{1,5} \models p^1 \wedge q^0 \wedge r^0$, $\mathcal{M}, t_{1,6} \models p^1 \wedge q^0 \wedge r^1$, $\ldots$, $\mathcal{M}, t_{m,1} \models p^0 \wedge q^0 \wedge r^0$, $\mathcal{M}, t_{m,2} \models p^0 \wedge q^0 \wedge r^1$, $\mathcal{M}, t_{m,3} \models p^0 \wedge q^1 \wedge r^0$, $\mathcal{M}, t_{m,4} \models p^0 \wedge q^1 \wedge r^1$, $\mathcal{M}, t_{m,5} \models p^1 \wedge q^0 \wedge r^0$, $\mathcal{M}, t_{m,6} \models p^1 \wedge q^0 \wedge r^1$.

[18] To see this, notice that

- by the definition of the valuation V'' on W'', for all t_0-tips $(u_0, a_1, u_1, \ldots, a_k, u_k)$ in W'', $(u_0, a_1, u_1, \ldots, a_k, u_k)$ and u_k satisfy the same atoms,
- for all $i \in \{1, 2\}$, by the definition of the equivalence relation R''_i on W'', knowing that R_i is an equivalence relation on W, for all t_0-tips $(u_0, a_1, u_1, \ldots, a_k, u_k), (v_0, b_1, v_1, \ldots, b_l, v_l)$ in W'', if $(u_0, a_1, u_1, \ldots, a_k, u_k) R''_i (v_0, b_1, v_1, \ldots, b_l, v_l)$ then $u_k R_i v_l$,
- for all $i \in \{1, 2\}$, by the definition of the equivalence relation R''_i on W'', knowing that R_i is an equivalence relation on W, for all t_0-tips $(u_0, a_1, u_1, \ldots, a_k, u_k)$ in W'' and for all v in W, if $u_k R_i v$ then there exists a t_0-tip $(v_0, b_1, v_1, \ldots, b_l, v_l)$ in W'' such that $(u_0, a_1, u_1, \ldots, a_k, u_k) R''_i (v_0, b_1, v_1, \ldots, b_l, v_l)$ and $v_l = v$.

This implies that W'' contains the t_0-tip $\mathbf{tt} = (t_0, 1, t_{1,1}, 2, t_{1,2}, 1, t_{1,3}, 2, t_{1,4}, 1, t_{1,5}, 2, t_{1,6}, \ldots, 1, t_{m,1}, 2, t_{m,2}, 1, t_{m,3}, 2, t_{m,4}, 1, t_{m,5}, 2, t_{m,6})$. Similarly, since $\mathcal{M}''', (t'_0) \not\models \boxminus^m \bot$, there exists $t'_{1,1}, t'_{1,2}, t'_{1,3}, t'_{1,4}, t'_{1,5}, t'_{1,6}, \ldots, t'_{m,1}, t'_{m,2}, t'_{m,3}, t'_{m,4}, t'_{m,5}, t'_{m,6} \in W'$ such that

- $t'_0 R_2 t'_{1,1} R_1 t'_{1,2} R_2 t'_{1,3} R_1 t'_{1,4} R_2 t'_{1,5} R_1 t'_{1,6} \ldots R_2 t'_{m,1} R_1 t'_{m,2} R_2 t'_{m,3} R_1 t'_{m,4} R_2 t'_{m,5} R_1$ $t'_{m,6}$,
- $\mathcal{M}', t'_0 \models p^1 \wedge q^0 \wedge r^1$, $\mathcal{M}', t'_{1,1} \models p^1 \wedge q^0 \wedge r^0$, $\mathcal{M}', t'_{1,2} \models p^0 \wedge q^1 \wedge r^1$, $\mathcal{M}', t'_{1,3} \models p^0 \wedge q^1 \wedge r^0$, $\mathcal{M}', t'_{1,4} \models p^0 \wedge q^0 \wedge r^1$, $\mathcal{M}', t'_{1,5} \models p^0 \wedge q^0 \wedge r^0$, $\mathcal{M}', t'_{1,6} \models p^1 \wedge q^0 \wedge r^1$, $\ldots$, $\mathcal{M}', t'_{m,1} \models p^1 \wedge q^0 \wedge r^0$, $\mathcal{M}', t'_{m,2} \models p^0 \wedge q^1 \wedge r^1$, $\mathcal{M}', t'_{m,3} \models p^0 \wedge q^1 \wedge r^0$, $\mathcal{M}', t'_{m,4} \models p^0 \wedge q^0 \wedge r^1$, $\mathcal{M}', t'_{m,5} \models p^0 \wedge q^0 \wedge r^0$, $\mathcal{M}', t'_{m,6} \models p^1 \wedge q^0 \wedge r^1$.

Similarly, this implies that W''' contains the t'_0-tip $\mathbf{tt}' = (t'_0, 2, t'_{1,1}, 1, t'_{1,2}, 2, t'_{1,3}, 1, t'_{1,4}, 2, t'_{1,5}, 1, t'_{1,6}, \ldots, 2, t'_{m,1}, 1, t'_{m,2}, 2, t'_{m,3}, 1, t'_{m,4}, 2, t'_{m,5}, 1, t'_{m,6})$. Let $\mathcal{M}^{\cup} = (W^{\cup}, R_1^{\cup}, R_2^{\cup}, V^{\cup})$ be the model obtained from the disjoint union of $\mathcal{M}''$ and $\mathcal{M}'''$ by deleting all possible worlds in $R''_1(\mathbf{tt})$ but $\mathbf{tt}$ and by deleting all possible worlds in $R'''_2(\mathbf{tt}')$ but $\mathbf{tt}'$. Notice that consequently, $R_1^{\cup}(\mathbf{tt}) = \{\mathbf{tt}\}$ and $R_2^{\cup}(\mathbf{tt}') = \{\mathbf{tt}'\}$. Obviously, the Kripke frame $(W^{\cup}, R_1^{\cup}, R_2^{\cup})$ is in $\mathcal{C}^{\leq}$. Moreover, notice that in this frame, the length of the shortest path from (t_0) to $\mathbf{tt}$ is equal to $6m$ and the length of the shortest path from (t'_0) to $\mathbf{tt}'$ is equal to $6m$. Since $\mathtt{deg}(v(x)) \leq 6m$, $\mathcal{M}'', (t_0) \models v(x)$ and $\mathcal{M}''', (t'_0) \not\models v(x)$, $\mathcal{M}^{\cup}, (t_0) \models v(x)$ and $\mathcal{M}^{\cup}, (t'_0) \not\models v(x)$. Let $\mathcal{M}^{\uplus} = (W^{\uplus}, R_1^{\uplus}, R_2^{\uplus}, V^{\uplus})$ be the least model model obtained from $\mathcal{M}^{\cup} = (W^{\cup}, R_1^{\cup}, R_2^{\cup}, V^{\cup})$ by adding new states u_1, u_2, u_3, u_4 and u_5 such that

($\star$) $\mathbf{tt} R_1^{\uplus} u_1 R_2^{\uplus} u_2 R_1^{\uplus} u_3 R_2^{\uplus} u_4 R_1^{\uplus} u_5 R_2^{\uplus} \mathbf{tt}'$,

($\star\star$) $R_1^{\uplus}$ and $R_2^{\uplus}$ are reflexive and symmetric[19],

($\star\star\star$) $\mathcal{M}^{\uplus}, u_1 \models p^0 \wedge q^0 \wedge r^0$, $\mathcal{M}^{\uplus}, u_2 \models p^0 \wedge q^0 \wedge r^1$, $\mathcal{M}^{\uplus}, u_3 \models p^0 \wedge q^1 \wedge r^0$, $\mathcal{M}^{\uplus}, u_4 \models p^0 \wedge q^1 \wedge r^1$ and $\mathcal{M}^{\uplus}, u_5 \models p^1 \wedge q^0 \wedge r^0$.

Notice that consequently, $R_1^{\uplus}(\mathbf{tt}) = \{\mathbf{tt}, u_1\}$ and $R_2^{\uplus}(\mathbf{tt}') = \{\mathbf{tt}', u_5\}$. Obviously, the Kripke frame $(W^{\uplus}, R_1^{\uplus}, R_2^{\uplus})$ is in $\mathcal{C}^{\leq}$. Moreover, notice that in this frame, the length of the shortest path from (t_0) to $\mathbf{tt}$ is still equal to $6m$ and the length of the shortest path from (t'_0) to $\mathbf{tt}'$ is still equal to $6m$. Since $\mathtt{deg}(v(x)) \leq 6m$, $\mathcal{M}^{\cup}, (t_0) \models v(x)$ and $\mathcal{M}^{\cup}, (t'_0) \not\models v(x)$, we obtain $\mathcal{M}^{\uplus}, (t_0) \models v(x)$ and $\mathcal{M}^{\uplus}, (t'_0) \not\models v(x)$. Since v is a unifier of φ and v is a unifier of ψ, $v(x) \rightarrow \boxplus v(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg v(x) \rightarrow \boxminus \neg\sigma(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since the Kripke frame $(W^{\uplus}, R_1^{\uplus}, R_2^{\uplus})$ is in $\mathcal{C}^{\leq}$, $\mathcal{M}^{\uplus}, (t_0) \models v(x)$ and $\mathcal{M}^{\uplus}, (t'_0) \not\models v(x)$, it follows that $\mathcal{M}^{\uplus}, \mathbf{tt} \models v(x)$ and

[19]The transitivity of $R_1^{\uplus}$ and $R_2^{\uplus}$ is a consequence of the definition of $\mathcal{M}^{\uplus}$.

$\mathcal{M}^{\uplus}, \mathbf{tt}' \not\models \upsilon(x)$. Since $\upsilon(x) \to \boxplus\upsilon(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\neg\upsilon(x) \to \boxminus\neg\sigma(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and the Kripke frame $(W^{\uplus}, R_1^{\uplus}, R_2^{\uplus})$ is in $\mathcal{C}^{\leq}$, then by $(\star)$, $(\star\star)$ and $(\star\star\star)$, $\mathcal{M}^{\uplus}, \mathbf{tt} \not\models \upsilon(x)$ and $\mathcal{M}^{\uplus}, \mathbf{tt}' \models \upsilon(x)$: a contradiction.

This ends the proof of Proposition 7. $\square$

Proposition 8. *If* $\mathbf{L}_1$ *and* $\mathbf{L}_2$ *are non-trivial extensions of* $\mathbf{S}5$ *then* $\varphi \wedge \psi$ *is of type* 0 *in* $\mathbf{L}_1 \otimes \mathbf{L}_2$*.*

Proof. Suppose $\mathbf{L}_1$ and $\mathbf{L}_2$ are non-trivial extensions of $\mathbf{S}5$. Suppose $\varphi \wedge \psi$ is not of type 0 in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, there exists a minimal complete set Σ of unifiers of $\varphi \wedge \psi$ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. By Lemma 28, σ_0 is a unifier of $\varphi \wedge \psi$ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Since Σ is a minimal complete set of unifiers of $\varphi \wedge \psi$ in $\mathbf{L}_1 \otimes \mathbf{L}_2$, let $\upsilon \in \Sigma$ be such that $\upsilon \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma_0$. Thus, by Proposition 7, let $k \geq 0$ be such that either $\sigma_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$, or $\tau_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$. In the former case, by Lemma 28, σ_{k+1} is a unifier of $\varphi \wedge \psi$ in $\mathbf{L}_1 \otimes \mathbf{L}_2$. Since Σ is a minimal complete set of unifiers of $\varphi \wedge \psi$ in $\mathbf{L}_1 \otimes \mathbf{L}_2$, let $\upsilon' \in \Sigma$ be such that $\upsilon' \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma_{k+1}$. Since $\sigma_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$, then by Lemma 24, $\upsilon' \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$. Since Σ is a minimal complete set of unifiers of $\varphi \wedge \psi$ in $\mathbf{L}_1 \otimes \mathbf{L}_2$, $\upsilon' = \upsilon$. Since $\sigma_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$ and $\upsilon' \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma_{k+1}$, $\sigma_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma_{k+1}$: a contradiction with Lemmas 3 and 25. In the latter case, since $\upsilon \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma_0$, $\tau_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \sigma_0$: a contradiction with Lemmas 3 and 26. $\square$

Proposition 9. *If* $\mathbf{L}_1$ *and* $\mathbf{L}_2$ *are non-trivial extensions of* $\mathbf{S}5$ *then* $\mathbf{L}_1 \otimes \mathbf{L}_2$ *is of type* 0 *for unification with parameters.*

Proof. By Proposition 8. $\square$

7 Conclusion

After Propositions 4 and 5, it is natural to ask whether if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type ∞ then for all $i \in \{1,2\}$, $\mathbf{L}_i$ is either of type 1, or of type ω, or of type ∞. Unfortunately, we have not been able to answer this question, seeing that in Proposition 2, it is not clear that if the set Σ considered there is an infinite minimal complete set of unifiers then, when $L_{\bar{i}} \subseteq \mathbf{Triv}_{\bar{i}}$, or $L_{\bar{i}} \subseteq \mathbf{Verum}_{\bar{i}}$, the corresponding set among $\{\sigma_i^T : \sigma \in \Sigma\}$ and $\{\sigma_{\emptyset,i}^V : \sigma \in \Sigma\}$ is minimal complete too.[20]

[20]By the way, no modal logic (either unimodal, or multimodal) is known to be infinitary and it is also an open problem to determine if such modal logic exists.

Following the same line of reasoning as the one used in Section 6, other fusions such as $\mathbf{K}4 \otimes \mathbf{K}4$ and $\mathbf{S}4 \otimes \mathbf{S}4$ can also be proved to be nullary. See [28, Chapter 6]. The results obtained there as well as the results obtained in Section 6 lead us to the conjecture that every non-trivial fusion is of type 0, that is to say: if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is not of type 0 then either $\mathbf{L}_1 = \mathbf{Triv}_1$, or $\mathbf{L}_1 = \mathbf{Verum}_1$, or $\mathbf{L}_2 = \mathbf{Triv}_2$, or $\mathbf{L}_2 = \mathbf{Verum}_2$. These results also lead us to the conjecture that if either $\mathbf{L}_1$ is of type 0, or $\mathbf{L}_2$ is of type 0 then $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type 0[21]. By Propositions 4 and 5, this conjecture is equivalent to the one saying that if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type ∞ then for all $i \in \{1, 2\}$, $\mathbf{L}_i$ is either of type 1, or of type ω, or of type ∞.

Finally, Proposition 9 only constitutes a partial answer to Dzik's conjecture that the fusion $\mathbf{S}5 \otimes \mathbf{S}5$ of $\mathbf{S}5$ with itself is either nullary, or infinitary [11, Chapter 6], seeing that it is still unknown when $\mathbf{L}_1$ and $\mathbf{L}_2$ are non-trivial extensions of $\mathbf{S}5$ whether $\mathbf{L}_1 \otimes \mathbf{L}_2$ is of type 0 for elementary unification. In the case of elementary unification, what will play the role of the parameters p, q and r used in the formulas φ and ψ? What will play the role of the formulas φ and ψ? Is $\mathbf{S}5 \otimes \mathbf{S}5$ itself of type 0 for elementary unification?

References

[1] Baader, F., and S. Ghilardi, 'Unification in modal and description logics', *Logic Journal of the IGPL* **19** (2011) 705–730.

[2] Baader, F., and W. Snyder, 'Unification theory', In: *Handbook of Automated Reasoning,* Elsevier (2001) 439–526.

[3] Balbiani, P., 'Remarks about the unification type of several non-symmetric non-transitive modal logics', *Logic Journal of the IGPL* **27** (2019) 639–658.

[4] Balbiani, P., and Ç. Gencer, '**KD** is nullary', *Journal of Applied Non-Classical Logics* **27** (2017) 196–205.

[5] Balbiani, P., and Ç. Gencer, 'About the unification type of modal logics between **KB** and **KTB**', *Studia Logica* **108** (2020) 941–966.

[6] Balbiani, P., and M. Mojtahedi, 'Unification with parameters in the implication fragment of Classical Propositional Logic', *Logic Journal of the IGPL* (to appear).

[7] Blackburn, P., M. de Rijke, and Y. Venema, *Modal Logic,* Cambridge University Press (2001).

[21] In specific cases when among $\mathbf{L}_1$ and $\mathbf{L}_2$ one is either **K**, or **KB**, or **KD**, or **KDB**, or **KT**, or **KTB** and the other is characterized by a class of Kripke frames closed under the formation of disjoint unions and isomorphic copies, the line of reasoning developed in [20] for **K** and adapted in [3, 4, 5] for **KB**, **KD**, **KDB**, **KT** and **KTB** can be used to show that the fusion of $\mathbf{L}_1$ and $\mathbf{L}_2$ is of type 0. See [28, Chapter 6].

[8] Chagrov, A., and M. Zakharyaschev, *Modal Logic,* Oxford University Press (1997).

[9] Van Ditmarsch, H., W. van der Hoek, and B. Kooi, *Dynamic Epistemic Logic,* Springer (2008).

[10] Dzik, W., 'Unitary unification of **S**5 modal logics and its extensions', *Bulletin of the Section of Logic* **32** (2003) 19–26.

[11] Dzik, W., *Unification Types in Logic,* Wydawnicto Uniwersytetu Slaskiego (2007).

[12] Dzik, W., and P. Wojtylak, 'Projective unification in modal logic', *Logic Journal of the IGPL* **20** (2012) 121–153.

[13] Fagin, R., J. Halpern, Y. Moses, and M. Vardi, *Reasoning About Knowledge,* MIT Press (1995).

[14] Fine, K., G. Schurz, 'Transfer theorems for multimodal logics', In: *Logic and Reality. Essays on the Legacy of Arthur Prior,* Cambridge University Press (1991) 169–213.

[15] Gabbay, D., A. Kurucz, F. Wolter, and M. Zakharyaschev, *Many-Dimensional Modal Logics: Theory and Applications,* Elsevier Science (2003).

[16] Gencer, Ç., 'Description of modal logics inheriting admissible rules for **K**4', *Logic Journal of the IGPL* **10** (2002) 401–411.

[17] Gencer, Ç., and D. de Jongh, 'Unifiability in extensions of **K**4', *Logic Journal of the IGPL* **17** (2009) 159–172.

[18] Ghilardi, S., 'Best solving modal equations', *Annals of Pure and Applied Logic* **102** (2000) 183–198.

[19] Iemhoff, R., 'A syntactic approach to unification in transitive reflexive modal logics', *Notre Dame Journal of Formal Logic* **57** (2016) 233–247.

[20] Jeřábek, E., 'Blending margins: the modal logic **K** has nullary unification type', *Journal of Logic and Computation* **25** (2015) 1231–1240.

[21] Kost, S., 'Projective unification in transitive modal logics', *Logic Journal of the IGPL* **26** (2018) 548–566.

[22] Kracht, M., *Tools and Techniques in Modal Logic,* Elsevier (1999).

[23] Kracht, M., and F. Wolter, 'Properties of independently axiomatizable bimodal logics', *Journal of Symbolic Logic* **56** (1991) 1469–1485.

[24] Kurucz, A., 'Combining modal logics', In: *Handbook of Modal Logic,* Elsevier (2007) 869–924.

[25] Makinson, D., 'Some embedding theorems for modal logic', *Notre Dame Journal of Formal Logic* **12** (1971) 252–254.

[26] Nagle, M., 'The decidability of normal **K**5 logics', *Journal of Symbolic Logic* **46** (1981) 319–328.

[27] Nagle, M., and S. Thomason, 'The extensions of the modal logic **K**5', *Journal of Symbolic Logic* **50** (1985) 102–109.

[28] Rostamigiv, M., *About the Type of Modal Logics for the Unification Problem,* Doctoral thesis of the University of Toulouse 3 (2020).

[29] Rybakov, V., *Admissibility of Logical Inference Rules,* Elsevier (1997).

[30] Thomason, S., 'Independent propositional modal logics', *Studia Logica* **39** (1980) 143–144.

[31] Wolter, F., 'Fusions of modal logics revisited', In: *Advances in Modal Logic,* CSLI Publications (1998) 361–379.

Appendix

This Appendix includes the proofs of some of our results. Most of these proofs are relatively simple and we have included them here just for the sake of the completeness.

Proof of Lemma 1: The proof is done by induction on φ.

Proof of Lemma 2: The proof is done by induction on φ.

Proof of Lemma 3: Suppose $\mathbf{L}_1$ and $\mathbf{L}_2$ are non-trivial extensions of $\mathbf{S}5$.

Firstly, we prove that $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed. Since $\mathbf{S}5_1$ contains all $\{1\}$-formulas of the form $\varphi \to \Box_1 \Diamond_1 \varphi$ and $\mathbf{S}5_2$ contains all $\{2\}$-formulas of the form $\varphi \to \Box_2 \Diamond_2 \varphi$, $\mathbf{L}_1$ contains all $\{1\}$-formulas of the form $\varphi \to \Box_1 \Diamond_1 \varphi$ and $\mathbf{L}_2$ contains all $\{2\}$-formulas of the form $\varphi \to \Box_2 \Diamond_2 \varphi$. Hence, $\mathbf{L}_1 \otimes \mathbf{L}_2$ contains all $\{1, 2\}$-formulas of the form $\varphi \to \Box_1 \Diamond_1 \varphi$ and $\varphi \to \Box_2 \Diamond_2 \varphi$. Thus, $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed.

Secondly, we prove that $\mathbf{L}_1 \otimes \mathbf{L}_2$ is smooth. More precisely, we prove that for all $k, l \geq 0$, if $k > l$ then $\boxplus^k \bot \to \boxplus^l \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$[22]. Let $k, l \geq 0$. Suppose $k > l$. Let $\mathcal{M} = (W, R_1, R_2, V)$ be a model such that

- $W = \{i \geq 0 : \ 0 \leq i \leq 6l\}$,
- for all $i, j \in W$, $iR_1 j$ iff $\mid j - i \mid \leq 1$ and either $i = j$, or $\max\{i, j\}$ is odd,
- for all $i, j \in W$, $iR_2 j$ iff $\mid j - i \mid \leq 1$ and either $i = j$, or $\max\{i, j\}$ is even,
- $V(p) = \{i \geq 0 : \ 0 \leq i \leq 6l$ and either $i \mod 6 = 0$, or $i \mod 6 = 5\}$,
- $V(q) = \{i \geq 0 : \ 0 \leq i \leq 6l$ and either $i \mod 6 = 3$, or $i \mod 6 = 4\}$,
- $V(r) = \{i \geq 0 : \ 0 \leq i \leq 6l$ and either $i \mod 6 = 0$, or $i \mod 6 = 2$, or $i \mod 6 = 4\}$.

[22] The proof that for all $k, l \geq 0$, if $k > l$ then $\boxminus^k \bot \to \boxminus^l \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$ can be similarly done.

Obviously, $\mathcal{M}, 0 \models \boxplus^k \bot$ and $\mathcal{M}, 0 \models \neg \boxplus^l \bot$. Consequently, the Kripke frame (W, R_1, R_2) does not validate $\boxplus^k \bot \to \boxplus^l \bot$. In other respects, each equivalence class modulo R_1 contains exactly two possible worlds and each equivalence class modulo R_2 contains exactly two possible worlds. Hence, the Kripke frame (W, R_1, R_2) validates $\mathbf{S5}_1^2 \otimes \mathbf{S5}_2^2$. Since the Kripke frame (W, R_1, R_2) does not validate $\boxplus^k \bot \to \boxplus^l \bot$, $\boxplus^k \bot \to \boxplus^l \bot \not\in \mathbf{S5}_1^2 \otimes \mathbf{S5}_2^2$. Since $\mathbf{L}_1$ and $\mathbf{L}_2$ are non-trivial extensions of $\mathbf{S5}$, $\mathbf{L}_1 \otimes \mathbf{L}_2 \subseteq \mathbf{S5}_1^2 \otimes \mathbf{S5}_2^2$. Since $\boxplus^k \bot \to \boxplus^l \bot \not\in \mathbf{S5}_1^2 \otimes \mathbf{S5}_2^2$, $\boxplus^k \bot \to \boxplus^l \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof of Lemma 4: The proof is done by induction on φ.

Proof of Lemma 5: The proof is done by induction on φ.

Proof of Lemma 6: The proof is done by induction on φ.

Proof of Lemma 7: The proof is done by induction on φ.

Proof of Lemma 8: The proof of Item (1) can be done as follows[23].

Suppose $i = 1$.

Suppose $\varphi \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\mathbf{L}_2 \subseteq \mathbf{Triv}_2$. Hence, $\mathbf{L}_1 \otimes \mathbf{L}_2 \subseteq \mathbf{L}_1 \otimes \mathbf{Triv}_2$. Since $\varphi \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\varphi \in \mathbf{L}_1 \otimes \mathbf{Triv}_2$. Thus, by Lemma 7, $\mathtt{tr}_1^T(\varphi) \in \mathbf{L}_1 \otimes \mathbf{Triv}_2$. Since $\mathbf{L}_1 \otimes \mathbf{Triv}_2$ is a conservative extension of $\mathbf{L}_1$, $\mathtt{tr}_1^T(\varphi) \in \mathbf{L}_1$.

Proof of Lemma 9: The proofs of Items (1)–(4) are left to the reader. The proofs of Items (5) and (6) are done by using the well-known fact that if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed then $\mathbf{L}_1 \otimes \mathbf{L}_2$ is closed under the rules $\frac{\neg\varphi \to \Box_1 \psi}{\neg\psi \to \Box_1 \varphi}$ and $\frac{\neg\varphi \to \Box_2 \psi}{\neg\psi \to \Box_2 \varphi}$.

Proof of Lemma 10: The proof is done by induction on k.

Proof of Lemma 11: The proof is done by induction on k.

Proof of Lemma 12: Suppose $\mathbf{L}_1 \otimes \mathbf{L}_2$ is smooth. Hence, $\boxplus^{k+1} \bot \to \boxplus^k \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^{k+1} \bot \to \boxminus^k \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $\boxplus^k \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^k \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof of Lemma 13: The proof is done by induction on k. Suppose for all $k' \geq 0$, if $k' < k$ then if $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed and smooth then for all $l \geq 0$, $\boxplus^{k'} \bot \vee \boxminus^l \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$.

[23]The proof of Item (2) can be similarly done.

Case $k = 0$: Suppose $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed and smooth. Let $l \geq 0$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is smooth, then by Lemma 12, $\boxminus^l \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, $\boxplus^k \bot \vee \boxminus^l \bot \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Case $k \geq 1$: Suppose $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed and smooth. Let $l \geq 0$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed and smooth, then by induction hypothesis, $(\boxplus^{k-1} \bot \vee \boxminus^{l+1} \bot) \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $(\neg \boxplus^{k-1} \bot \to \boxminus \boxminus^l \bot) \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is tensed, then by Lemma 9, $(\neg \boxminus^l \bot \to \boxplus \boxplus^{k-1} \bot) \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, $(\boxplus^k \bot \vee \boxminus^l \bot) \not\in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof of Lemma 14: The proof is done by induction on k. Suppose for all $k' \geq 0$, if $k' < k$ then $\boxplus^{<k'} x \wedge \boxplus^{k'} \bot \to \sigma_{k'}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^{<k'} \neg x \wedge \boxminus^{k'} \bot \to \neg \tau_{k'}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Case $k = 0$: This case is left to the reader.

Case $k \geq 1$: By induction hypothesis, $\boxplus^{<k-1} x \wedge \boxplus^{k-1} \bot \to \sigma_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^{<k-1} \neg x \wedge \boxminus^{k-1} \bot \to \neg \tau_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, $\boxplus \boxplus^{<k-1} x \wedge \boxplus \boxplus^{k-1} \bot \to \boxplus \sigma_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus \boxminus^{<k-1} \neg x \wedge \boxminus \boxminus^{k-1} \bot \to \boxminus \neg \tau_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $x \wedge \boxplus \boxplus^{<k-1} x \wedge \boxplus \boxplus^{k-1} \bot \to x \wedge \boxplus \sigma_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg x \wedge \boxminus \boxminus^{<k-1} \neg x \wedge \boxminus \boxminus^{k-1} \bot \to \neg x \wedge \boxminus \neg \tau_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, by Lemma 11, $\boxplus^{<k} x \wedge \boxplus^k \bot \to \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^{<k} \neg x \wedge \boxminus^k \bot \to \neg \tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof of Lemma 15: The proof is left to the reader.

Proof of Lemma 16: The proof is done by induction on k. Suppose for all $k' \geq 0$, if $k' < k$ then $\sigma_{k'}(x) \to \boxplus \sigma_{k'}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg \tau_{k'}(x) \to \boxminus \neg \tau_{k'}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Case $k = 0$: This case is left to the reader.

Case $k \geq 1$: By induction hypothesis, $\sigma_{k-1}(x) \to \boxplus \sigma_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg \tau_{k-1}(x) \to \boxminus \neg \tau_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, by Lemma 15, $\sigma_{k-1}(x) \to x \wedge \boxplus \sigma_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg \tau_{k-1}(x) \to \neg x \wedge \boxminus \neg \tau_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $\sigma_{k-1}(x) \to \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg \tau_{k-1}(x) \to \neg \tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, $\boxplus \sigma_{k-1}(x) \to \boxplus \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus \neg \tau_{k-1}(x) \to \boxminus \neg \tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\sigma_k(x) \to \boxplus \sigma_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg \tau_k(x) \to \boxminus \neg \tau_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\sigma_k(x) \to \boxplus \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg \tau_k(x) \to \boxminus \neg \tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof of Lemma 17: The proof is done by induction on k. Suppose for all $k' \geq 0$, if $k' < k$ then for all $l \geq 0$, if $k' \leq l$ then $\sigma_{k'}(x) \to \boxplus^l \bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and

$\neg\tau_{k'}(x) \to \boxminus^l\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Let $l \geq 0$. Suppose $k \leq l$.

Case $k = 0$: This case is left to the reader.

Case $k \geq 1$: Since $k \leq l$, $k-1 \leq l-1$ and by induction hypothesis, $\sigma_{k-1}(x) \to \boxplus^{l-1}\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg\tau_{k-1}(x) \to \boxminus^{l-1}\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, $\boxplus\sigma_{k-1}(x) \to \boxplus\boxplus^{l-1}\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus\neg\tau_{k-1}(x) \to \boxminus\boxminus^{l-1}\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $x \wedge \boxplus\sigma_{k-1}(x) \to \boxplus\boxplus^{l-1}\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg x \wedge \boxminus\neg\tau_{k-1}(x) \to \boxminus\boxminus^{l-1}\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, $\sigma_k(x) \to \boxplus^l\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg\tau_k(x) \to \boxminus^l\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof of Lemma 18: The proof is done by induction on k. Suppose for all $k' \geq 0$, if $k' < k$ then for all $l \geq 0$, if $k' \leq l$ then $\boxplus^{k'}\bot \wedge \sigma_l(x) \leftrightarrow \sigma_{k'}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^{k'}\bot \wedge \neg\tau_l(x) \leftrightarrow \neg\tau_{k'}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Let $l \geq 0$. Suppose $k \leq l$.

Case $k = 0$: This case is left to the reader.

Case $k \geq 1$: Since $k \leq l$, $k-1 \leq l-1$ and by induction hypothesis, $\boxplus^{k-1}\bot \wedge \sigma_{l-1}(x) \leftrightarrow \sigma_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^{k-1}\bot \wedge \neg\tau_{l-1}(x) \leftrightarrow \neg\tau_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, $\boxplus\boxplus^{k-1}\bot \wedge x \wedge \boxplus\sigma_{l-1}(x) \leftrightarrow x \wedge \boxplus\sigma_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus\boxminus^{k-1}\bot \wedge \neg x \wedge \boxminus\neg\tau_{l-1}(x) \leftrightarrow \neg x \wedge \boxminus\neg\tau_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $\boxplus^k\bot \wedge \sigma_l(x) \leftrightarrow \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^k\bot \wedge \neg\tau_l(x) \leftrightarrow \neg\tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof of Lemma 19: The proof is done by induction on k. Suppose for all $k' \geq 0$, if $k' < k$ then for all $l \geq 0$, if $k' \leq l$ then $\lambda_l(\sigma_{k'}(x)) \leftrightarrow \sigma_{k'}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\mu_l(\tau_{k'}(x)) \leftrightarrow \tau_{k'}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Let $l \geq 0$. Suppose $k \leq l$.

Case $k = 0$: This case is left to the reader.

Case $k \geq 1$: Since $k \leq l$, $k-1 \leq l$ and by induction hypothesis, $\lambda_l(\sigma_{k-1}(x)) \leftrightarrow \sigma_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\mu_l(\tau_{k-1}(x)) \leftrightarrow \tau_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, $x \wedge \boxplus^l\bot \wedge \boxplus\lambda_l(\sigma_{k-1}(x)) \leftrightarrow \boxplus^l\bot \wedge x \wedge \boxplus\sigma_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg x \wedge \boxminus^l\bot \wedge \boxminus\neg\mu_l(\tau_{k-1}(x)) \leftrightarrow \boxminus^l\bot \wedge \neg x \wedge \boxminus\neg\tau_{k-1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $\lambda_l(x \wedge \boxplus\sigma_{k-1}(x)) \leftrightarrow \boxplus^l\bot \wedge \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\mu_l(\neg x \wedge \boxminus\neg\tau_{k-1}(x)) \leftrightarrow \boxminus^l\bot \wedge \neg\tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $k \leq l$, by Lemma 17, $\lambda_l(\sigma_k(x)) \leftrightarrow \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\mu_l(\tau_k(x)) \leftrightarrow \tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof of Lemma 20: The proof is done by induction on k. Suppose for all $k' \geq 0$, if $k' < k$ then for all $l \geq 0$, if $k' \geq l$ then $\lambda_l(\sigma_{k'}(x)) \leftrightarrow \sigma_l(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and

$\mu_l(\tau_{k'}(x)) \leftrightarrow \tau_l(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Let $l \geq 0$. Suppose $k \geq l$.

Case $k = l$: This case is left to the reader.

Case $k \geq l+1$: Hence, $k-1 \geq l$ and by induction hypothesis, $\lambda_l(\sigma_{k-1}(x)) \leftrightarrow \sigma_l(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\mu_l(\tau_{k-1}(x)) \leftrightarrow \tau_l(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $x \wedge \boxplus^l \bot \wedge \boxplus \lambda_l(\sigma_{k-1}(x)) \leftrightarrow \boxplus^l \bot \wedge x \wedge \boxplus \sigma_l(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg x \wedge \boxminus^l \bot \wedge \boxminus \neg \mu_l(\tau_{k-1}(x)) \leftrightarrow \boxminus^l \bot \wedge \neg x \wedge \boxminus \neg \tau_l(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, $\lambda_l(x \wedge \boxplus \sigma_{k-1}(x)) \leftrightarrow \boxplus^l \bot \wedge \sigma_{l+1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\mu_l(\neg x \wedge \boxminus \neg \tau_{k-1}(x)) \leftrightarrow \boxminus^l \bot \wedge \neg \tau_{l+1}(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, by Lemma 18, $\lambda_l(\sigma_k(x)) \leftrightarrow \sigma_l(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\mu_l(\tau_k(x)) \leftrightarrow \tau_l(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof of Lemma 21: Suppose $\mathbf{L}_1 \otimes \mathbf{L}_2$ is smooth. Let $l \geq 0$. Suppose $k > l$. Let υ and θ be the $\{1,2\}$-substitutions defined as follows:

- $\upsilon(x) = \top$,
- for all variables y distinct from x, $\upsilon(y) = y$,
- $\theta(x) = \bot$,
- for all variables y distinct from x, $\theta(y) = y$.

By Lemma 14, $\boxplus^{<k} x \wedge \boxplus^k \bot \to \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^{<k} \neg x \wedge \boxminus^k \bot \to \neg \tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, $\boxplus^{<k} \upsilon(x) \wedge \boxplus^k \bot \to \upsilon(\sigma_k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^{<k} \neg \theta(x) \wedge \boxminus^k \bot \to \neg \theta(\tau_k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\upsilon(x) = \top$ and $\theta(x) = \bot$, then by Lemma 10, $\boxplus^k \bot \to \upsilon(\sigma_k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^k \bot \to \neg \theta(\tau_k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\mathbf{L}_1 \otimes \mathbf{L}_2$ is smooth and $k > l$, $\boxplus^k \bot \to \boxplus^l \bot \notin \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^k \bot \to \boxminus^l \bot \notin \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\boxplus^k \bot \to \upsilon(\sigma_k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\boxminus^k \bot \to \neg \theta(\tau_k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\upsilon(\sigma_k(x)) \to \boxplus^l \bot \notin \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg \theta(\tau_k(x)) \to \boxminus^l \bot \notin \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $\sigma_k(x) \to \boxplus^l \bot \notin \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg \tau_k(x) \to \boxminus^l \bot \notin \mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof of Lemma 22: Suppose $\mathbf{L}_1 \otimes \mathbf{L}_2$ is smooth. Let $l \geq 0$. Let υ and θ be the $\{1,2\}$-substitutions defined as follows:

- $\upsilon(x) = \top$,
- for all variables y distinct from x, $\upsilon(y) = y$,
- $\theta(x) = \bot$,
- for all variables y distinct from x, $\theta(y) = y$.

Received 30 May 2020

Since $\mathbf{L}_1\otimes\mathbf{L}_2$ is smooth, then by Lemma 12, $\boxplus^k\bot\not\in\mathbf{L}_1\otimes\mathbf{L}_2$ and $\boxminus^k\bot\not\in\mathbf{L}_1\otimes\mathbf{L}_2$. Since $\upsilon(x)=\top$ and $\theta(x)=\bot$, $\boxplus^k\bot\vee\neg\upsilon(x)\not\in\mathbf{L}_1\otimes\mathbf{L}_2$ and $\boxminus^k\bot\vee\theta(x)\not\in\mathbf{L}_1\otimes\mathbf{L}_2$. Hence, $\boxplus^k\bot\vee\neg x\not\in\mathbf{L}_1\otimes\mathbf{L}_2$ and $\boxminus^k\bot\vee x\not\in\mathbf{L}_1\otimes\mathbf{L}_2$. By Lemma 15, $\neg\tau_k(x)\to\neg x\in\mathbf{L}_1\otimes\mathbf{L}_2$ and $\sigma_l(x)\to x\in\mathbf{L}_1\otimes\mathbf{L}_2$. Since $\boxplus^k\bot\vee\neg x\not\in\mathbf{L}_1\otimes\mathbf{L}_2$ and $\boxminus^k\bot\vee x\not\in\mathbf{L}_1\otimes\mathbf{L}_2$, $\boxplus^k\bot\vee\neg\tau_k(x)\not\in\mathbf{L}_1\otimes\mathbf{L}_2$ and $\boxminus^k\bot\vee\sigma_l(x)\not\in\mathbf{L}_1\otimes\mathbf{L}_2$.

Proof of Lemma 23: By Lemma 20.

Proof of Lemma 24: By Lemma 23.

Proof of Lemma 25: Suppose $\mathbf{L}_1\otimes\mathbf{L}_2$ is smooth. Let $l\geq 0$. Suppose $k<l$. Suppose either $\sigma_k\preceq^{\{x\}}_{\mathbf{L}_1\otimes\mathbf{L}_2}\sigma_l$, or $\tau_k\preceq^{\{x\}}_{\mathbf{L}_1\otimes\mathbf{L}_2}\tau_l$. In the former case, let λ be a $\{1,2\}$-substitution such that $\sigma_k\circ\lambda\simeq^{\{x\}}\sigma_l$. Hence, $\lambda(\sigma_k(x))\leftrightarrow\sigma_l(x)\in\mathbf{L}_1\otimes\mathbf{L}_2$. Since $\mathbf{L}_1\otimes\mathbf{L}_2$ is smooth and $k<l$, then by Lemma 21, $\sigma_l(x)\to\boxplus^k\bot\not\in\mathbf{L}_1\otimes\mathbf{L}_2$. By Lemma 17, $\sigma_k(x)\to\boxplus^k\bot\in\mathbf{L}_1\otimes\mathbf{L}_2$. Thus, $\lambda(\sigma_k(x))\to\boxplus^k\bot\in\mathbf{L}_1\otimes\mathbf{L}_2$. Since $\sigma_l(x)\to\boxplus^k\bot\not\in\mathbf{L}_1\otimes\mathbf{L}_2$, $\lambda(\sigma_k(x))\leftrightarrow\sigma_l(x)\not\in\mathbf{L}_1\otimes\mathbf{L}_2$: a contradiction. In the latter case, let μ be a $\{1,2\}$-substitution such that $\tau_k\circ\mu\simeq^{\{x\}}\tau_l$. Consequently, $\mu(\tau_k(x))\leftrightarrow\tau_l(x)\in\mathbf{L}_1\otimes\mathbf{L}_2$. Since $\mathbf{L}_1\otimes\mathbf{L}_2$ is smooth and $k<l$, by Lemma 21, $\neg\tau_l(x)\to\boxminus^k\bot\not\in\mathbf{L}_1\otimes\mathbf{L}_2$. By Lemma 17, $\neg\tau_k(x)\to\boxminus^k\bot\in\mathbf{L}_1\otimes\mathbf{L}_2$. Hence, $\neg\mu(\tau_k(x))\to\boxminus^k\bot\in\mathbf{L}_1\otimes\mathbf{L}_2$. Since $\neg\tau_l(x)\to\boxminus^k\bot\not\in\mathbf{L}_1\otimes\mathbf{L}_2$, $\mu(\tau_k(x))\leftrightarrow\tau_l(x)\not\in\mathbf{L}_1\otimes\mathbf{L}_2$: a contradiction.

Proof of Lemma 26: Suppose $\mathbf{L}_1\otimes\mathbf{L}_2$ is smooth. Let $l\geq 0$. Suppose either $\sigma_k\preceq^{\{x\}}_{\mathbf{L}_1\otimes\mathbf{L}_2}\tau_l$, or $\tau_k\preceq^{\{x\}}_{\mathbf{L}_1\otimes\mathbf{L}_2}\sigma_l$. In the former case, let λ be a $\{1,2\}$-substitution such that $\sigma_k\circ\lambda\simeq^{\{x\}}\tau_l$. Hence, $\lambda(\sigma_k(x))\leftrightarrow\tau_l(x)\in\mathbf{L}_1\otimes\mathbf{L}_2$. By Lemma 17, $\sigma_k(x)\to\boxplus^k\bot\in\mathbf{L}_1\otimes\mathbf{L}_2$. Since $\mathbf{L}_1\otimes\mathbf{L}_2$ is smooth, then by Lemma 22, $\boxplus^k\bot\vee\neg\tau_l(x)\not\in\mathbf{L}_1\otimes\mathbf{L}_2$. Since $\lambda(\sigma_k(x))\leftrightarrow\tau_l(x)\in\mathbf{L}_1\otimes\mathbf{L}_2$, then $\lambda(\sigma_k(x))\to\boxplus^k\bot\not\in\mathbf{L}_1\otimes\mathbf{L}_2$. Thus, $\sigma_k(x)\to\boxplus^k\bot\not\in\mathbf{L}_1\otimes\mathbf{L}_2$: a contradiction. In the latter case, let μ be a $\{1,2\}$-substitution such that $\tau_k\circ\mu\simeq^{\{x\}}\sigma_l$. Consequently, $\mu(\tau_k(x))\leftrightarrow\sigma_l(x)\in\mathbf{L}_1\otimes\mathbf{L}_2$. By Lemma 17, $\neg\tau_k(x)\to\boxminus^k\bot\in\mathbf{L}_1\otimes\mathbf{L}_2$. Since $\mathbf{L}_1\otimes\mathbf{L}_2$ is smooth, then by Lemma 22, $\boxminus^k\bot\vee\sigma_l(x)\not\in\mathbf{L}_1\otimes\mathbf{L}_2$. Since $\mu(\tau_k(x))\leftrightarrow\sigma_l(x)\in\mathbf{L}_1\otimes\mathbf{L}_2$, $\neg\mu(\tau_k(x))\to\boxminus^k\bot\not\in\mathbf{L}_1\otimes\mathbf{L}_2$. Hence, $\neg\tau_k(x)\to\boxminus^k\bot\not\in\mathbf{L}_1\otimes\mathbf{L}_2$: a contradiction.

Proof of Lemma 27: The proof is done by induction on k. Suppose for all $k'\geq 0$, if $k'<k$ then for all $\mathbf{L}_1\otimes\mathbf{L}_2$-unifiers σ of φ, $\sigma(x)\to\boxplus^{<k'}\sigma(x)\in\mathbf{L}_1\otimes\mathbf{L}_2$ and for all $\mathbf{L}_1\otimes\mathbf{L}_2$-unifiers τ of ψ, $\neg\tau(x)\to\boxminus^{<k'}\neg\tau(x)\in\mathbf{L}_1\otimes\mathbf{L}_2$.

Case $k=0$: This case is left to the reader.

Received 30 May 2020

Case $k \geq 1$: Let σ be an $\mathbf{L}_1 \otimes \mathbf{L}_2$-unifier of φ and τ be an $\mathbf{L}_1 \otimes \mathbf{L}_2$-unifier of ψ. By induction hypothesis, $\sigma(x) \rightarrow \boxplus^{<k-1}\sigma(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg\tau(x) \rightarrow \boxminus^{<k-1}\neg\tau(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since σ is an $\mathbf{L}_1 \otimes \mathbf{L}_2$-unifier of φ and τ is an $\mathbf{L}_1 \otimes \mathbf{L}_2$-unifier of ψ, $\sigma(x) \rightarrow \boxplus^{k-1}\sigma(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg\tau(x) \rightarrow \boxminus^{k-1}\neg\tau(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\sigma(x) \rightarrow \boxplus^{<k-1}\sigma(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg\tau(x) \rightarrow \boxminus^{<k-1}\neg\tau(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\sigma(x) \rightarrow \boxplus^{<k-1}\sigma(x) \wedge \boxplus^{k-1}\sigma(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg\tau(x) \rightarrow \boxminus^{<k-1}\neg\tau(x) \wedge \boxminus^{k-1}\neg\tau(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, $\sigma(x) \rightarrow \boxplus^{<k}\sigma(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$ and $\neg\tau(x) \rightarrow \boxminus^{<k}\neg\tau(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

Proof of Lemma 28: By Lemma 16.

Proof of Lemma 29: Suppose υ is an $\mathbf{L}_1 \otimes \mathbf{L}_2$-unifier of φ. Let $k \geq 0$.

$(a) \Rightarrow (b)$: Suppose $\sigma_k \circ \upsilon \simeq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$. Hence, $\sigma_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$.

$(b) \Rightarrow (c)$: Suppose $\sigma_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$. Let υ' be a $\{1,2\}$-substitution such that $\sigma_k \circ \upsilon' \simeq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$. Thus, $\upsilon'(\sigma_k(x)) \leftrightarrow \upsilon(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. By Lemma 17, $\sigma_k(x) \rightarrow \boxplus^k\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, $\upsilon'(\sigma_k(x)) \rightarrow \boxplus^k\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\upsilon'(\sigma_k(x)) \leftrightarrow \upsilon(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\upsilon(x) \rightarrow \boxplus^k\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

$(c) \Rightarrow (a)$: Suppose $\upsilon(x) \rightarrow \boxplus^k\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since υ is an $\mathbf{L}_1 \otimes \mathbf{L}_2$-unifier of φ, by Lemma 27, $\upsilon(x) \rightarrow \boxplus^{<k}\upsilon(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\upsilon(x) \rightarrow \boxplus^k\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\upsilon(x) \rightarrow \boxplus^{<k}\upsilon(x) \wedge \boxplus^k\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$. By Lemma 14, $\boxplus^{<k}x \wedge \boxplus^k\bot \rightarrow \sigma_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, $\boxplus^{<k}\upsilon(x) \wedge \boxplus^k\bot \rightarrow \upsilon(\sigma_k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\upsilon(x) \rightarrow \boxplus^{<k}\upsilon(x) \wedge \boxplus^k\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\upsilon(x) \rightarrow \upsilon(\sigma_k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. By Lemma 15, $\sigma_k(x) \rightarrow x \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $\upsilon(\sigma_k(x)) \rightarrow \upsilon(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\upsilon(x) \rightarrow \upsilon(\sigma_k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\upsilon(\sigma_k(x)) \leftrightarrow \upsilon(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, $\sigma_k \circ \upsilon \simeq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$.

Suppose υ is an $\mathbf{L}_1 \otimes \mathbf{L}_2$-unifier of ψ. Let $k \geq 0$.

$(d) \Rightarrow (e)$: Suppose $\tau_k \circ \upsilon \simeq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$. Hence, $\tau_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$.

$(e) \Rightarrow (f)$: Suppose $\tau_k \preceq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$. Let υ' be a $\{1,2\}$-substitution such that $\tau_k \circ \upsilon' \simeq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$. Thus, $\upsilon'(\tau_k(x)) \leftrightarrow \upsilon(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. By Lemma 17, $\neg\tau_k(x) \rightarrow \boxminus^k\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, $\upsilon'(\neg\tau_k(x)) \rightarrow \boxminus^k\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\upsilon'(\tau_k(x)) \leftrightarrow \upsilon(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\neg\upsilon(x) \rightarrow \boxminus^k\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$.

$(f) \Rightarrow (d)$: Suppose $\neg\upsilon(x) \rightarrow \boxminus^k\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since υ is an $\mathbf{L}_1 \otimes \mathbf{L}_2$-unifier of ψ, by

Received 30 May 2020

Lemma 27, $\neg\upsilon(x) \to \boxminus^{<k}\neg\upsilon(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\neg\upsilon(x) \to \boxminus^{k}\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\neg\upsilon(x) \to \boxminus^{<k}\neg\upsilon(x) \wedge \boxminus^{k}\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$. By Lemma 14, $\boxminus^{<k}\neg x \wedge \boxminus^{k}\bot \to \neg\tau_k(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Hence, $\boxminus^{<k}\neg\upsilon(x) \wedge \boxminus^{k}\bot \to \upsilon(\neg\tau k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\neg\upsilon(x) \to \boxminus^{<k}\neg\upsilon(x) \wedge \boxminus^{k}\bot \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\neg\upsilon(x) \to \upsilon(\neg\tau_k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. By Lemma 15, $\neg\tau_k(x) \to \neg x \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Thus, $\upsilon(\neg\tau_k(x)) \to \neg\upsilon(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Since $\neg\upsilon(x) \to \upsilon(\neg\tau_k(x)) \in \mathbf{L}_1 \otimes \mathbf{L}_2$, $\upsilon(\tau_k(x)) \leftrightarrow \upsilon(x) \in \mathbf{L}_1 \otimes \mathbf{L}_2$. Consequently, $\tau_k \circ \upsilon \simeq^{\{x\}}_{\mathbf{L}_1 \otimes \mathbf{L}_2} \upsilon$.

 Received 30 May 2020

Proving Cleanthes Wrong

Laureano Luna

laureanoluna@yahoo.es

Abstract

Hume's famous character Cleanthes claims that there is no difficulty in explaining the existence of causal chains with no first cause since in them each item is causally explained by its predecessor. Relying on logico-mathematical resources, we argue for two theses: (1) if the existence of Cleanthes' chain can be explained at all, it must be explained by the fact that the causal law ruling it is in force, and (2) the fact that such a causal law is in force cannot explain the occurrence of the events in the chain. In order to perform (1), we manage to express in mathematical terms the intuitive idea that indefinitely delayed explanation is ultimately no explanation. In order to achieve (2), we identify a logical relation we can prove to be as strong as the causal relation at issue in the Cleanthes passage, according to a precise notion of strength of relations.
Keywords: cause; chain; autonomous causal chain; ungrounded causal chain; causal law; *L*-nomologically possible worlds.

1 Introduction

Let us remember the words of Hume's character Cleanthes [17, p. 59]:

> "Add to this, that in tracing an eternal succession of objects, it seems absurd to enquire for a general cause or first author. How can any thing, that exists from eternity, have a cause, since that relation implies a priority in time, and a beginning of existence?
>
> In such a chain, too, or succession of objects, each part is caused by that which preceded it, and causes that which succeeds it. Where then is the difficulty? But the WHOLE, you say, wants a cause. I answer, that the uniting of these parts into a whole, like the uniting of several distinct countries into one kingdom, or several distinct members into one body, is performed merely by an arbitrary act of the mind, and has no

> influence on the nature of things. Did I shew you the particular causes of each individual in a collection of twenty particles of matter, I should think it very unreasonable, should you afterwards ask me, what was the cause of the whole twenty. This is sufficiently explained in explaining the cause of the parts".

Note that Cleanthes is not necessarily rejecting as meaningless to ask for an explanation of the chain, his claim is that if every item in the chain is explained, then it makes no sense to ask in addition for an explanation of the chain as a whole: the explanation of every item in the chain amounts to an explanation of the chain. This is sometimes called the Hume-Edwards principle (see [31]): the explanation of every part of a whole explains the whole. We wish to remark that our argumentation is compatible with the Hume-Edwards principle. "Where then is the difficulty?" Though Cleanthes asks rhetorically the difficulty might be real: intuitively, if any item that acts as a ground or an explanation for another needs itself to be grounded or explained, there is no *ultimate* explanation or, equivalently, there is *ultimately* no explanation for any item in the chain. It is intuitive that delaying the explanation indefinitely cannot count as actually providing it.

We intend to express this intuition in logico-mathematical terms and prove it from extremely plausible assumptions.

We are assuming, as the context in which Cleanthes speaks suggests, that there is no ground or explanation for the causal chain external to the chain itself; we say that the chain is *autonomous*. As possible grounds for the chain we only have the items in it and the causal law taking from each item to the next. Cleanthes' claim is that there is no difficulty in explaining the existence of an *autonomous* beginningless causal chain because each item in it is explained by its predecessor.

Consider a causal chain ruled by a causal law. Supose the chain to have no first item and suppose it to be *autonomous*, that is, lacking any explanation, cause, ground or support exterior to the chain itself. Our first claim is that if the chain has any explanation at all, that explanation must be the fact that the causal law ruling it is is force or, equivalently, that the causal powers of the items of the chain are what they are; for this we argue in section 3. Our second claim is that the fact that a causal law is in force cannot bring any event into existence, hence does not serve as an explanation of any causal chain; we elaborate on this in section 4. Our conclusion in section 5 will be that if an autonomous ungrounded chain exists, it exists for no reason, that is, without any ultimate cause or explanation.

Focusing on our second question, it is intuitive that the fact that a causal law is in force, that is, the fact that certain events would cause some others if they occurred, can bring nothing into existence by itself. The law can only return outputs if it acts on some previously existing input or, equivalently, an event can only bring another into existence once it has occurred. However, one can ask whether the question is susceptible of rigorous logical treatment. We contend it is, at least for the type of problem posed by Cleanthes' words, in which the focus is on an isolated beginningless causal chain containing temporal slices of a world's history. We consider an ungrounded causal chain $S_L = (Q_L, <_L)$ whose set of items Q_L is endowed with a strict linear ordering $<_L$. So, the logical form of the causal law L ruling the chain should be that of a sequence of some kind of concatenated conditionals:

$$... \Rightarrow p_i \Rightarrow p_{i+1} \Rightarrow p_{i+2} \Rightarrow ...$$

where for each i, p_i is the assertion that q_i occurs at time t_i. It is intuitive that from such a sequence of conditionals alone no p_i follows unless for logical reasons unrelated to any causal relation (e.g. if p_i is a validity), which we can safely assume to never be the case when the q_i are worldly events. The problem is that in Cleanthes' context material implication is too weak a relation to represent the causal relation: sometimes $p_i \rightarrow p_j$ is true even if q_i does not cause q_j. We manage to identify a logical relation such that if the relation holds between p_i and p_j, then q_i, if it occurred, would cause q_j; that is, we find a logical relation that is *in that sense* (see definition 2.iv. below) at least as strong as the causal relation involved in Cleanthes' words.

As we identify a logical relation among propositions that we can prove to be at least as strong as the causal relation and show that the fact that such a relation obtains does not entail the occurrence of any member of a certain set of events, we show that a causal law's being in force is also unable to explain the occurrence of such events.

2 Causality in Cleanthes' Universe: the Background of our Argument and its Connections to Some of the Recent Literature.

Cleanthes' universe is an isolate, i.e. context-free, and beginningless temporal chain of causes and effects, in which each member of the chain causes its immediate successor in time, if it exists, so being a sufficient condition for its occurrence given the causal laws in force, in such a way that the occurrence of a member of the chain together with the fact that the causal laws are such and such logically entails the occurrence of its immediate successor in time, if there is one. As the chain is isolated, we can think of it as the history of a universe or world and we can regard its members as the contents of infinitely many finite and disjunct time intervals whose mereological fusion yields exactly the history of the world.

Assuming this universe settles many disputed aspects of the nature of causality, thus rendering many controversies about the nature of causality largely irrelevant for our purpose: adicity of the causal relation, finer-grained vs. coarser-grained individuation of causes and effects, token vs. type causal relation, distinction between cause and background conditions, causes as sufficient conditions for their effects vs. causes as merely raising the probabilities of their effects, local vs. global nature of causal relation,[1] relationship between temporal and causal order, etc. We are aware of the subtleties surrounding the metaphysics of causation (see e.g. [12]) but we do believe that the very scenario we are dealing with settles most of the questions they pose.

The events that make up Cleanthes' causal chain are worldly events, slices of the universal history, hence it is not possible for them (or the propositions stating their occurrences) to logically entail other events of the chain (or the propositions stating their occurrences). In this, they are unlike the events in probabilistic spaces — which may be Boolean compounds of simpler events — or the Lewisian events [20] — which are properties of spatiotemporal regions and behave like classes, both of which can enter relations of logical implication. However, for the sake of the argument, we will allow not just events but also facts (e.g. the fact that a causal law is in force) as explanations *even if they are not members of the Cleanthian causal chain.*

For simplicity, we can assume that Cleanthes' universe is ruled by one causal law that endows each member of the causal chain with the power to bring its immediate successor into existence. We need not make substantial ontological claims concerning causal laws: the fact that a causal law is in force in Cleanthes' universe or actually rules it means only that the members of the causal chain have certain causal powers and not other. However, the assumption made by Cleanthes himself that the occurrence of a member of the chain must be able to furnish an explanation of the

[1] This aspect for instance is important for a number of questions (see e.g. [4]) but irrelevant here.

occurrence of its immediate successor forces upon us a conception of causality in which causal laws are able to provide explanations. This means that the fact that a causal law is in force in Cleanthes' universe cannot reduce to which the actual facts are in that universe because the statement of which the facts are cannot suffice as an explanation of them, unless we wish to strip the word "explanation" of its usual meaning. The description of the causal chain is by no means an explanation thereof. For A to count as a causal explanation of B, it is not enough if A is always followed by B as far as we know; we must attribute A the causal power to bring about B or equivalently we must believe that a causal law exists according to which A has the causal power to make B occur. Therefore, the need to understand causation as a form of explanation of occurrence or existence compels us to a nomological understanding of causality. This in turn renders the talk about nomologically possible worlds meaningful.

These will be our assumptions about causation and the causal relation, all of them necessary to fit the scenario put forward by Cleanthes. Whether our results here would stand if causation is so understood that a cause is not a sufficient condition for its effect may be of little interest, because so weak a notion of causation makes little sense in a scenario in which causes are proposed as the sources of complete explanations of their effects; indeed a cause that is not even a sufficient condition for its effect could hardly provide a complete explanation of it. Since we are interested precisely in figuring out whether ungroundedness prevents the existence of complete explanation, we eschew scenarios in which other circumstances would thwart it. Withal, we sketch in footnotes variants of our proofs showing that our result holds as well if we assume that all a cause has to do to act as such is raise the probability of its effect.

This seems to be the place to make some clarifications concerning the argument we develop in the following pages and its connections to recent literature, without any hope of being exhaustive. First of all, we do not think ours to be an *infinite regress argument*. Maurin [23, p. 422], drawing on Gratton [16], writes that infinite regress arguments typically have the following ingredients:

1. The assumptions necessary for the generation of an infinite regress.
2. First conclusion: the infinite regress.
3. The premises necessary to show that the first conclusion is unacceptable.
4. Second conclusion: the rejection of one or more of the assumptions in 1.

Therefore, such arguments use the existence of an infinite regress to refute some assumptions. In contrast, we start from a given infinite regress and argue for a claim about it but we do not use it to discard previous assumptions.

In particular, our argument has no essential relation to the claims made by Atkinson and Peijnenburg [1, 2], who argue for the possibility of probabilistic justification through an infinite regress of justifying items. These authors essentially rely on the fact that certain functions defined by ungrounded recursion may be well-defined but this fact has no obvious bearing on our argument.

It is also convenient to make a distinction between our argument and arguments for *fundamentalism* (e.g. [9, 21, 10, 15]; for contrary positions see e.g. [14, 24, 5], understood as the claim that the grounding relation must be well-founded. We do not argue for the impossibility that a grounding relation be non well-founded; we do not even argue for the narrower claim that the relation of causal grounding must be well-founded; we do not even argue against the existence of causal chains with no first cause. We just argue that if such chains exist (and are of the kind we deal here with, i.e. they are Cleanthian causal chains), there is no explanation for their existence.

The main intuition for the position that the grounding relation must be well-founded is — in Cameron's words [10, p. 3] — that:

> "... it is hard to see how things could get off the ground in the first place.",

Or as Schaffer 2010 (p. 37) puts it:

> "if one thing exists only in virtue of another, then there must be something from which the reality of the derivative entities ultimately derives.",

We think these intuitions involve an assumption that fundamentalists usually make, namely, that no part of reality can be unexplained; one can almost read the following implicit in Schaffer's words: 'otherwise there would be no explanation of why anything exists at all'. Fundamentalists typically argue from that assumption. We do not assume in this paper anything like that. Despite this, fundamentalists who argue *from* the assumption that nothing can be unexplained may profit from our argument that Cleanthian causal chains are unexplained, if they exist, to the end of arguing for the impossibility of their existence.

Rota [30], for instance, does argue for the same conclusion as we do: causal chains with no first cause are unexplained but his argument differs from ours in that Rota essentially relies on dismissing the Hume-Edwards Principle, since he takes the existence of an infinite causal chain as a complex fact on its own that requires explanation and cannot be explained by its constituent parts. Dumsday [15], on his way to conclude that certain causal chains (transitive ones) cannot lack first causes, argues as well for the claim that they could not provide causal explanation if they had no first causes. However, for this he essentially relies on the intuition that real explanation is ultimate explanation (see for this our section 5). Without denying either the appeal or the significance of that intuition, we intend to offer here a more rigorous logico-mathematical treatment of the question.

Support for the possibility of Cleanthian causal chains comes mostly from two philosophical positions. Firstly, the denial that everything must have an explanation, as in Russell's [32, p. 134] famous words:

> "I should say that the universe is just there, and that's all."

This position is widespread among those philosophers who regard Leibniz' Principle of Sufficient Reason and the related principles of causality either as suspect or as false as metaphysical principles. Many logical positivists were in this class; for example, Hans Reichenbach [28, p. 4] wrote:

> "According to the verifiability theory of meaning, which has been generally accepted for the interpretation of physics, the statement that there are causal laws therefore must be considered as physically meaningless."

Outside logical positivism a classical is Bunge [8], especially Part IV, but see also (in Bunge's footsteps) Romero [29]; also Brown [7, p. 525].

Secondly, support for the possibility of Cleanthian causal chains can rely on the belief that the Hume-Edwards Principle is enough to supply explanation for them. This position is held for instance by Russell himself [32, p. 134] and recently for instance by Paul Edwards [14]. As said above, this paper has nothing to say as regards the former position; however, it presents an argument against the latter since it argues against the possibility that Cleanthian causal chains have explanation and does it without rejecting the Hume-Edwards Principle.

There exist logico-mathematical arguments against the possibility of ungrounded causal chains, which — as already said — is not exactly our topic here, for instance, Craig and Sinclair 2012, [13]. Some of these arguments rely on *finitism* understood as the thesis that there can be no actual infinity of concrete entities (so Craig and Sinclair). Pruss [27] features a number of paradoxical situations with one common trait: in all of them infinitely many items are causally previous to some other item. The author harnesses those examples to build an argument for *causal finitism*, that is, the claim that such a causal structure is impossible. Our argument here does not rely on finitism or causal finitism nor employs any of the logico-mathematical resources these authors utilize. However, as we comment at the end of section 3, our first result could be seen as a paradox of causal infinitism.

Let us finally mention an attempt [22] at proving on logico-mathematical grounds that an ungrounded causal chain cannot exist because it cannot determine the content of its constituents. The argument is based on these two facts: it is impossible to determine a value by means of an infinitely regressing computation and from an infinite chain of conditionals no categorical proposition follows. Intuitive as this approach may be, it seems inconclusive because we have no certainty that an ungrounded causal chain would have to perform either a computation or a logical deduction to give determinate contents to its constituents.

3 First Claim: the Sole Possible Explanation.

We argue in this section for the claim that the occurrence of an autonomous beginningless causal chain can only be explained, if at all, by the fact that the causal law ruling it is in force.

Definition 1.

i. By a *strict countable chain* we understand a pair $S = (Q, <)$, where Q is a countable set with at least two members and $<$ is a strict total order on Q.

The order among the members of Q can be expressed by indices from a subset I of $\mathbb{Z}$, endowed with the natural order among the integers. Let q_B be the $<$-first member of Q, if it exists.

ii. A strict countable chain is *ungrounded* iff the corresponding relation $<$ is ungrounded, that is, iff

$$\forall j_{\in I} \exists i_{\in I} [q_i < q_j].$$

iii. Our mathematical representation of a *causal law* L is a relation holding between members of two non empty sets of events, namely, Q_C (the set of causes) and Q_E (the set of effects): $L \subseteq Q_C \times Q_E$.

iv. A *causal chain* $S_L = (Q_L, <_L)$ *ruled by a causal law* L is a strict countable chain of events in which for any $q_i, q_j \in Q_L$, $(q_i, q_j) \in L$ iff q_i is the immediate $<_L$-predecessor of q_j in Q_L.

v.We say that x explains the occurrence of event y iff x is a sufficient condition for the occurrence of y. We will assume that if x causes the occurrence of y according to L, then x and the fact that L is in force make up a sufficient condition for y's occurrence.

vi. A causal chain ruled by a causal law L is *autonomous* iff nothing external to the causal chain itself — Äthat is, nothing else than L or the members of Q_L — causes or explains the existence of the chain, not even partially.

Remarks of definition 1.

Chains are usually defined as totally ordered subsets of partially ordered sets: strictness and countability are usually not included. However, we wish to adapt our frame to the kind of chain Hume seems to have had in mind in the quote above. For brevity, we will omit the adjectives "strict" and "countable" from now on. The order relation $<_L$ in a causal chain being irreflexive, it will contain no pair like (q, q). As $<_L$ is transitive, this rules out causal loops. Causal loops merit specific discussion (see e.g. [27, p. 152]; [25]) but will not be considered here.

Note also that as the ordering relation $<$ in a chain has been defined as a total ordering in Definition 1.i, each causal chain can contain at most one independently given event, which we call q_B. In so doing, we restrict ourselves for simplicity to the simplest case but nothing substantial depends on this.

Note that we do not intend to define causation; this is beyond the point here; so, we take it as a primitive concept. ■

Let us now address our first claim. After providing an intuitive approach to the question we bring up, we will look for a more conclusive way of settling it and we will find a handy mathematical fact.

Consider an arbitrary item $q_i \in Q_L$ of an autonomous S_L. The causal production of q_i is explained both by the occurrence of q_{i-1} and the fact that q_{i-1} is able to cause the occurrence of q_i, that is, that the relevant part of L is in force. These are *distinguishable moments* -not *separable parts*- obtained by abstraction from one single act of causation. However, distinguishing different causal contributions of inseparable aspects of a cause is not unusual; for example, in collisions, the capacity of a moving ball to convey momentum to another ball at rest depends (among other things) on the mass of the moving ball and its velocity, although the latter is but a property of the former.

Note that if S_L is ungrounded, there is no item in Q_L that occurs independently of L's being in force, no item is given to L as a starting point from which it could bring forth a chain $q_B, q_1, q_2, ...$ If q_B existed, we would explain the chain from its occurrence and the fact that L is in force. If we remove q_B so that the chain is ungrounded, everything is produced through L out of no independent event. Note that only the items in Q_L and the law L are internal to S_L. This suggests that in an *autonomous ungrounded* chain S_L it must be the fact that L is in force that explains the chain, if anything at all. Let us see if we can approach this in more rigorous, mathematical terms.

Splitting the responsibility for the production of q_i into two different quantities, namely, what the occurrence of q_{i-1} contributes and what the fact that L is in force does, may seem to make little sense because both things are simultaneously necessary, namely, that q_{i-1} occurs and that it can in fact cause the occurrence of q_i. There is no such thing as *marginal productivity* here permitting to isolate the contribution of each factor because the two ingredients are simultaneously required in their integrity, fused in one single act of causation, for anything to be produced at all. Nevertheless, we can argue by other means for a quantitative approach to the respective contributions of L's being in force and the occurrence of q_{i-1} [2]. Consider that in a longer chain, for instance

$$q_{i-2}, q_{i-1}, q_i$$

instead of just

$$q_{i-1}, q_i$$

[2]For brevity, sometimes we will simply speak of the contributions of L and q_{i-1}.

where q_{i-1} is assumed to *not* depend on L (i.e. to be q_B), L's contribution must be larger because L is involved in producing the occurrence not just of q_i but of q_{i-1} itself, in such a way that the contribution of q_{i-1} to the occurrence of q_i is also partially due to L; thus, an additional contribution to the production of q_i must accrue to L for its involvement in the production of q_{i-1}.There must be two quantities k_1 and k_2, corresponding to L's contributions in the shorter and the longer chain, such that $k_2 > k_1$. We enunciate this as

Assumption 1.

If given some event q_i, exactly two conditions C_1 and C_2 can be distinguished that must be satisfied for q_i to occur and we consider two occasions O_1 and O_2, such that in O_1 no condition contributes to the fulfillment of the other condition and in O_2 C_2 does contribute to the fulfillment of C_1, then -*ceteris paribus*- the contribution of C_2 to the occurrence of q_i in O_2 increases with respect to its contribution in O_1 in a mathematically consistent way, that is to say, in such a way that if C_2 contributes to the fulfillment of C_1 k parts per unit, then k times the per unit value of the contribution of C_1 to the occurrence of q_i must be added to the per unit value of the contribution of C_2 to the occurrence of q_i. ■

It is commonsense that if something acts $n + 1$ times, then it contributes more than if it acts just n times; furthermore, the sole way we have to measure the difference is in real numbers and as a difference in the percent or per unit value of the total contribution.

The assumption is extremely general and we need to work out how it applies to our case, for it is Assumption 1 as applied to our particular case that yields the particular number established in Theorem 1 below.

Let the two-item and the three-item chains considered above be respectively O_1 and O_2. As in this case the contribution of L's being in force (which is our C_2 here) to the causation of the occurrence of q_{i-1} (which is C_1) is exactly the same as its contribution to the occurrence of q_i in O_1, the total contribution of L's being in force to the occurrence of q_i in O_2 must be computed by adding to its contribution in O_1 the same percentage of C_1's contribution in O_1 that was used to compute C_2's contribution in O_1. This generalizes in the obvious way to n-item chains and attains a limit value for an ungrounded chain, as we show next.

Let us consider an ungrounded S_L and let us *initially* say that L contributes a portion $0 < x_0 < 1$ of the total explanation of the occurrence of q_i and $1 - x_0$ is the part due to q_{i-1}; as q_{i-1} itself is explained by L according to the same *ratio* (for in ungrounded chains all items occupy structurally identical *loci*), we must add $(1-x_0)x_0$ to x_0 to approach the real share of L; we get $-x_0^2+2x_0$; then such is also the part of explanation of q_{i-1} due to L and we must add $(1-(-x_0^2+2x_0))(-x_0^2+2x_0)$ to $-x_0^2+2x_0$; and so on. We are iterating the function

$$f(x) = x + (1-x)x = -x^2 + 2x$$

Starting from any $0 < x_0 < 1$, the iteration yields 1 in the limit, so that *ultimately* the explanation of the occurrence of q_i — hence of Q_L — is all due to L. The mathematical fact is elementary[3] but we provide below an easy proof.

Theorem 1.

Let $f(x) = -x^2 + 2x$ and $0 < x_0 < 1$. Then

$\lim_{n\to\infty} f^n(x_0) = 1.$

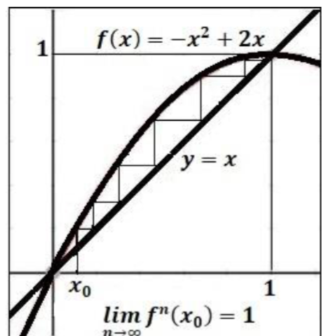


FIGURE 1

Proof. f^n is a dynamical system with seed x_0 such that

$f^0(x_0) = x_0$,
$f^{n+1}(x_0) = 2f^n(x_0) - f^n(x_0)^2 = f^n(x_0) + f^n(x_0) - f^n(x_0)^2$.

So, $f^{n+1}(x_0)$ adds $f^n(x_0) - f^n(x_0)^2$ to $f^n(x_0)$. As $0 < x_0 < 1$, $f^n(x_0) - f^n(x_0)^2 > 0$ up to $f^n(x_0) = 1$, where $f^n(x_0) - f^n(x_0)^2 = 0$, $f^{n+1}(x_0)$ adds 0 and f^n has a fixed point (FIGURE 1). As $(f^n(x_0))_{n\in\mathbb{N}}$ converges, it is a Cauchy sequence, hence it approaches a limit only as $f^{n+1}(x_0) - f^n(x_0)$ approaches 0, i.e. as $f^n(x_0)$ tends to 1 so that $f^n(x_0) - f^n(x_0)^2$ approaches 0.[4]

[3] Let k be L's contribution; whatever k is, the same part of the rest, i.e. $(1-k)k$, goes as well to L; so, $k = k + (1-k)k$ and $k = 1$. Whoever sees this as a correct approach can skip Theorem 1. This version of the theorem reveals that any number system for which the ordinary arithmetical operations are defined and that is such that for all $x > 0$, $x/x = 1$ will do the job (and not just the real numbers), since this is all that is required to extract that $k = 1$ from '$k = k + (1-k)k$'. Here '0' and '1' are respectively the additive identity and the multiplicative identity.

[4] Again the reader might wonder how much this result depends on the possibility of expressing

Remarks on theorem 1.

Theorem 1 shows that 1 minus L's contribution is less than ε for any $\varepsilon > 0$. This implies that L's contribution equals 1. Limits, however, are not values that are reached at infinity but values we can get arbitrarily close to. Though infinite summations (i.e. *series*) are usually thought of as limits, one does have the intuition that if all summands of a summation are in place, its value has to be reached. So, perhaps we can make the most of the fact that there are infinitely many q_i if S_L is ungrounded by making the theorem state the value of a series.[5]

We have seen that in each iteration of f we add $(1-x)x$ to x, where x is the sum of all previous results;[6] thus, the iteration of f carries over the sum of all previous results, so that if we define recursively

$$g(0) = x_0$$

$$g(n+1) = f^{n+1}(x_0) - f^n(x_0)$$

we have that

$$\lim_{n\to\infty} f^n(x_0) = \textstyle\sum_{n=0}^{\infty} g(n) = 1.$$

Let us explore how the summands $g(0), g(1), g(2), \ldots$ arise. Let the "$x \Rightarrow y$" denote the causation of y by x. As L partakes in $q_{i-1} \Rightarrow q_i$, we assign it the initial share $0 < x_0 < 1$ and we assign the rest, that is, $1 - x_0$, to q_{i-1}; however, as L partakes as well in $q_{i-2} \Rightarrow q_{i-1}$, we must transfer to L in addition a part of q_{i-1}'s share, namely, $(1-x_0)x_0 = -x_0^2 + x_0$, so that L's share amounts now to $-x_0^2 + 2x_0$ and q_{i-1}'s share is $1 - (-x_0^2 + 2x_0)$; however, as L partakes in $q_{i-3} \Rightarrow q_{i-2}$ too, the part accorded

partial contributions to total explanation as real numbers. For this, we refer the reader to footnote 3. Cleanthians may object to our modeling of the problem but as ours is the usual treatment of partial contributions to a total quantity, the burden of the proof that the modeling of this problem should be special lies with the Cleanthian.

[5]Some readers might suspect that taking limit could be a misguided attempt at cooking up a first cause as some sort of 'point at negative infinity'. We show below that the summands of the series arise from the items in the causal chain; if this is Cleanthian, hence actually infinite for any q_i, the summands must actually make up a series and yield its value.

[6]Proof for our f^n by induction on n. It is obvious for $f^1(x_0) = x_0 + (1-x_0)x_0$; assume it holds for $n = m$; then the argument of f^{m+1} is the sum of all results prior to f^m plus what f^m adds; call it k; then $f^{m+1}(k) = k + (1-k)k$.

to L in $q_{i-2} \Rightarrow q_{i-1}$ should not be just x_0 as previously assumed but $-x_0^2 + 2x_0$, hence the part accorded to L in $q_{i-1} \Rightarrow q_i$ should not be $-x_0^2 + 2x_0$ but in addition $(1 - (-x_0^2 + 2x_0))(-x_0^2 + 2x_0)$; and son on (FIGURE 2).

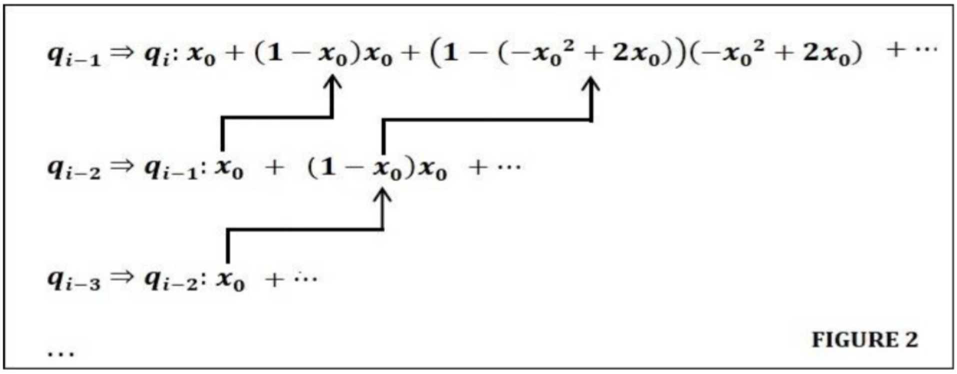


FIGURE 2

The deeper we go into the chain's tail, the greater L's contribution grows. The number of summands $g(0), g(1), g(2), ...$ depends on the number of items $<_L$-prior to q_i in Q_L; so, if S_L is ungrounded, the summation tops out at 1.[7]

The theorem implies that on Assumption 1 if an autonomous ungrounded causal chain S_L were to exist, its existence would be either unexplained or explained solely by L's being in force. The striking fact that in such a chain the contribution of the mere occurrence of q_{i-1} to the explanation of the occurrence of q_i vanishes may be taken as the mathematical expression of the fact that indefinitely delayed explanation is ultimately no explanation or as the mathematical expression of the old thesis that there is no cause at all if there is no first cause (Aristotle, Metaphysics, 994a10-19; see Aristotle [3, p. 37]. It can also be taken as a kind of paradox of infinity, since it actually looks paradoxical that though q_{i-1} is still causing q_i according to L, the contribution of the occurrence of q_{i-1} to the explanation of the occurrence of q_i is null. So interpreted, the situation may suggest an additional conclusion, other than the one we are arguing for here, namely, that autonomous ungrounded causal chains can only exist on pain of paradox. That point, however, is outside the scope of this paper.

[7]The reader may have noticed that the situation is different for dense causal chains because any such chain with more than one item involves always infinitely many of them. Note that definition 1.i excludes dense causal chains, which we deem to be incompatible with Cleanthes' universe.

4 Second Question: the Failure of the Sole Possible Explanation.

In this section, we argue for the claim that a causal law's being in force cannot explain on its own the existence of a causal chain. Let us clarify that the fact that L's being in force explains the occurrence of event q_i cannot mean that L becomes a cause that according to some other causal law $L*$ causes q_i. On the one hand, L is not the kind of thing we have admitted as possibly entering a causal relation. On the other hand, if we invoke $L*$, which is an element external to the chain S_L, then L's being in force is not enough to explain the occurrence of q_i: $L*$'s being in force, that is. an element external to the chain, would be required as well. If L's being in force alone is to explain the occurrence of q_i, it has to do so by conferring the causal powers it confers by being in force: it is the fact that these causal powers exist that has to explain the occurrence of q_i.

Now, which causal powers L bestows by being in force and upon which events it confers them is a matter of definition of L: it is essential to L's identity; L would not be L if it did not endow the events it endows with the causal powers it actually endows them with; so, which these powers are must be either explicit in each complete definition of L or logically entailed by this definition. In addition, such powers are defined by what they are powers to achieve. As a consequence, if L's being in force explains the occurrence of q_i, then L's being in force logically entails the occurrence of q_i, so that if we can prove for an arbitrary L that its being in force is logically compatible with q_i's not occurring, we are proving that no L is able — by being in force — of explaining the occurrence of q_i. And this is what we intend to do here.

It is convenient to keep in mind the following distinction: in the Cleanthian universe, the causal relation between say q_{i-1} and q_i is surely not logically necessary;[8] even if $(q_{i-1}, q_i) \in L$ and L is in force, it surely does not follow from any complete definition of q_{i-1} that it brings q_i into existence; it is indeed logically possible for the sort of causal laws acting in a Cleanthian universe that at some possible world q_{i-1} does not cause q_i; what comes next in Cleanthes' universe can hardly be a logical consequence of what is there currently the case. The point is ultimately that worldly events are not the kind of things that could enter relations of logical entailment. However,

[8]Some authors (e.g. [34, 6]) support the existence of *a priori* causal laws but not of the kind of the laws ruling causal chains in Cleanthian universes: to the best of our knowledge no author presently claims that the universe's unfolding in time may be just executing a logical deduction.

to repeat, the relation between L's being in force and the occurrence of q_i, if the former is to explain the latter, has to logically follow from any complete definition of L because what specifies a causal law is precisely the powers its being in force bestows and these powers are in turn defined by which things they are powers to achieve.

As we have already said, it is commonsensical that causal law L cannot — by just being in force — bring an event q_i into existence. By being in force L can endow certain events with certain causal powers; for instance, it can endow event q_{i-1} with the power to bring q_i into existence but the fact that L is in force does not ensure the occurrence of q_{i-1}, hence does not ensure the effective causation of q_i. As already suggested, what L's being in force guarantees is the truth of some conditional implying that if q_{i-1} occurs, then so does q_i but it does not entail any categorical statement that some q_i does occur. To be able to address the question formally, we should know what kind of conditionals causal laws consist of. This we do not claim to have figured out but we do claim to have found a logical relation[9] at least as strong as the causal relation, so that if our relation's being in force does not explain the occurrence of q_i, neither does the fact that the causal relation encapsulated in L is in force, whatever this relation actually is.

Definition 2.

i. By a *state-of-affairs*, we understand a function $\sigma : P \to \{T, F\}$, where P is a set of propositions[10] and T and F are the classical truth value , such that the truth value assignment made by σ is logically possible.

ii. Let P_1 and let P_2 be sets of propositions; P_2 is a *logical consequence* of P_1 (written as $P_1 \vDash P_2$) iff each state-of-affairs σ rendering all propositions in P_1 true — which we symbolize as "$\sigma \Vdash P_1$" — renders all propositions in P_2 true:

$$(P_1 \vDash P_2) \leftrightarrow_{def} \forall\sigma[(\sigma \Vdash P_1) \to (\sigma \Vdash P_2)].$$

We read "$P_1 \vDash P_2$" as "P_2 is a logical consequence of P_2" or "P_1 entails P_2".[11]

[9]Namely, the relation of contingent but L-nomologically necessary material implication; see Remarks on Lemma 3 below.

[10]Note that we think of propositions as semantic objects, not as closed formulas. See below.

[11]In this context, we usually replace singletons by their members and write e.g. "$\varphi \vDash \psi$", "$\sigma \Vdash \varphi$" instead of "$\{\varphi\} \vDash \{\psi\}$","$\sigma \Vdash \{\varphi\}$".

iii. Let P_1 and P_2 be two sets of propositions. We say that P_1 is *logically as strong* as P_2 iff

$$P_1 \vDash P_2$$

iv. In a derivative sense, we say that relation R_1 is *as strong as* relation R_2 iff

$$\forall xy[R_1(x,y) \vDash R_2(x,y)]$$

Of course, this implies that for all x, y

$$R_1(x,y) \rightarrow R_2(x,y).$$

Remarks on Definition 2.

The concept of logical possibility requires some comments.

Consider first that logical possibility may differ from metaphysical possibility: some logically possible states of affairs could be metaphysically impossible:[12] for instance, a state-of-affairs according to which things pop out of nowhere or time is reversible may be metaphysically impossible although it is logically possible since logic is not concerned with relations among events or other *concrete* objects.

As a consequence of the incompleteness of higher order logic, the concept of logical possibility cannot be given a syntactical definition. A mathematically precise semantical definition is available for formalized languages by means of the usual Tarskian notion of logical consequence: a sentence s is a logical consequence of a set S of sentences iff all models of S are models of $\{s\}$. A set S of sentences is logically possible iff it is consistent, that is, if no contradiction is a Tarskian logical consequence of S.

Nevertheless, Definition 2 is not about formal sentences but about propositions, which are semantical objects. We could attempt to define logical possibility of states-of-affairs in that context as follows: let $\sigma : P \rightarrow \{T, F\}$, let $P = (p_i)_{i \in I}$ and let $S = (s_i)_{i \in I}$ be a set of formal sentences such that for each $i_{\in I}$, s_i is the formalization of p_i. σ is logically possible iff the following set is consistent in the Tarskian sense:

$$S^* = \{s_i | \sigma(p_i) = T\} \cup [\neg s_i | \sigma(p_i) = F].$$

[12]We take metaphysical modality and the associated possible worlds as primitive concepts.

However, we would have to add that the formalizations of the p_i into the s_i must be such that all logical relations holding among the former hold also among the latter due to their logical form and this would surely have us circling around because Definition 2 defines logical (un)relatedness in terms of logical possibility.

So, we must take logical possibility as a primitive concept. This situation, however, poses no real problem for our arguments because the use they make of logical possibility is very restricted. All our arguments require is the following:

1. Propositions that are true by virtue of definitions are logically necessary.
2. Propositions describing the state of (sa Cleanthian universe at different times are logically unrelated (this we will call Assumption 2), which is intuitive indeed since worldly events are not the kind of things logic is about.

Condition 1 is uncontroversial since all it amounts to is the claim that tautological propositions are logically necessary. We enunciate condition 2 as

Assumption 2.

Propositions describing the state of a Cleanthian universe at different times are logically unrelated.

Lemma 1.

Transitivity of $\vDash$: $(P_1 \vDash P_2 \vDash P_3) \rightarrow (P_1 \vDash P_3)$

Proof.

Assume the antecedent "$P_1 \vDash P_2 \vDash P_3$". By definition 2.ii (of logical consequence), we have that for any state-of-affairs σ,

$$((\sigma \Vdash P_1) \rightarrow (\sigma \Vdash P_2)) \ \& \ ((\sigma \Vdash P_2) \rightarrow (\sigma \Vdash P_3))$$

By transitivity of material implication, for any σ

$$(\sigma \Vdash P_1) \rightarrow (\sigma \Vdash P_3))$$

which, by definition 2.ii (of logical consequence), is equivalent to $P_1 \vDash P_3$.

Definition 3.

Let P be a (non-empty) set of propositions. We say that the propositions in P are *logically unrelated* iff any truth-value assignment to them is logically possible. That p and q are logically unrelated in this sense can be expressed by this conjunction:

$$p \nvDash q \;\&\; p \nvDash \neg q \;\&\; \neg p \nvDash q \;\&\; \neg p \nvDash \neg q^{13}.$$

Definition 4.

If W is a set of possible worlds and P is a set of propositions, a *Kripke valuation function* $V : P \times W \to \{T, F\}$ is a function assigning each proposition in P a classical truth value from $\{T, F\}$ at each possible world in W. Kripke valuation functions are subject to certain rules in order to comply with logic and the meaning of modal operators. Instead of specifying them all here, we will specify the ones we need to in the proof of Lemma 4. As usual, we assume all possible worlds to be logically possible.

Remarks on Definition 4.

A pair $(p, w) \in P \times W$ can be interpreted as a proposition stating that proposition p is true at world w. As a consequence, each Kripke valuation function is a state-of-affairs as defined in Definition 2.i. This will be relevant for the proof of Lemma 4 below.

Lemma 2.

Let W be a set of possible worlds and P a set of propositions. If the members of P are logically unrelated, any Kripke valuation function $V : P \times W \to \{T, F\}$ is logically possible.

Proof.

A Kripke valuation function $V : P \times W \to \{T, F\}$ can be thought of as a collection of truth value assignments $(\sigma_i)_{i \in I}$ to the members of P in which the σ_i are paired to possible worlds w_j from $W = (w_j)_{j \in J}$:

$$V = \{(\sigma_{i_\alpha}, w_{j_\alpha}), (\sigma_{i_\beta}, w_{j_\beta}), \ldots\}$$

[13]The converses of the members of the conjunction follow from these by $\nvDash$-Contraposition and Double Negation: consider e.g. "$p \nvDash \neg q$", which implies that p is consistent with q, so that $q \nvDash \neg p$.

If the members of P are logically unrelated, then each σ_i is logically possible and then it is so with logical necessity. As all the w_j are logically possible, it is logically possible for any of the σ_i to obtain at any of the w_j. Thus, the entire V is logically possible.

Definition 5.

An L-*nomologically* possible world is any possible world in which the causal law L is in force iff it is in the actual world. We denote L-nomological necessity by the operator "$\Box_L$".

Definition 6.

We only introduce some notation.

i. Let "p_x" mean "event q_x occurs at time t_x" and let "$C(p_x, p_y)$" express the relationship holding between propositions p_x and p_y whenever $(q_x, q_y) \in L$ and L is in force. Now we can express the fact that a causal law L is in force (see Definition 1.iii) by

$$\forall ij_{\in I}[(q_i, q_j) \in L \rightarrow C(p_i, p_j)]$$

It follows from our considerations in section 2 that C is such that

$$\Box[C(p_i, p_j) \rightarrow (p_i \rightarrow p_j)],$$

where "$\Box$" is the metaphysical necessity operator; that is, the relation C between propositions induced by L's being in force is necessarily at least as strong (in the sense of Definition 2.iv) as the relation of material implication.

ii. Let $P_L = \{p_i \mid q_i \in Q_L\}$, that is, the set of all propositions asserting exactly the occurrence of a member of Q_L at its corresponding time.

Remarks on definition 6.

Note that L is a relation between certain events while C is a relation between the propositions asserting the occurrence of those events at certain times.

In order to isolate what L's being in force entails by itself from what the possible logical relations existing between the members of P_L could entail, it is useful to suppose that all members of P_L are *logically unrelated.* To see why this matters (see e.g. [18]), assume for instance that, for some pair $(p_i, p_j) \in C$, $p_i \vDash p_j$; then L's being in force would entail $p_i \rightarrow p_j$, for logical reasons alone, without any relation to any causal powers encapsulated in L. If L's being in force has to entail $p_i \rightarrow p_j$ due to the causal powers of some q_i, it has to entail it even if p_i and p_j are logically unrelated. For causal chains of worldly events, our assumption seems possible WLOG because worldly events are not the kind of things that can maintain logical relations. This is what renders Assumption 2 extraordinarily plausible.

Lemma 3.

Let L be a causal law and let C_L be the set of all conditionals of the form

$$p_i \rightarrow p_j$$

with one such conditional for any pair (p_i, p_j) such that $(q_i, q_j) \in L$. Let $C_{\Box L}$ be a set of conditionals of the form

$$\Box_L[p_i \rightarrow p_j]$$

with one such modal conditional for any pair (p_i, p_j) such that $(q_i, q_j) \in L$. Thus:

$$C_{\Box L} = \{x \mid \exists y[y \in C_L \ \& \ x = \Box_L[y]]\}$$

Let K be the proposition

$$\neg\Box[p_1 \rightarrow p_2]$$

for p_1, p_2 such that $(q_1, q_2) \in L$. Let $C_{\Box LK} = C_{\Box L} \cup \{K\}$ and let "F_L" denote the proposition that L is in force:

$$F_L = \text{“}\forall ij_{\in I}[(q_i, q_j) \in L \rightarrow C(p_i, p_j)]\text{”}$$

Accordingly,

$$C_{\Box LK} \vDash F_L.$$

Proof.

We first prove the mere conditional that if all members of $C_{\Box LK}$ are true, then L is in force. Assume the antecedent and that L is not in force for a *reductio.* As K is true, there is a possible world w' at which "$p_1 \rightarrow p_2$" is false; w' is L-nomologically impossible[14] because $\Box_L[p_1 \to p_2]$ is in $C_{\Box L}$ and has been assumed to be true. As L is not in force in the actual world, it is in force in all L-nomologically impossible worlds (by Definition 6), hence also in w'; since $(q_1, q_2) \in L$ and L is in force in w', it is true at w' that $(p_1, p_2) \in C$; now, by Definition 7.i, C is at least as strong as material implication; then "$p_1 \to p_2$" is true at w'; contradiction.

To see that the entailment holds, note that in the preceding part of the proof, F_L has been logically deduced from the truth of all members of $C_{\Box LK}$ and the premise that for any $(q_x, q_y) \in L$, if L is in force in some world w, then "$p_x \to p_y$" is true at w, which is analytical for our concept of causal law (see Definition 7) and any concept of causation admissible in Cleanthes' universe.[15]

Remarks on Lemma 3.

The fact that all propositions in the set $C_{\Box LK}$ are true is equivalent to some relation holding among the members of P_L : L-nomologically necessary but contingent material implication, which we by no means claim to exactly correspond to the causal relation but which we have shown to be as strong as this.

Let R_L be defined by

$$\forall xy[(R_L(x, y) \leftrightarrow_{def} (x, y) \in L \ \& \ C_{\Box LK})].$$

For those who believe that causal laws are metaphysically contingent and that causes are sufficient conditions for their effects, R_L could be a candidate for representing the relation of causality according to law L. In a nomological understanding of causation, a general binary causality relation K could then be defined by

$$\forall xy[K(x, y) \leftrightarrow_{def} \exists L[R_L(x, y)]].$$

Lemma 4.

Let C_L, $C_{\Box L}$, K, and $C_{\Box LK} = C_{\Box L} \cup \{K\}$ be as defined in Lemma 3. Let P_L be the set of all antecedents and consequents of the members of C_L (as in Definition 6.ii). Then

[14]Note that K being true together with all members of $C_{\Box L}$ requires L-nomologically necessary material implication to be *contingent*.

[15]If we do not believe "$C(p_i, p_j)$" should imply "$p_i \to p_j$", because we believe causes are not sufficient conditions for their effects — they just raise their probabilities, we can replace "$\neg\Box[p_1 \to p_2]$" in K with "$\Diamond[Pr(p_2) = 0]$", a probability assignment which is incompatible with $Pr(p_2|p_1) > Pr(p_2)$, so that we obtain all the same that $C_{\Box LK}$ entails F_L.

$$\forall\varphi_{\in P_L}[F_L \nvDash \varphi].$$

Proof.

We first prove that

$$\forall\varphi_{\in P_L}[C_{\Box LK} \nvDash \varphi].$$

The argument shows that for each φ in P_L there is a state-of-affairs rendering φ false and all members of $C_{\Box LK}$ true. As the members of P_L are logically unrelated, we can treat them as atomic propositions; thus, all we need now is modal propositional logic, where the states of affairs are represented by valuation functions in Kripke models (see above Remark on Definition 4). A Kripke model is a triple

$$M = \{W, R, V\},$$

where W is the set of all possible worlds, $R \subseteq W \times W$ is an accessibility relation, and $V : P_L \times W \to \{T, F\}$ is a Kripke valuation function (see Definition 4) assigning each φ in P_L a truth value at each $w \in W$. We will show that a Kripke model M^* exists that renders all members of $C_{\Box LK}$ true and all members of P_L false (at the actual world). Valuation functions in Kripke models for modal propositional logic are only subject to the usual recursive clauses for molecular formulae:

4.i. $V(\neg\varphi, w) = T$ iff $V(\varphi, w) = F$.
4.ii. $V(\varphi \vee \psi, w) = T$ iff $V(\varphi, w) = T$ or $V(\psi, w) = T$.
4.iii. $V(\Box\varphi, w) = T$ iff $\forall w'_{\in W}[R(w, w') \to V(\varphi, w') = T]$.

Let $M^* = \{W, R^*, V^*\}$; let $W_L \subseteq W$ be the set of all L-nomologically possible worlds; let V^* render all members of P_L false at all L-nomologically possible worlds:

$$\forall\varphi_{\in P_L}\ \forall w_{\in W_L}[V^*(\varphi, w) = F]$$

but let it be such that

$$\exists w'_{\in W}[V^*(p_1, w') = T \ \&\ V^*(p_2, w') = F],^{16,17}$$

where p_1 and p_2 are the propositions involved in K. Let R^* be such that

[16] Consider instead extending V^* by $V^*(Pr(p_2) = 0, w') = T$; whatever "p_2 has probability 0 at w'" means in terms of possible worlds (e.g. "p_2 is false at all worlds nomologically compatible with w'" or "p_2 is false at all worlds sufficiently similar to w'", etc.), by Lemma 2 it is logically compatible with p_2's being false at all L-nomologically possible worlds.

[17] Note that $w' \in W \setminus W_L$.

$$\forall w_{\in W}[R^*(w_A, w)],$$

where w_A is the actual world.

As the members of P_L are logically unrelated (as stated by Assumption 2), by Lemma 2, V^* is a logically possible state-of-affairs. V^* renders all members of P_L false and, by clauses 4.i, 4.ii, 4.iii, all members of $C_{\Box LK}$ true: it renders the members of $C_{\Box L}$ true by rendering all antecedents of the members of C_L false at all L-nomologically possible worlds, and it renders K — which was "$\neg\Box[p_1 \to p_2]$" — true by rendering "$p_1 \to p_2$" false at some $w' \in W$ such that $R^*(w_A, w')$.

Thus, V^* renders $C_{\Box LK}$ true though it makes all p_i false. From this and Definition 2.ii (of logical consequence), we have

$$\forall \varphi_{\in P_L}[C_{\Box LK} \not\models \varphi].$$

From Lemma 3, we have

$$C_{\Box LK} \models F_L.$$

Thus, by transitivity of $\models$ (Lemma 1), we obtain

$$\forall \varphi_{\in P_L}[F_L \not\models \varphi].$$

Remarks on Lemma 4.

Arguably, it is metaphysically impossible for all members of P_L to be false at all L-nomologically possible worlds if L is in force, since worldly events are commonly regarded as contingent and there seems to be no metaphysical incompatibility between L's being in force and the occurrence of the members of Q_L. It might be the case as well that causal laws are metaphysically necessary, so that $W = W_L$, and no such Kripke valuation function as V^* is metaphysically possible since it renders p_1 false at all L-nomologically possible worlds but true at some possible world. If so, the situation used to show that all members of $C_{\Box LK}$ may be true while all members of P_L are false would be *metaphysically* impossible. This, however, is irrelevant to our proof, which only requires that situation to be *logically* possible; that it is in fact so follows from Assumption 2.

Note that the lemma shows that it is possible for L to be in force even if none of the members of Q_L takes place. This — together with the assumption that the relation $C(p_x, p_y)$ is stronger than material implication — implies that the causal relation involves counterfactual conditional of the form "if x occurred, x would cause the occurrence of y".

Theorem 2.

The fact that a causal law is in force is unable by itself to cause any of the events in the chain it rules.

Proof.

Lemma 4 states that for all p_i in P_L, $F_L \nvDash p_i$; if L's being in force explains the occurrence of q_i, then it does so as a matter of definition, hence with logical necessity, so that $F_L \vDash p_i$. As a consequence, for all q_i in Q_L, the fact that L is in force does not explain the occurrence of q_i.

Remarks on Theorem 2.

It is an easy corollary of Lemma 4 that the relation $C(x, y)$ has a counterfactual nature, so that for any $ij_{\in I}$, L can be in force even if $(q_i, q_j) \in L$ but neither q_i nor q_j occurs, so that for all $i_{\in I}$ it is logically possible that F_L & $\neg p_i$. If we assume this as a premise, a much easier proof of Theorem 2 becomes possible: if L's being in force is a sufficient condition for the occurrence of q_i, then it is so as a *matter of definition*, so that $F_L \vDash p_i$ and it is logically impossible that F_L & $\neg p_i$. Thus, L's being in force is a sufficient condition for the occurrence of no q_i.

5 Conclusions.

Cleanthes claims that an autonomous ungrounded causal chain finds sufficient explanation for its existence in the fact that each item is caused by its immediate predecessor, that the existence of such a chain is thereby explained without difficulty. Theorem 1 entails that — on the very plausible Assumption 1 — this claim amounts to the contention that the existence of the chain is explained by the fact that the causal law ruling it is in force. This is, however, what Theorem 2 — on the equally plausible Assumption 2 — shows impossible. As a consequence, we can conclude that on very plausible assumptions, if autonomous ungrounded causal chains exist, they ultimately exist for no reason at all, they exist without ultimate explanation. So, contrary to what Cleanthes contends there is indeed a difficulty in explaining why they exist. The fact that every item in them is caused by an immediate predecessor is not enough to fill the explanatory gap.

In this context, it may be convenient to draw an expeditious distinction[18] between a scientific and a metaphysical sense of *explanation*. Empirical science usually understands that something is explained as soon as a cause of it according to some known causal law is discovered even if the cause and the causal law themselves remain unexplained. Scientific explanation is usually partial explanation. In contrast, metaphysical explanation (at least in the tradition of the Principle of Sufficient Reason, as it appears in Leibniz) requires *ultimate* or *complete* explanation. It is the latter kind of explanation that an autonomous ungrounded causal chain would lack even if each of its items could be given an explanation of the former kind.

Drawing a difference between partial and ultimate explanation by no means compromises our neutrality as regards the Hume-Edwards Principle. These are disparate questions. Note that we can indeed endorse the claim that it does not make sense to require explanation for a whole when there is available explanation for each of its parts (so supporting the Principle) while at the same time advocating the view that any real explanation of the parts must be complete or ultimate explanation.[19] In fact, we do not commit ourselves here to the claim that real explanation is ultimate explanation; we just draw the distinction between partial and ultimate explanation to clarify the kind of explanation our argument is about.

Thus, have we proven Cleanthes wrong? Almost.

Acknowledgments

I wish to thank Jacobus Erasmus for meticulous reading followed by helpful advice about earlier versions of this article.

References

[1] Atkinson, D; Peijnenburg, J. 2008. Probabilistic justification and the regress problem. *Studia Logica* 89(3): 333-341.

[18] So, Cole [11, p. 18] writes that “a scientist does not think he has failed in explanation because he makes use of a theory about which further questions can be raised”. The literature on causality and causal explanation is humongous. I beg the reader’s pardon for daring to tackle this difference without offering a revision of the literature, which would transcend the scope and space of this article.

[19] One author who may have criticized the Principle without observing the distinction we propose is Pruss [27].

[2] Atkinson, D; Peijnenburg, J. 2014. Can infinite regress justify everything?. In J. Turri, P. Klein (eds.) *Ad Infinitum: New Essays on Epistemological Infinitism*, pp. 162-178. Oxford, Oxford University Press.

[3] Aristotle. 2007. *Metaphysics*. Translated by J. H. McMahon. Mineola (NY), Dover.

[4] Bigelow, J.; Pargetter, R. 1975. Metaphysics of causation. *Erkenntnis* 33(1): 89-119.

[5] Bohn, E. D. 2018. Indefinitely descending ground. In R. Bliss, G. Priest (eds.) *Reality and its Structure: Essays in Fundamentality*. Oxford, Oxford University Press, pp. 167-181.

[6] Bradley, D. 2017. A priori causal laws. *Inquiry* 60(4): 358-370.

[7] Brown, P. 1996. Infinite causal regression. *The Philosophical Review* 75(4): 510-525.

[8] Bunge, M. 1959. Causality. *The Place of the Causal Principle in Modern Science*. Cambridge (MA), Harvard University Press.

[9] Callender, C. 2001. Why be a fundamentalist, Reply to Schaffer. Presented at the Pacific APA, San Francisco, March 2001. Text available at PhilSci Archive: `http://philsci-archive.pitt.edu/215/`

[10] Cameron R. P. 2008. Turtles all the way down: Regress, priority and fundamentality. *The Philosophical Quarterly* 58(230): 1-14.

[11] Cole, R. 1974. Causality and sufficient reason. *The Review of Metaphysics* 28(1): 3-23.

[12] Collins, J.; Hall, N.; Paul, L. (eds.). 2004. *Causation and counterfactuals*. Cambridge (MA), The MIT Press.

[13] Craig, W. L.; Sinclair J. D. 2012. The kalam cosmological argument. In W. L. Craig and J. P. Moreland (eds.) *The Blackwell Companion to Natural Theology*, pp. 101–201. Oxford, Wiley-Blackwell.

[14] Edwards, P. 2008. A critique of the cosmological argument. In L. P. Pojman, M. Rea (eds.) *Philosophy of Religion. An Anthology*, 5th edition. Belmont (CA), Thomson Wadsworth, pp. 15-23.

[15] Dumsday, T. 2014. Can causal chains extend back infinitely? Entailment, determinism, and a cosmological argument. *Forum Philosophicum* 19(2): 193-208.

[16] Gratton, C. 2010. *Infinite Regress Arguments*. Dordrecht, Springer.

[17] Hume, D. 1779. *Dialogue concerning natural religion*. H. D. Aiken (ed.), New York, Hafner, 1966.

[18] Kim, J. 1973. Causes and Counterfactuals. *The Journal of Philosophy* 70(17): 570-572.

[19] Lewis, D.K. 1983. *Philosophical Papers. Volume 1*. Oxford, Oxford University Press.

[20] Lewis, D.K. 1986. *Philosophical Papers. Volume 2*. Oxford, Oxford University Press.

[21] Lowe, E. J. 2001. *The Possibility of Metaphysics: Substance, Identity, and Time*. Oxford, Clarendon Press.

[22] Luna, L. 2014. No successful infinite regress. *Logic and Logical Philosophy* 23(2): 189-201.

[23] Maurin, A-S. 2013. Infinite regress arguments. In C. Svennerlind, J. AlmÃďng, R. Ingthorsson (eds.) *Johanssonian Investigations*, pp. 421-438. Heusenstamm, Ontos Verlag.

[24] Orilia, F. 2009. Bradley's regress and ungrounded dependence chains: A reply to

Cameron. *Dialectica* 63(3): 333-341.

[25] Meyer, U. 2012. Explaining causal loops. *Analysis* 72(2): 259-264.

[26] Pruss, A.R. 1998. The Hume-Edwards principle and the cosmological argument. *International Journal for Philosophy of Religion* 43(3): 149-165.

[27] Pruss, A. R. 2006. *The Principle of Sufficient Reason. A Reassessment.* Cambridge, Cambridge University Press.

[28] Reichenbach, H. 1944. *Philosophic Foundations of Quantum Mechanics.* Mineola (NY), Dover, 1998.

[29] Romero, G. 2016. Sufficient reason and reason enough. *Foundations of Science* 21(3): 455-460.

[30] Rota, M. 2007. Infinite Causal Chains and Explanation. *Proceedings of the American Catholic Philosophical Association 81*: 109-122.

[31] Rowe, W.L. 1989. Two criticisms of the cosmological argument. In W.L. Rowe and W.J. Wainwright (eds.) *Philosophy of Religion: Selected Readings.* Second edition, pp. 142-156. New York, Harcourt Brace Jovanovic.

[32] Russell, B. 2004. *Why I am not a Christian.* London, New York, Routledge.

[33] Schaffer, J. 2010. Monism: The priority of the whole. *Philosophical Review* 119(1): 31-76.

[34] Sober, E. 2011. A priori causal models of natural selection. *Australasian Journal of Philosophy* 89(4): 571âĂŞ589.

Appendix

To see not just that a probabilistic L's being in force cannot bring into existence the members of Q_L but that it cannot even probabilistically explain them, consider the following scenario. Let $P_L = (p_i)_{i \in I}$ and $W_L = (w_{L_j})_{j \in J}$ Let U be a Kripke valuation function. Let us assume F_L, so that

$$\forall i_{\in I} \forall j_{\in J}[U(Pr(p_{i+1}|p_i) > Pr(p_{i+1}), w_{L_j}) = T].$$

The question is whether F_L is logically compatible with the following:

$$Pr(p_i|F_L) \leq Pr(p_i), \tag{1}$$

that is, with the proposition that L's being in force does not raise p_i's probability.

If L's being in force probabilistically explains q_i, then F_L and (1) are logically incompatible. To see that F_L and (1) are in fact logically compatible, consider that q_i could be an event q_i^* in the set Q_{L^*} of items of some other causal chain S_{L^*} ruled by another causal law L^*, so that even if the occurrence of q_{i-1} is a necessary condition for q_i to be caused in the worlds in which L is in force, it may not be a metaphysically necessary condition for q_i to be caused: q_i as q_i^* in Q_{L^*} could be caused in some L-nomologically impossible world by a different event q_{i-1}^*. Now, suppose that as expected in a Cleanthian scenario, L's being in force precludes L^*'s being in force and suppose further that

$$Pr(p_i^*|p_{i-1}^*) > Pr(p_i|p_{i-1}),$$

so that by preventing a greater rise in its probability, L's being in force in fact lowers p_i's probability. This is indeed logically possible and would render both F_L and (1) simultaneously true.

Received 3 September 2020

A Comparative Study of Assumption-based Argumentative Approaches to Reasoning with Priorities

Jesse Heyninck*
Technische Universität Dortmund
jesse.heyninck@tu-dortmund.de

Christian Strasser
Ruhr-Universität Bochum
christian.strasser@rub.de

Abstract

In this paper we study formal properties of approaches to the reasoning with prioritized defeasible assumptions. We focus on methods proposed in formal argumentation, more specifically in the context of assumption-based argumentation.

We systematically compare two approaches for handling conflicts: preference-based defeats and preference-based defeats extended with reverse defeat. We investigate under which conditions these approaches give rise to the same output. We study several meta-theoretical properties including argumentation theoretical properties (such as Dung's Fundamental Lemma and the consistency of extensions) and properties for nonmonotonic reasoning (such as Cautious Monotony and Cut) in a parametrized way, i.e., relative to specific constraints on the underlying deducability relation. Finally, we study the relationship between these approaches and preferred subtheories, a nonmonotonic reasoning formalism that is based on maximal consistent subsets of a totally ordered knowledge base.

In the parametrized setting we study different sub-classes of assumption-based argumentation frameworks. For instance, we identify a particularly well-behaved sub-class of argumentation-based frameworks for which the different conflict-handling mechanisms coincide, which give the same outcomes as preferred subtheories and for which core properties of nonmonotonic logic are valid.

We thank the anonymous reviewers for helpful comments on previous versions of this paper. The research of both authors was supported by a Sofja Kovalevskaja award of the Alexander von Humboldt-Foundation, funded by the German Ministry for Education and Research.

*The research of Jesse Heyninck was additionally supported by the Deutsche Forschungsgemeinschaft under grant KE 1413/11-1

1 Introduction

We reason with defeasible information nearly all the time, be it the weather report or a doctor's diagnosis. Moreover, defeasible information usually comes in varying degrees of reliability depending on many factors, such as the quality of the underlying method by means of which some information has been obtained or the trustworthiness of a person's testimony. Several models have been proposed in the context of non-monotonic logic to model this type of reasoning (see [3] for an overview). In this paper our main aim is to provide some meta-theoretic transparency by studying formal properties of some of these methods. For this, our main focus will mainly be on the paradigm of structured argumentation.

Assumption-based argumentation (ABA) [9] is a formal model of defeasible reasoning with strict rules and defasible assumptions. Sets of defeasible assumptions can be in conflict with one another. To represent and resolve such conflicts, a formal argumentation framework is constructed on the basis of the strict rules and defeasible assumptions. In particular, *argumentative attacks* represent conflicts between sets of assumptions. Preferences over the plausible assumptions can be used to refine conflicts between assumptions by turning attacks into defeats. The orthodox approach (see [23, 33, 31]) in argumentation to resolve conflicts using preferences is to allow defeats only when an attacking set is not less preferred than the attacked assumption.

Example 1. Björn is throwing a party. His best friends are Anni-Frid, Benny and Agnetha, in order of affection. Björn told you that if Agnetha is around, Anni-Frid and Benny only talk about business and so she will not invite Agnetha in case she invites Björn and Benny. Defeasible assumptions in this scenario are $\texttt{Anni-Frid}, \texttt{Benny}$, and $\texttt{Agnetha}$, expressing that the person in question is invited. They are ordered according to plausibility by $\texttt{Agnetha} \prec \texttt{Benny} \prec \texttt{Anni-Frid}$ given Björn's respective affection. We also know that $\texttt{Anni-Frid}, \texttt{Benny} \rightarrow \overline{\texttt{Agnetha}}$: he will not invite Agnetha if he invites Anni-Frid and Benny ($\overline{\texttt{Agnetha}}$ denotes the *contrary* of $\texttt{Agnetha}$, which expresses that $\texttt{Agnetha}$ is not acceptable).

Now any set of assumptions $\{\texttt{Agnetha}, \dots\}$ is attacked by $\{\texttt{Benny}, \texttt{Anni-Frid}\}$ (among others). The reason is that given $\texttt{Benny}$ and $\texttt{Anni-Frid}$ we can derive $\overline{\texttt{Agnetha}}$. It is also defeated since neither $\texttt{Benny}$ nor $\texttt{Anni-Frid}$ is less preferred than $\texttt{Agnetha}$.

Recently, ABA^+ was proposed in [16], where *reverse defeats* are added as a passive counterpart to *direct defeats*: if an assumption is attacked by a set of assumptions one of which is strictly less preferred than the attacked assumption, a reverse attack is initiated.

Example 2. Suppose in our previous example the order of affection is inverse: $\texttt{Anni-Frid} \prec \texttt{Benny} \prec \texttt{Agnetha}$. In this case $\{\texttt{Agnetha}\}$ is not (directly) defeated by $\{\texttt{Benny}, \texttt{Anni-Frid}\}$ since the assumptions $\texttt{Benny}$ and $\texttt{Anni-Frid}$ have lower priorities than $\texttt{Agnetha}$. On the other hand, $\texttt{Agnetha}$ reverse defeats $\{\texttt{Benny}, \texttt{Anni-Frid}\}$.

Absent reverse-defeats, in examples such as Example 2 conflicts between assumptions may not get tracked properly and as a consequence consistency may get violated when selecting assumptions. Avoiding such scenarios is one of the main motivations behind the notion of reverse-defeat.

In this contribution we offer a comparative study of the approach based on direct defeat and the approach based on reverse defeat. For instance, we investigate whether closing the given inference rules under contraposition (e.g., in the given example we could add the rules $\texttt{Benny}, \texttt{Agnetha} \rightarrow \overline{\texttt{Björn}}$ and $\texttt{Björn}, \texttt{Agnetha} \rightarrow \overline{\texttt{Benny}}$) leads to the same or similar outcomes.

We proceed as follows: In Section 2 we give the necessary background on ABA with direct defeats ($\mathrm{ABA}^{\texttt{d}}$) and reverse defeats ($\mathrm{ABA}^{\texttt{r}}$). In Section 3 we review some argumentation theoretical properties of the frameworks, such as consistency and the Fundamental Lemma and in Section 4 we investigate under which conditions these approaches give rise to the same output. In the second part of the paper we relate the approaches based on direct and reverse defeats to the study of nonmonotonic logic more generally. In Section 5, we offer a systematic study of properties of the resulting consequence relations, such as cautious monotonicity and cut. The study proceeds in a parametrized way, in the sense that the results are presented relative to specific constraints on the underlying deducibility relation or relative to specific argumentation semantics.

Argumentative approaches are closely related to reasoning with consistent subsets of a given knowledge base. In the final Section 6 we validate this claim by presenting a characterization theorem for preferred subtheories [10], a nonmonotonic reasoning formalism based on (maximally) consistent subsets of a given totally ordered knowledge base.

Altogether, with this paper we hope to enrich our understanding of prioritized assumption-based defeasible reasoning by comparing two paradigmatic approaches, by studying important nonmonotonic reasoning properties of these approaches, and by characterizing them in terms of reasoning with maximally consistent subsets.

ABA was chosen as a core system for this study for several reasons. First, it is one of the paradigmatic approaches to formal argumentation and offers an intuitive and well-behaved approach [38]. Second, ABA is related to other argumentative approaches in a transparent way. For instance, ABA can be seen a special case of ASPIC^{+} [33], a prominent argumentation-based formalism for reasoning with knowledge bases that can contain defeasible rules in addition to strict rules, as well as strict and defeasible premises. Moreover, in [24] it was shown that (absent priorities) ASPIC-like defeasible rules can be represented by means of defeasible assumptions within ABA via a translation, making ASPIC$^+$ (without priorities) a special case of ABA. ABA also has strong connections to abstract argumentation [14, 27] and logic-programming [12, 36, 27]. Finally, in [9] ABA has been shown to capture autoepistemic logic [30] and default logic [34].

2 Preliminaries

ABA, thoroughly described in [9], is a formal model on the use of plausible assumptions used "to extend a given theory" [9, p.70] unless and until there are good arguments for not using (some of) these assumptions.

Inferences in ABA are implemented in ABA by means of rules over a formal language. Furthermore, defeasible assumptions are introduced, together with a contrariness operator to express argumentative attacks. We adapt the definition from [16] for an ABA$^+$ assumption-based framework.

Definition 1 (Assumption-based framework (ABF))**.** *An* assumption-based framework *is a tuple of the form* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$*, where:*

- $\mathcal{L}$ *is a formal language given by a countable set of sentences* $\{A, A_1, \ldots, B, \ldots\}$.
- $\mathcal{R}$ *is a set of inference rules of the form* $A_1, \ldots, A_n \to A$ *or* $\to A$.
- $Ab \subseteq \mathcal{L}$ *is a non-empty, finite set of plausible assumptions.*
- $\overline{} : Ab \to \mathcal{L}$ *is a contrariness operator.*
- $\mathbb{V}$ *is a non-empty, finite set of elements (called* values*).*
- $\leq \subseteq \mathbb{V} \times \mathbb{V}$ *is a total preorder over the values.*[1]

[1]We note that we do not require anti-symmetry to hold for $\leq$. Due to the totality of $\leq$, this allows for equally preferred but still different values. Nevertheless, whether one uses total preorders or total orders is inconsequential for the results of this paper. One can easily see that for any of the semantics sem defined below (see Definition 7) and where $\mathtt{ABF}_\sim = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}_\sim, \leq_\sim, v_\sim)$ is the quotient framework to $\mathtt{ABF}$ (i.e., $\mathbb{V}_\sim = \{[v]_\sim \mid v \in \mathbb{V}\}$, $[v]_\sim = \{v' \mid v \leq v', v' \leq v\}$, $\leq_\sim = \{([v]_\sim, [v']_\sim) \mid v \leq v'\}$ and $v_\sim : A \mapsto [v(A)]_\sim$), $\mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}) = \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}_\sim)$ (where $\mathtt{x} \in \{\mathtt{f}, \mathtt{d}, \mathtt{r}\}$). Note that $\leq_\sim$ is a total order.

- *$\upsilon : Ab \to \mathbb{V}$ is a total function assigning values to the assumptions.*[2]

As usual, we denote by $\geq$ the inverse of $\leq$, and define $\alpha < \beta$ iff $\alpha \leq \beta$ and $\beta \not\leq \alpha$.

In some presentations of ABA, deductions are obtained from a set of strict premises $\Gamma \subseteq \mathcal{L}$, a set of assumptions $Ab \subseteq \mathcal{L}$ and a set of rules $\mathcal{R}$. Here we follow [16] and rewrite each strict premise $A \in \Gamma$ as an empty-bodied rule $\to A$ in the set $\mathcal{R}$.

Definition 2 ($\mathcal{R}$-deduction)**.** *Given* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$*, a set* $\Delta \subseteq \mathcal{L}$*, and a sentence* $A \in \mathcal{L}$*, an* $\mathcal{R}$-deduction *from* Δ *of* A*, written* $\Delta \vdash_{\mathcal{R}} A$*, is a finite labeled tree such that*

- *the root is labeled A,*
- *each leaf is labeled with an element in Δ or with the empty string ϵ,*
- *for each parent node labeled with B there is a rule $B_1, \ldots, B_m \to B \in \mathcal{R}$ such that its children node are labeled $B_1, \ldots, B_m$ (in case of a rule with empty body its only child is labeled with ϵ),*
- *Δ is the set of all labels of leaf nodes.*

Definition 3. *Given* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$ *and* $\Delta \subseteq \mathcal{L}$*, we define*

$$\mathrm{Cn}_{\mathcal{R}}(\Delta) = \{A \in \mathcal{L} \mid \textit{there is a } \Delta' \subseteq \Delta, \Delta' \vdash_{\mathcal{R}} A\}.$$

We will restrict attention to so-called *flat* ABFs, i.e. ABFs for which $Ab \cap Cn_{\mathcal{R}}(\Delta) = \Delta$ for any $\Delta \subseteq Ab$ (cf. [15, Definition 2.5]) (this restriction is also made in e.g. [9, 39, 16, 18]). The investigation of non-flat frameworks is left for future work. Clearly, an ABF whose rule base $\mathcal{R}$ contains no rules whose consequents are assumptions is flat.

Although we have defined the deducability relation $\vdash_{\mathcal{R}}$ over $\wp(\mathcal{L}) \times \mathcal{L}$, in order to define ABA-frameworks it is sufficient to consider the restriction of $\vdash_{\mathcal{R}}$ to $\wp(Ab) \times \mathcal{L}$. The more general form is, however, useful to define Contraposition (see Def. 9). Note that $\vdash_{\mathcal{R}}$ need not be monotonic in the antecedent as witnessed by Example 3. This non-monotonicity arises in view of a relevancy requirement that comes with Definition 2 (according to which only those formulas occur on the left side of $\vdash$ which are labels in a deduction tree).

[2] In [16], a preference order $\leq \subseteq Ab \times Ab$ is defined directly over the assumptions. It will, however, greatly increase readability to use values to express priorities in this paper. Clearly, these modes of expression are equivalent.

Example 3. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where $v(p) < v(r) < v(q)$, $\overline{p} = u$, $\overline{q} = v$, $\overline{r} = w$ and:

$$Ab = \{p, q, r\} \quad \mathcal{L} = \{p, q, r, u, v, w\} \quad \mathcal{R} = \{q \to w\}$$

Then $\{q\} \vdash_{\mathcal{R}} w$ but $\{p, q\} \not\vdash_{\mathcal{R}} w$ (since p does not label any leaves in the $\mathcal{R}$-deduction tree), while $w \in \mathrm{Cn}_{\mathcal{R}}(\{q\}) \cap \mathrm{Cn}_{\mathcal{R}}(\{p, q\})$. [3]

Remark 1. *Often derivability/deducability relations $\vdash$ are defined in a monotonic way: so if $\Delta \vdash A$ then also $\Delta \cup \Delta' \vdash A$. Clearly, a monotonic counterpart to $\vdash_{\mathcal{R}}$ is readily available (in the style of the* $\mathrm{Cn}_{\mathcal{R}}$*-operator from Definition 3 below). Our choice to define $\vdash_{\mathcal{R}}$ in a non-monotonic way is really a question of ease of presentation: it will come in handy when defining reverse-defeat (see Definition 6 and the follow-up Remark 2) and when defining contraposition (see Definition 9 and the discussion following it). For both, it is important to track that every element in the support Δ of $\Delta \vdash_{\mathcal{R}} A$ is really needed to derive A. To remove all potential doubts on the side of the reader, we provide Appendix A where we present alternative definitions based on a monotonic derivability relation and prove that it gives rise to the same semantic selections (Theorem 14).*

The total preorder $\leq$ in an ABF encodes a preference between some of the assumptions in Ab via the map $v : Ab \to \mathbb{V}$. The intuitive reading of the preferences is as follows: if $v(A) \leq v(B)$ then B is at least as preferred as A. In order to compare arguments (i.e. $\mathcal{R}$-deductions), the order $\leq$ needs to be lifted to $\wp(Ab) \times Ab$. The following Definitions 4–8 are relative to a given assumption-based framework $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$.

Definition 4 ($\leq$-minimal set)**.** *Where $\Delta \subseteq Ab$ is a set of assumptions, we define* $\min(\Delta) = \{A \in \Delta \mid$ *there is no* $B \in \Delta$ *such that* $v(B) < v(A)\}$.[4]

Definition 5 (Lifting of $<$)**.** *Given some $\Delta \cup \{A\} \subseteq Ab$, we define*

$$\Delta < A \qquad \textit{iff} \qquad \textit{for some } B \in \min(\Delta),\ v(B) < v(A)$$

Definition 6 (Attack, defeat, reverse defeat)**.** *Given $\Delta \cup \Theta \cup \{A\} \subseteq Ab$,*

[3] Note that our relevancy requirement is not the same as minimality of support which is sometimes assumed in structured argumentation. To see this we slightly alter the example by adding non-assumptive atoms q_1 and q_2 and the assumption q' to our language $\mathcal{L}$. Additionally we add $q \to q_1$, $q' \to q_2$, and $q_1, q_2 \to w$ to our set of rules $\mathcal{R}$ above. Now we have: $\{q\} \vdash_{\mathcal{R}} w$ and $\{p, q\} \vdash_{\mathcal{R}} w$ as before. However, we also have $\{q, q'\} \vdash_{\mathcal{R}} w$ which would not be the case if we were to require minimality of support.

[4] Since we assume a total order and thus $v(A) = v(B)$ for any $A, B \in \min(\Delta)$, one could alternatively define $\min(\Delta)$ as one of the minimal elements of Δ. See also Footnote 1.

- Δ attacks A *(with Δ') iff there is $\Delta' \subseteq \Delta$ such that $\Delta' \vdash_{\mathcal{R}} \overline{A}$.*
- Δ attacks Θ *iff Δ attacks some $A \in \Theta$.*
- Δ $\mathtt{d}$-defeats A *iff Δ attacks A with some Δ' such that $\Delta' \not< A$.*
- Δ $\mathtt{d}$-defeats Θ *iff Δ $\mathtt{d}$-defeats some $A \in \Theta$.*
- Δ $\mathtt{r}$-defeats Θ *iff Δ $\mathtt{d}$-defeats Θ, or for some $\Theta' \subseteq \Theta$ and $A \in \Delta$, Θ' attacks A with $\Theta' < A$.*

If some $\{A\}$ $\mathtt{r}$-defeats Θ we will also say A $\mathtt{r}$-defeats Θ.

Remark 2. *When defining Δ attacks A* with *Δ' we essentially make use of the non-monotonicity of $\vdash_{\mathcal{R}}$: every member of Δ' is relevant for deriving $\overline{A}$. This is especially important in the definition of $\mathtt{r}$-defeat. Consider for this again Example 3. Absent the relevancy requirement r were to $\mathtt{r}$-defeat $\{q, p\}$ because $\{p, q\} < r$ (since $p < r$) and $\overline{r} \in \mathrm{Cn}_{\mathcal{R}}(\{p, q\})$. However, p is not relevant in the derivation of $\overline{r}$ (while $\{q\} \vdash_{\mathcal{R}} \overline{r}$, we have $\{p, q\} \nvdash_{\mathcal{R}} \overline{r}$), only q is and $r < q$.*

In the context of ABFs for which all assumptions are equally preferred (i.e. an $\mathtt{ABF}$ for which $\leq\, = \mathbb{V}^2$),[5] attack coincides with $\mathtt{d}$-defeat, so we will sometimes write $\mathtt{f}$-*defeat* (where the $\mathtt{f}$ abbreviates *flat*) instead of *attack* to avoid confusion. From here on, $\mathrm{ABA}^{\mathtt{f}}$, $\mathrm{ABA}^{\mathtt{d}}$ and $\mathrm{ABA}^{\mathtt{r}}$ denote assumption-based argumentation using, respectively $\mathtt{f}$-, $\mathtt{d}$- and $\mathtt{r}$-defeats. We observe that for any $\mathtt{x} \in \{\mathtt{f}, \mathtt{d}, \mathtt{r}\}$, $\mathtt{x}$-defeat is *monotonic* on both sides of the defeat relation.

Fact 1. *Where $\mathtt{x} \in \{\mathtt{f}, \mathtt{d}, \mathtt{r}\}$, $\Delta', \Theta \subseteq Ab$ and $\Delta \subseteq \Delta'$:*

- *if Δ $\mathtt{x}$-defeats Θ then Δ' $\mathtt{x}$-defeats Θ*
- *if Θ $\mathtt{x}$-defeats Δ then Θ $\mathtt{x}$-defeats Δ'.*[6]

We also note that $\mathtt{r}$-defeats preserve conflicts between sets of assumptions, as witnessed by the following fact.

Fact 2. *If $\Delta \vdash_{\mathcal{R}} \overline{A}$ then either A $\mathtt{r}$-defeats Δ or Δ $\mathtt{d}$-defeats (and therefore also $\mathtt{r}$-defeats) A.*

Example 4 witnesses that a similar fact does not hold for $\mathtt{d}$-defeat.

The consequences of a given ABF are determined by the argumentation semantics.

[5] Clearly, if $\leq\, = \mathbb{V}^2$, $v(A) \leqslant v(B)$ and $v(B) \leqslant v(A)$ for every $A, B \in Ab$.

[6] This fact and other results in this section are proven in Appendix B.

Definition 7 (Argumentation semantics [9])**.** *Given some sets $\Delta, \Delta' \subseteq Ab$, we define for each* $\mathtt{x} \in \{\mathtt{f}, \mathtt{d}, \mathtt{r}\}$*:*

- Δ *is* x-conflict-free *iff* Δ *does not* x*-defeat itself.*
- Δ x-defends Δ' *iff for any* $\Delta'' \subseteq Ab$ *that* x*-defeats* Δ'*,* Δ x*-defeats* Δ''*.*
- Δ *is* x-admissible *iff* Δ *is* x*-conflict-free and* Δ x*-defends itself.*
- Δ *is* x-complete *iff* Δ *is* x*-admissible and* Δ *contains every* $\Delta' \subseteq Ab$ *it* x*-defends.*
- Δ *is* x-grounded *iff* Δ *is* $\subseteq$*-minimally* x*-complete.*
- Δ *is* x-preferred *iff* Δ *is* $\subseteq$*-maximally* x*-admissible.*
- Δ *is* x-stable *iff* Δ *is* x*-conflict-free and* Δ x*-defeats every* $A \in Ab \setminus \Delta$*.*

We will denote x*-conflict-free,* x*-admissible,* x*-complete,* x*-grounded,* x*-preferred resp.* x*-stable by* $\mathtt{x}$-cf*,* $\mathtt{x}$-adm*,* $\mathtt{x}$-comp*,* $\mathtt{x}$-grou*,* $\mathtt{x}$-pref*,* $\mathtt{x}$-stab*. For any semantics* $\mathsf{sem} \in \{\mathsf{cf}, \mathsf{adm}, \mathsf{comp}, \mathsf{grou}, \mathsf{pref}, \mathsf{stab}\}$*, we define* $\mathtt{x}$-$\mathsf{sem}(\mathtt{ABF})$ *as the set of all sets of assumptions in Ab that are* $\mathtt{x}$-sem*, as defined above.*

The (skeptical) consequence relations based on the various semantics from Definition 7 are defined as follows:

Definition 8. *For any* $\mathsf{sem} \in \{\mathsf{grou}, \mathsf{pref}, \mathsf{stab}\}$*, let* $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} A$ *iff* $A \in \mathrm{Cn}_{\mathcal{R}}(\Delta)$ *for every* $\Delta \in \mathtt{x}$-$\mathsf{sem}(\mathtt{ABF})$*.*

Remark 3. *In Definition 8, we restrict attention to the* x*-preferred,* x*-stable and* x*-grounded semantics. The reason why we omit* x*-complete and* x*-admissible semantics from our discussion is that (1) entailment relations for* x*-admissible semantics are trivial and non-informative*[7] *and (2) by Definition 7 entailment relations based on the* x*-complete semantics coincide with those based on* x*-grounded semantics (for skeptical consequence).*

Example 4. Björn wants to go out with his friends Agnetha (a), Benny (b) and Anni-Frid (f). Also, if Benny is with Anni-Frid, Benny does not want to go out with Agnetha ($f, b \to \overline{a}$). Furthermore, Björn likes Benny more then Anni-Frid,

[7]Notice that for any $\mathtt{x} \in \{\mathtt{f}, \mathtt{d}, \mathtt{r}\}$, $\emptyset$ is x-admissible. This means that A is a consequence of ABF (based on the x-admissible semantics) iff $\emptyset \vdash_{\mathcal{R}} A$.

($v(f) = 1$ and $v(b) = 2$) and Björn likes Agnetha more than Benny ($v(a) = 3$). Who should Björn take out? We have the following $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where:[8]

$$Ab = \{a, b, f\} \qquad \mathcal{L} = Ab \cup \overline{Ab} \qquad \mathcal{R} = \{f, b \to \overline{a}\}$$
$$\mathbb{V} = \{1, 2, 3\} \quad v(f) = 1,\ v(b) = 2,\ v(a) = 3$$

We have the following defeats and failures thereof:

- $\{f, b\}$ $\texttt{f}$-defeats $\{a\}$ (since $\{f, b\} \vdash_{\mathcal{R}} \overline{a}$),
- $\{f, b\}$ does not $\texttt{d}$-defeat $\{a\}$ (since $\{f, b\} < a$),
- $\{a\}$ $\texttt{r}$-defeats $\{f, b\}$ (since $\{f, b\}$ $\texttt{f}$-defeats $\{a\}$ and $\{f, b\} < a$)
- $\{a\}$ does not $\texttt{d}$-defeat $\{f, b\}$ (since a does not $\texttt{f}$-defeat $\{f, b\}$).

This results in:

- $\{f, b\}$ being $\texttt{f}$-grounded, $\texttt{f}$-preferred and $\texttt{f}$-stable.
- $\{a, b, f\}$ being $\texttt{d}$-grounded, $\texttt{d}$-preferred and $\texttt{d}$-stable.
- $\{a, f\}$ and $\{a, b\}$ being $\texttt{r}$-preferred. Note that there are no $\texttt{r}$-complete, $\texttt{r}$-grounded and $\texttt{r}$-stable extensions (since e.g. $\{a, f\}$ $\texttt{r}$-defends b (since b has no $\texttt{r}$-defeater, but adding b to $\{a, f\}$ would result in $\{a, b, f\}$ not being $\texttt{r}$-admissible).

This means that, for instance, $\texttt{ABF} \mid\!\sim^{\mathsf{pref}}_{\texttt{f}} f$ and $\texttt{ABF} \mid\!\sim^{\mathsf{grou}}_{\texttt{d}} a$ while $\texttt{ABF} \not\mid\!\sim^{\mathsf{pref}}_{\texttt{f}} a$ and $\texttt{ABF} \not\mid\!\sim^{\mathsf{pref}}_{\texttt{r}} f$.

Remark 4. *While for Dung's seminal abstract argumentation frameworks from [20], the set of $\subseteq$-maximal admissible sets of arguments is identical to the set of $\subseteq$-maximal complete sets of arguments, Example 4 shows that this does not in general hold for ABFs based on* $\texttt{r}$*-defeat. Indeed, notice that $\{a, f\}$ and $\{b, f\}$ are not* $\texttt{r}$*-complete since e.g. $\{a, f\}$* $\texttt{r}$*-defends the unattacked* b *yet $\{a, b, f\}$ is* not $\texttt{r}$*-complete since it is not* $\texttt{r}$*-conflict-free. This behaviour is caused by a failure of the so-called* Fundamental Lemma of Abstract Argumentation *for* ABA$^{\texttt{r}}$*, which is the subject of Section 3.*

[8]Unless mentioned otherwise, we will assume that when $\mathbb{V}$ consists of a set of natural numbers, $\leq$ is the canonical order over $\mathbb{V}$. Moreover, we will often slightly abuse notation by defining $\mathcal{L}$ as $Ab \cup \overline{Ab}$ (or similar), meaning that $\mathcal{L}$ consists of all members of Ab and for each member $A \in Ab$ there is a unique element $A' \notin Ab$ in $\mathcal{L}$ for which $\overline{A} = A'$.

		Contraposition
Consistency for d-cf	Ex 5	Thm 1
Consistency for r-cf	Thm 2	
DFL for d-def.	Thm 3	
DFL for r-def.	[17, Ex 10]	Thm 4

Table 1: Argumentation theoretic properties of $\mathrm{ABA}^{\mathtt{d}}$ and $\mathrm{ABA}^{\mathtt{r}}$.

3 Argumentation-Theoretic Properties

In this section we investigate some properties that are interesting from an argumentative perspective. In particular, we consider (1) the rationality postulate of consistency, known from structured argumentation [11] and (2) the Fundamental Lemma (DFL) [20] known from abstract argumentation. Table 1 summarizes the results of this section. The top line lists the assumptions on the ABF under which a claim (in the lefmost column) is proved or shown to fail (the empty column meaning that contraposition is not supposed to hold). The table refers to the positive results, or to counter-examples (the latter have a grey background).

3.1 $\mathrm{ABA}^{\mathtt{d}}$ and Conflict Preservation

In [11], several *rationality postulates* were proposed for structured argumentation systems. The only postulate proposed in [11] that doesn't trivially hold for flat $\mathrm{ABA}^{\mathtt{d}}$ and $\mathrm{ABA}^{\mathtt{r}}$ frameworks is the postulate of *consistency*:

> No set of assumptions Δ selected by a given semantics contains an assumption A for which $\overline{A}$ is derivable from some $\Delta' \subseteq \Delta$.[9]

One of the reasons for introducing reverse defeats in $\mathrm{ABA}^{\mathtt{r}}$ is to avoid violations of the postulate of consistency by preserving conflicts between assumptions even if the attacking assumptions are strictly less preferred than the attacked assumption. The following example shows that for $\mathrm{ABA}^{\mathtt{d}}$ conflicts are not necessarily preserved.

Example 5. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$ where:[10]

$$Ab = \{p, q\} \qquad \mathcal{L} = Ab \cup \overline{Ab} \qquad \mathcal{R} = \{p \to \overline{q}\}$$
$$\mathbb{V} = \{1, 2\} \quad \upsilon(p) = 1,\ \upsilon(q) = 2$$

[9] See Theorem 1 below for a formal statement.

[10] Here and in examples below, we let $\overline{Ab} = \{\overline{A} \mid A \in Ab\}$ and suppose that $Ab \cap \overline{Ab} = \emptyset$ and for all $A, B \in Ab$, if $A \neq B$ then $\overline{A} \neq \overline{B}$.

Note that $\{p\}$ does not $\mathtt{d}$-defeat q. As a consequence, $\{p,q\}$ is $\mathtt{d}$-conflict-free, but at the same time it entails $\overline{q}$. In contrast, note that in $\mathrm{ABA}^{\mathtt{r}}$, consistency will be preserved since $\{q\}$ $\mathtt{r}$-defeats $\{p\}$, rendering $\{p,q\}$ not $\mathtt{r}$-conflict-free.

Accordingly, one might ask under which conditions consistency is preserved in the context of $\mathrm{ABA}^{\mathtt{d}}$. As done in ASPIC^{+} [33, 32], ASPIC^{-} [13] or $\mathrm{ASPIC}^{\ominus}$ [25] we focus our attention on contraposition-like properties.

Definition 9 (Contraposition). $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *is* closed under contraposition *if for every $\Delta \subseteq \mathcal{L}$ and every $A \in Ab$:*

> *If $\Delta \vdash_{\mathcal{R}} \overline{A}$ then for every $B \in \Delta \cap Ab$ there is a $\Theta \subseteq \{A\} \cup (\Delta \setminus \{B\})$ for which $\Theta \vdash_{\mathcal{R}} \overline{B}$.*

Note that the two formulas A and B that change position in our definition of contraposition are both assumptive. This distinguishes our definition from the stronger form of contraposition e.g. known from classical logic.

Although our definition of contraposition is similar to the definition found in [33, 32], it is less demanding. The difference is that [33, 32] define it by using a monotonic deducibility relation. In our notation:

Contraposition*: If $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta)$ then for every $B \in \Delta \cap Ab$, $\overline{B} \in \mathrm{Cn}_{\mathcal{R}}(\{A\} \cup (\Delta \setminus \{B\}))$.

Any framework that satisfies Contraposition* also satisfies Contraposition.

Fact 3. *Any $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ that satisfies Contraposition* also satisfies Contraposition.*

Proof. Suppose $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ satisfies Contraposition*. Assume further that $\Delta \vdash_{\mathcal{R}} \overline{A}$ and let $B \in \Delta \cap Ab$. Thus, $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta)$. By Contraposition*, $\overline{B} \in \mathrm{Cn}_{\mathcal{R}}(\{A\} \cup (\Delta \setminus \{B\}))$. Thus, there is a $\Theta \subseteq \{A\} \cup (\Delta \setminus \{B\})$ for which $\Theta \vdash_{\mathcal{R}} \overline{B}$. □

Not every framework that satisfies Contraposition also satisfies Contraposition*, as the following example shows.

Example 6. Consider $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ with $Ab = \{p, q, r\}$ and $\mathcal{R} = \{q \to \overline{r},\ r \to \overline{q}\}$. Then $\overline{r} \in \mathrm{Cn}_{\mathcal{R}}(\{p,q\})$, but $\overline{p} \notin \mathrm{Cn}_{\mathcal{R}}(\{q,r\})$.

Thus, all results below that hold for contrapositive ABFs will also hold for ABFs that satisfy the more demanding notion (here dubbed Contraposition*) from [33, 32].

A related condition on knowledge bases is *transposition* (see e.g. [33, 32]), which is a condition on the rule base $\mathcal{R}$, not on the resulting consequence relation $\vdash_{\mathcal{R}}$. In more detail, we can define a variant of transposition in our setting as follows:

Transposition: Where $A \in Ab$, if $A_1, \ldots, A_n \to \overline{A} \in \mathcal{R}$ then for every $1 \leqslant i \leqslant n$ for which $A_i \in Ab$: $A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n, A \to \overline{A_i} \in \mathcal{R}$.

Our variant restricts transposition to formulas A_i that are assumptions. In the context of ABA, transposition as originally defined is problematic, since in ABA contraries are only defined for Ab but not necessarily for every element of $\mathcal{L}$.

Moreover, neither does transposition imply contraposition nor vice versa. We give two examples:

Example 7. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ with $Ab = \{q, p\}$, $\mathcal{L} = Ab \cup \overline{Ab} \cup \{t\}$ and $\mathcal{R} = \{p \to t; t \to \overline{q}\}$. Notice that $\mathcal{R}$ is closed under transposition. However, $\vdash_{\mathcal{R}}$ is not closed under contraposition, since $\{p\} \vdash_{\mathcal{R}} \overline{q}$ yet $\{q\} \nvdash_{\mathcal{R}} \overline{p}$. The problem is that the derivation of $\overline{q}$ from p makes use of the intermediate step $\{p\} \vdash_{\mathcal{R}} t$, and since $t \notin Ab$, transposition does not enforce $\{q\} \nvdash_{\mathcal{R}} \overline{p}$.

Example 8. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ with $Ab = \{q, p\}$, $\mathcal{L} = Ab \cup \overline{Ab} \cup \{q'\}$ and $\mathcal{R} = \{p \to \overline{q}; q \to q'; q' \to \overline{p}\}$. Then $\vdash_{\mathcal{R}}$ is closed under contraposition, but $\mathcal{R}$ is not closed under transposition since $p \to \overline{q} \in \mathcal{R}$ but $q \to \overline{p} \notin \mathcal{R}$.

We close the discussion of transposition with the observation that transposition and contraposition are closely related as soon as $\mathcal{R}$ is closed under Cut. We say that $\mathcal{R}$ is *closed under Cut* iff whenever $A_1, \ldots, A_n \to B \in \mathcal{R}$ and $C_1, \ldots, C_m \to A_i \in \mathcal{R}$ then also $C_1, \ldots, C_m, A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n \to B \in \mathcal{R}$.

Fact 4. *If $\mathcal{R}$ is closed under Cut, then $A_1, \ldots, A_n \to B \in \mathcal{R}$ iff $A_1, \ldots, A_n \vdash_{\mathcal{R}} B$.*

Proof. The ($\Rightarrow$) direction is trivial. The ($\Leftarrow$) direction is shown inductively over the length of a derivation of B from $A_1, \ldots, A_n$. The base case is trivial. For the inductive step suppose $C_1, \ldots, C_m \to B$ is the last rule applied in the derivation of B from $A_1, \ldots, A_n$. Thus, for each C_i there are $A^i_1, \ldots, A^i_{k_i} \in \{A_1, \ldots, A_n\}$ such that $A^i_1, \ldots, A^i_{k_i} \vdash_{\mathcal{R}} A_i$. Also $\bigcup_{i=1}^m \{A^i_1, \ldots, A^i_{k_i}\} = \{A_1, \ldots, A_n\}$. By the inductive hypothesis, $A^i_1, \ldots, A^i_{k_i} \to A_i \in \mathcal{R}$. By applying Cut m times, $A_1, \ldots, A_n \to B \in \mathcal{R}$. □

Fact 5. *If $\mathcal{R}$ is closed under Cut and under transposition, then $\vdash_{\mathcal{R}}$ is closed under contraposition.*

Proof. Suppose $\mathcal{R}$ is closed under cut and transposition. Suppose $A_1, \ldots, A_n \vdash_{\mathcal{R}} \overline{A}$ and $A_i \in Ab$. By Fact 4, $A_1, \ldots, A_n \to \overline{A} \in \mathcal{R}$. Thus, by transposition, $A_1, \ldots, A_{i-1}, A, A_{i+1}, \ldots, A_n \to \overline{A_i}$. Hence, again by Fact 4,
$A_1, \ldots, A_{i-1}, A, A_{i+1}, \ldots, A_n \vdash_{\mathcal{R}} \overline{A_i}$. □

Indeed, contraposition guarantees consistency, as shown next.[11]

Theorem 1 (Consistency). *Let* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *be closed under contraposition. For any* $\Delta \subseteq Ab$*, if* Δ *is* $\mathtt{d}$*-conflict-free, then there is no* $A \in \Delta$ *for which* $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta)$*.*[12]

Note that Theorem 1 implies the consistency of an extension in any of the $\mathtt{d}$-semantics: admissible, complete, preferred, grounded, stable.

Consistency for $\mathrm{ABA}^{\mathtt{r}}$ follows immediate in view of Fact 2.[13]

Theorem 2 (Consistency). *Where* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$*, for any* $\Delta \subseteq Ab$*, if* Δ *is* $\mathtt{r}$*-conflict-free, then there is no* $A \in \Delta$ *for which* $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta)$*.*

3.2 Fundamental Lemma

In [20] we find the *Fundamental Lemma of Abstract Argumentation* (DFL, short for Dung's Fundamental Lemma). Informally, this result says that if a set of assumptions Δ is admissible and it defends another assumption A, $\Delta \cup \{A\}$ is admissible as well. This result is "fundamental" since it guarantees that one can build up complete extensions from admissible extensions Δ in an incremental way by simply adding assumptions that are defended by Δ. Where $\Delta = \emptyset$ this will give rise to the unique grounded extension (see Appendix C.3.1 for technical details).

As such, DFL assures us that different argumentation semantics will have all the properties that were investigated in [20] (besides the existence of a unique grounded extension also the fact that every preferred extension is complete, see Lemma 2 below). Indeed, when the DFL is violated, some of these properties might be violated, as is the case of $\mathrm{ABA}^{\mathtt{f}}$ for non-flat $\mathtt{ABF}$s (see [15, Example 2.15]). Example 4 is an example of the failure of DFL for $\mathrm{ABA}^{\mathtt{r}}$ for flat $\mathtt{ABF}$s not closed under contraposition (this behaviour was first observed by [18, Example 12]). Note that there the admissible set $\{a, b\}$ defends the assumption f but $\{a, b, r\}$ is not admissible. The example shows that in cases in which the Fundamental Lemma is violated we may loose other properties such as preferred extensions being complete or the existence of complete/grounded extensions. Even when complete extensions exist, the grounded extension may not be unique since there may be more than one minimal complete extensions. Our next example illustrates such a case.

[11] In ASPIC$^+$ [33], contraposition together with various other conditions are sufficient for consistency. For $\mathrm{ABA}^{\mathtt{d}}$ it turns out that contraposition alone guarantees consistency.

[12] The proof of this theorem can be found in Appendix C.1.

[13] This was also observed in [16, Lemma 6].

Example 9. Let $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$ (see Fig. 1) where

$$\mathcal{R} = \{a_1 \to \overline{a_2};\ a_2 \to \overline{a_1};\ a_1 \to \overline{a_3};\ a_3 \to \overline{a_1};\ a_2 \to \overline{a_3};\ a_3 \to \overline{a_2};\ b_1, b_2 \to \overline{a_1}\}$$
$$Ab = \{a_1, a_2, a_3, b_1, b_2\} \qquad \mathcal{L} = Ab \cup \overline{Ab} \qquad \mathbb{V} = \{1, 2\}$$
$$\upsilon(b_1) = \upsilon(b_2) = 1 \qquad \upsilon(a_1) = \upsilon(a_2) = \upsilon(a_3) = 2$$

In this case both $\{a_2, b_1, b_2\}$ and $\{a_3, b_1, b_2\}$ are $\subseteq$-minimal $\texttt{r}$-complete extensions and therefore $\texttt{r}$-grounded. Note for this that $\emptyset$ defends b_1 and b_2, $\{b_1\}$ defends b_2 and $\{b_2\}$ defends b_1 but $\{b_1, b_2\}$ is not admissible. So neither $\emptyset$ nor $\{b_1\}$ nor $\{b_2\}$ nor $\{b_1, b_2\}$ are complete. Similarly, neither $\{a_1\}$ nor $\{a_1, b_1\}$ nor $\{a_1, b_2\}$ nor $\{a_1, b_1, b_2\}$ are complete. Now, also both a_2 and a_3 defend the assumptions b_1 and b_2. Therefore neither $\{a_2\}$ nor $\{a_3\}$ are complete. Both $\{a_2, b_1, b_2\}$ and $\{a_3, b_1, b_2\}$ are complete.

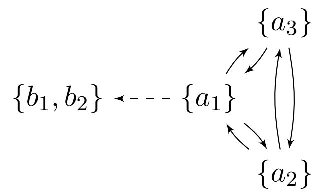


Figure 1: A fragment of the defeat diagram for Example 9. Here and in all the defeat diagrams below, full lines represent $\texttt{d}$-defeat whereas dashed lines represent (proper) $\texttt{r}$-defeats.

In view of these considerations, we prove the Fundamental Lemma for $\mathrm{ABA}^{\texttt{d}}$ for any $\texttt{ABF}$ and for $\mathrm{ABA}^{\texttt{r}}$ for $\texttt{ABF}$s *closed under contraposition.*

As explained above, for $\texttt{d}$-defeat, DFL holds for any $\texttt{ABF}$.

Theorem 3 (Fundamental Lemma, $\texttt{d}$-defeat)**.** *For any* $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$, *if the* $\texttt{d}$*-admissible* $\Delta \subseteq Ab$ $\texttt{d}$*-defends* $A \in Ab$ *then* $\Delta \cup \{A\}$ *is* $\texttt{d}$*-admissible.*[14]

The situation is more complicated when $\texttt{r}$-defeats are involved. As is shown by Example 4, for $\texttt{ABF}$s that are not closed under contraposition, DFL does not necessarily hold. However, when an $\texttt{ABF}$ is closed under contraposition it will satisfy DFL. In [18] the following formulation of DFL is proven.

Theorem 4 (Fundamental Lemma for Single Assumptions, $\texttt{r}$-defeat, [18])**.** *For any* $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$ *that is closed under contraposition, if the* $\texttt{r}$*-admissible* $\Delta \subseteq Ab$ $\texttt{r}$*-defends* $A \in Ab$ *then* $\Delta \cup \{A\}$ *is* $\texttt{r}$*-admissible.*

[14] The proof of this theorem can be found in Appendix C.2.

The Fundamental Lemma can easily be generalized to sets of defended assumptions.[15]

Lemma 1. *If* ABF $= (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *satisfies the Fundamental Lemma, then also for all* $\Delta, \Theta \subseteq Ab$*, if* Δ *is* r*-admissible and* r*-defends* Θ*, then* $\Delta \cup \Theta$ *is* r*-admissible as well.*

We immediately get the following corollary:[16]

Corollary 1 (Fundamental Lemma for Sets of Assumptions, r-defeat)**.** *For any* ABF $= (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *that is closed under contraposition, if the* r*-admissible* $\Delta \subseteq Ab$ r*-defends* $\Theta \subseteq Ab$ *then* $\Delta \cup \Theta$ *is* r*-admissible.*

We now state an important corollary, given the following insight:[17]

Lemma 2. *If* ABF $= (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *satisfies the Fundamental Lemma then (where* $\mathtt{x} \in \{\mathtt{r}, \mathtt{d}, \mathtt{f}\}$*):*

1. *there is a unique* x*-grounded extension and*
2. *every* x*-preferred extension is* x*-complete.*

We state an immediate consequence of Theorem 3, Corollary 1 and Lemma 2.

Corollary 2. *For any* ABF $= (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$,

1. *there is a unique* d*-grounded extension and*
2. *every* d*-preferred extension is* d*-complete.*

If ABF *is closed under contraposition,*

3. *there is a unique* r*-grounded extension and*
4. *every* r*-preferred extension is* r*-complete.*

[15] The proof can be found in Appendix C.3.

[16] In Appendix E we further generalize the result by showing that both versions of the Fundamental Lemma hold for r-defeat and for weakly contrapositive ABFs (see Lemmas 10 and 4 and Section 7 for more discussion on and a definition of weak contraposition).

[17] Item 1 of Lemma 2 is shown in Appendix C.3.1. Item 2 is a even more direct consequence: Suppose Δ is preferred. By the Fundamental Lemma $F(\Delta) \cup \Delta \supseteq \Delta$ is admissible where $F(\Delta)$ denotes the set of all assumptions defended by Δ. Since Δ is maximally admissible, $\Delta = F(\Delta) \cup \Delta$.

4 On the relation between $\mathrm{ABA^d}$ and $\mathrm{ABA^r}$

In this section we systematically compare $\mathrm{ABA^d}$ and $\mathrm{ABA^r}$. We will show that for a specific class of frameworks, dubbed *well-behaved* below, the two systems have the same preferred and stable extensions. We will proceed step-wise: first we do not consider any restrictions on the class of frameworks considered and subsequently we impose several restrictions. We first define the restrictions we consider.

In addition to contraposition, we introduce a new property called Sanity. The basic idea is that if an assumption seems 'paradoxical' and attacks itself, e.g. if there is a rule $A \to \overline{A}$ or more general if $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\{A\})$, then the rule-system should be equipped to allow filtering out these assumptions on independent grounds, i.e., by enforcing $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\emptyset)$. This is similar to classical logic where we have $\vdash_{\mathbf{CL}} \neg(A \wedge \neg A)$. We can generalize this basic requirement for our rule set $\mathcal{R}$ by demanding that whenever $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta)$, then $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta \setminus \{A\})$. The generalization disallows for assumptions to be in a more strict sense paradoxical, namely to be "essentially" involved in the demonstration of their own contrary:

Definition 10. *Where* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *and* $A \in Ab$*,* A *is a* paradoxical assumption *iff there is a* $\Delta \subseteq \mathcal{L} \setminus \{A\}$ *such that* $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta \cup \{A\})$ *and* $\overline{A} \notin \mathrm{Cn}_{\mathcal{R}}(\Delta)$*.*

Definition 11 (Sanity)**.** $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *is* sane *iff for all* $\Delta \subseteq \mathcal{L}$ *and* $A \in Ab$*,* $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta)$ *implies* $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta \setminus \{A\})$*.*

In the following remark we establish the above discussed relation between sanity and the absence of paradoxical assumptions and we additionally express sanity in terms of $\vdash_{\mathcal{R}}$.

Fact 6. *Where* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$*, the following statements are equivalent:*

1. $\mathtt{ABF}$ *is sane,*
2. *there are no paradoxical assumptions in* Ab*,*
3. *for all* $\Delta \subseteq \mathcal{L}$*, if* $\Delta \vdash_{\mathcal{R}} \overline{A}$ *for some* $A \in \Delta$ *then there is a* $\Delta' \subseteq \Delta \setminus \{A\}$ *for which* $\Delta' \vdash_{\mathcal{R}} \overline{A}$*.*

Proof. The equivalence of 1 and 2 follows immediately in view of the Definitions 10 and 11. Here is the proof of the "1 $\Rightarrow$ 3" direction. Suppose $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ is sane and suppose $\Delta \vdash_{\mathcal{R}} \overline{A}$ for some $A \in \Delta$. Thus, $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta)$. Thus, by sanity, $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta \setminus \{A\})$. Hence, there is a $\Delta' \subseteq \Delta \setminus \{A\}$ for which $\Delta' \vdash_{\mathcal{R}} \overline{A}$. The other direction is similar. $\square$

Remark 5. *Note that sanity does* not *require that there are no self-attacking sets of assumptions $\Delta \subseteq Ab$. All that is required is that there are no derivations of the form $\Delta \vdash_{\mathcal{R}} \overline{A}$, where $A \in \Delta$, that are not reducible to $\Delta' \vdash_{\mathcal{R}} \overline{A}$ for some $\Delta' \subseteq \Delta \setminus \{A\}$.*

Example 10. To further clarify the definition of sane ABFs, let us consider $\mathtt{ABF}_1 = (\mathcal{L}, \mathcal{R}_1, Ab, \overline{}, \mathbb{V}, \mathbb{V}^2, v)$ and $\mathtt{ABF}_2 = (\mathcal{L}, \mathcal{R}_2, Ab, \overline{}, \mathbb{V}, \mathbb{V}^2, v)$ where:

$$Ab = \{p, q, s\} \quad \mathcal{L} = Ab \cup \overline{Ab} \quad \mathcal{R}_1 = \{\ p, q \to \overline{s}\ \} \quad \mathcal{R}_2 = \{\ p, q, s \to \overline{s}\ \}$$

$\mathtt{ABF}_2$ is not sane since $\overline{s} \in \mathrm{Cn}_{\mathcal{R}_2}(\{p, q, s\}) \setminus \mathrm{Cn}_{\mathcal{R}_2}(\{p, q\})$ and so s is paradoxical. Note that this is not the case for $\mathtt{ABF}_1$ since even though $\{p, q, s\}$ is self-attacking, $\{p, q, s\} \nvdash_{\mathcal{R}_1} \overline{s}$ and since $\{p, q\} \vdash_{\mathcal{R}_1} \overline{s}$ there is an independent attacker of s which renders s non-paradoxical.

Remark 6. *Sanity does also not require that there are no odd defeat cycles and as such it does not imply that every preferred extension is automatically stable or that stable sets exist (Example 11). However, as will be demonstrated (Theorem 6), in the presence of contraposition, sanity is sufficient to guarantee the latter two properties.*

Example 11. Consider $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$ and $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{})$ where $Ab = \{s, q, p, r\}$, $\mathcal{L} = Ab \cup \overline{Ab}$ and $\mathcal{R} = \{p \to \overline{q};\ \ q \to \overline{s};\ \ s \to \overline{p}\}$. Then $\{r\}$ is the only $\mathtt{x}$-preferred extension of $\mathtt{ABF}$ and there is no $\mathtt{x}$-stable set. Note that $\mathtt{ABF}$ is sane.

For the sake of completeness we also give a simple example demonstrating the contraposition is not sufficient to warrant the existence of stable extensions.

Example 12. Consider $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$ and $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{})$ where $Ab = \{s\}$, $\mathcal{L} = Ab \cup \overline{Ab}$ and $\mathcal{R} = \{s \to \overline{s}\}$. Then $\emptyset$ is the only $\mathtt{x}$-preferred extension of $\mathtt{ABF}$ and there is no $\mathtt{x}$-stable set. Note that $\mathtt{ABF}$ is closed under contraposition.

Since $\mathtt{r}$-defeat is essentially a form of contrapositive reasoning, one could ask whether a given $\mathtt{ABF}$ closed under contraposition gives the same outcomes under $\mathrm{ABA}^{\mathtt{d}}$ and under $\mathrm{ABA}^{\mathtt{r}}$. Thus, another central property of interest is the closure of the underlying deducability relation $\vdash_{\mathcal{R}}$ of a given $\mathtt{ABF}$ under contraposition (see Definition 9).

Putting these requirements together we end up with a notion of *well-behaved* frameworks.

Definition 12 (Well-Behaved)**.** *We call an ABF* well-behaved *if it is sane and closed under contraposition.*[18]

[18]The name *well-behaved* was chosen since this class of ABFs behaves particularly well with respect to the meta-theoretic properties studied in this paper. To avoid being misunderstood, this does not mean that ABFs outside of this class cannot also have very useful for specific applications.

Our results are summarized in Table 2. The top line lists the assumptions on the ABF framework under which a claim (in the leftmost column) is proved or shown to fail. The table refers to the results containing these proofs, or to counter-examples to the corresponding claims (in gray). An empty cell to the left of a counter-example means that the example holds for the cell under consideration as well, whereas an empty cell to the right of a positive result means that the result applies to the cell under consideration as well.

When imposing no restrictions or merely sanity on the class of frameworks under consideration, $\text{ABA}^{\mathtt{d}}$ and $\text{ABA}^{\mathtt{r}}$ give rise to different extensions with all the standard semantics (see the 1st column in Table 2). The comparability improves when considering contraposition (see the 2nd column of the table). For instance, stable extensions then coincide (see also Theorem 5). Finally when considering well-behaved frameworks the preferred and stable extensions of the two approaches are identical (see the 3rd column of Table 2 and also Theorem 6).

The remainder of this section consists of two parts: first we state our positive results in Fact 7, Theorem 5 and Theorem 6. Then we state negative results in terms of (counter-)examples.

The following fact follows immediately due to the fact that every $\mathtt{d}$-defeat is also a $\mathtt{r}$-defeat.

Fact 7. *For any* $\mathtt{ABF}$ *we have that every* $\mathtt{r}$*-conflict-free set is* $\mathtt{d}$*-conflict-free.*

Theorem 5. *If* $\mathtt{ABF}$ *is closed under contraposition,*

(1) every $\mathtt{d}$*-conflict-free set is* $\mathtt{r}$*-conflict-free and vice versa;*

(2) every $\mathtt{d}$*-stable set is* $\mathtt{r}$*-stable and vice versa;*

(3) every $\mathtt{d}$*-admissible set is an* $\mathtt{r}$*-admissible set;*

(4) every $\mathtt{d}$*-complete set is a subset of an* $\mathtt{r}$*-complete set.*[19]

Theorem 6. *For any well-behaved* $\mathtt{ABF}$ *we have:*

$$\mathtt{r}\text{-}\mathsf{pref}(\mathtt{ABF}) = \mathtt{d}\text{-}\mathsf{pref}(\mathtt{ABF}) = \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF}) = \mathtt{d}\text{-}\mathsf{stab}(\mathtt{ABF}).^{20}$$

Example 13. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$ where:

$$\begin{array}{lll} Ab = \{p, r\} & \mathcal{L} = Ab \cup \overline{Ab} & \mathcal{R} = \{p \rightarrow \overline{r}\} \\ \mathbb{V} = \{1, 2\} & \upsilon(p) = 1,\ \upsilon(r) = 2 & \end{array}$$

[19] This theorem is proven in Appendix F.

[20] This theorem is an immediate consequence of Theorem 13, proven in Appendix D.

	no restr.	sane	contrap.	well-beh.
d-pref(ABF) ⊆ d-comp(ABF)	Cor. 2			
r-pref(ABF) ⊆ r-comp(ABF)		Ex. 4	Cor. 2	
r/d-pref(ABF) ⊆ r/d-stab(ABF)		Ex. 11	Ex. 12	Thm. 6
Uniqueness/Existence of d-grou(ABF)	Cor. 2			
Uniqueness/Existence of r-grou(ABF)		Ex. 9/ Ex. 4	Cor. 2	
d-cf(ABF) ⊆ r-cf(ABF)		Ex. 13	Thm. 5	
r-cf(ABF) ⊆ d-cf(ABF)	Fact 7			
d-adm(ABF) ⊆ r-adm(ABF)		Ex. 13	Thm. 5	
r-adm(ABF) ⊆ d-adm(ABF)				Ex. 14
d-pref(ABF) ⊆ r-pref(ABF)		Ex. 13	Ex. 15	Thm. 6
r-pref(ABF) ⊆ d-pref(ABF)		Ex. 13	Ex. 15	Thm. 6
every d-comp(ABF) is subset of some r-comp(ABF)		Ex. 13	Thm. 5	
r-comp(ABF) ⊆ d-comp(ABF)				Ex. 14
d-stab(ABF) ⊆ r-stab(ABF)		Ex. 13	Thm. 5	
r-stab(ABF) ⊆ d-stab(ABF)		Ex. 13	Thm. 5	
every d-grou(ABF) is subset of some r-grou(ABF)		Ex. 13	Thm. 5	
every r-grou(ABF) is subset of some d-grou(ABF)				Ex. 14

Table 2: A comparison of $\mathrm{ABA}^{\mathtt{d}}$ and $\mathrm{ABA}^{\mathtt{r}}$.

Notice that ABF is sane. $\{p\}$ does not d-defeat r since $v(p) < v(r)$. Consequently, $\{p, r\}$ is d-stable, d-preferred, and d-grounded. However, $\{p, r\}$ is not r-admissible. To see this, notice that r r-defeats $\{p\}$. Moreover, $\{r\}$ is r-stable, r-preferred, and r-grounded, while not being d-complete.

For a set of assumptions $\{A_1, \ldots, A_n\} \subseteq Ab$, we will use $\overline{A_1, \ldots, A_n}$ to denote the list of all the contrapositions of the rule $A_1, \ldots, A_{n-1} \rightarrow \overline{A_n}$. That is:

$$\overline{A_1, \ldots, A_n} \quad = \quad A_1, \ldots, A_{n-1} \rightarrow \overline{A_n}\,; \quad \ldots\,; \quad A_2, \ldots, A_n \rightarrow \overline{A_1}$$

Example 14. Let $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, \overline{}, \mathbb{V}, \leq, v)$ where (see Figure 2):

$$Ab = \{p, r, s, q\} \qquad \mathcal{L} = Ab \cup \overline{Ab} \qquad \mathcal{R} = \{\ \overline{p, q, r};\ \ \overline{p, q, s}\ \}$$
$$\mathbb{V} = \{1, 2, 3\} \quad v(s) = 1,\ v(p) = v(q) = 2,\ v(r) = 3$$

Note that ABF is well-behaved. Note that $\{r, s\}$ is r-complete but not d-admissible. To see that $\{r, s\}$ is r-complete note that r r-defeats $\{p, q\}$, the only d-defeater of s. Note that $\{r, s\}$ is r-grounded whereas $\{r\}$ is d-complete and d-grounded.

Example 15. Let $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, \overline{}, \mathbb{V}, \leq, v)$ where (see Figure 2):

$$Ab = \{s, p, q, r\} \quad \mathcal{L} = Ab \cup \overline{Ab} \quad \mathcal{R} = \{\ \overline{p, q, r};\ \overline{p, q, s};\ p \rightarrow \overline{p};\ q \rightarrow \overline{q}\ \}$$
$$\mathbb{V} = \{1, 2, 3\} \quad v(s) = 1 \quad v(p) = v(q) = 2 \quad v(r) = 3$$

Observe that ABF is closed under contraposition yet not sane (e.g., p and q are paradoxical). Note that $\{r, s\}$ is r-preferred. However $\{r\}$ is d-preferred, showing that a maximally d-admissible set (like $\{r\}$) need not be maximally r-admissible.

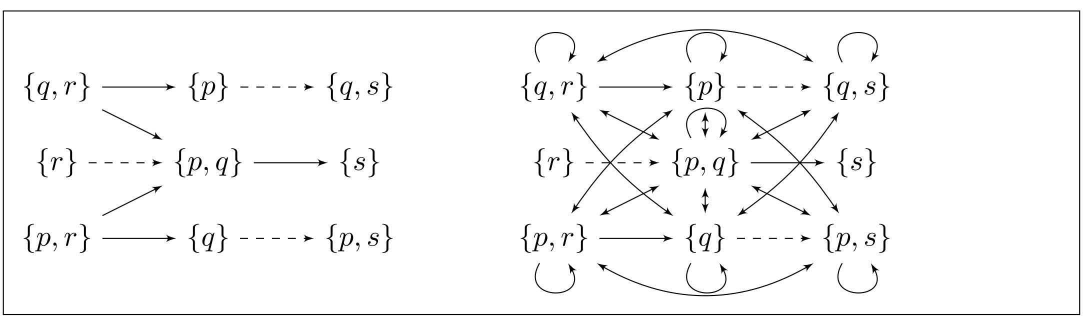


Figure 2: Fragments of the defeat diagrams for Example 14 (left) resp. Example 15 (right).

5 Properties for Nonmotonic Reasoning

Structured argumentation has been proposed as a powerful framework to model defeasible reasoning. Clearly, also assumption-based argumentation gives rise to non-monotonic consequence relations.

Example 16. Let $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where:

$$\begin{array}{llll} Ab = \{p\} & \mathcal{L} = \{p, q\} & \overline{p} = q & \mathcal{R} = \emptyset \\ \mathbb{V} = \{1\} & v(p) = 1 & & \end{array}$$

Let, moreover, $\texttt{ABF}^q = (\mathcal{L}, \mathcal{R} \cup \{\to q\}, Ab, \overline{}, \mathbb{V}, \leq, v)$ be the result of enriching ABF with the information that q holds (modeled in terms of a strict rule).

Clearly, for any $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}, \mathsf{grou}\}$ and $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$ we have $\texttt{ABF} \mathrel{|\!\sim}^{\mathsf{sem}}_{\mathtt{x}} p$ while $\texttt{ABF}^q \mathrel{|\!\not\sim}^{\mathsf{sem}}_{\mathtt{x}} p$.

This motivates studying properties for non-monotonic reasoning for assumption-based argumentation in this section. In particular we will investigate properties such as Cautious Monotony, Cautious Cut, Cumulativity, Rational Monotony, Rationality and two properties that concern Monotony and Cautious Cut under the addition of assumptions, namely *Ab-Monotony* and *Cautious Ab-Montony*.

Probably the most well-known properties studied here are *Cautious Monotony* and *Cautious Cut* [21, 29, 26]. Cautious Cut requires that adding information to an ABF that is derivable from the knowledge base does not result in new consequences. In other words, if A follows from ABF and adding $\to A$ to the strict rule base of ABF (resulting in $\texttt{ABF}^A$) allows to derive B, then B should have already been derivable from ABF. The reverse of Cautious Cut is known as *Cautious Monotony* and requires that if A and B are derivable from an ABF then adding $\to A$ to the strict rule base of ABF should not influence the derivability of B. In other words, no information is lost when adding information that is already derivable from the ABF under consideration. Consequence relations that satisfy Cautious Cut and Cautious Monotony are called *cumulative*. A weaker version of Cautious Monotony is *Cautious Ab-Monotony*, which additionally requires that A is an assumption. Finally, we study two strengthenings of Cautious Monotony: Rational Monotony and Ab-Monotony. *Rational Monotony* [26] is a somewhat controversial rule (see e.g. [37]) which requires that if a contrary of an assumption B is not derivable, adding the assumption B as a strict premise (i.e. adding the rule $\to B$ to $\mathcal{R}$) does not result in any loss of derivable information. Finally, the property *Ab-monotony* has, to the best of our knowledge, not been defined before.[21] It expresses that an ABF is robust under

[21]It is similar to *semi-monotonicity* known from default logic [28], which concerns monotonicity under the addition of default rules.

adding assumptions as strict rules: i.e. removing an assumption A from Ab and adding it as a rule $\to A$ instead preserves the consequences from $\texttt{ABF}$.

Definition 13. *Where* $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *and given a set of formulas* $\Delta \cup \{A\} \subseteq \mathcal{L}$, *let* $\Delta^{-A} = \Delta$ *if* $A \notin Ab$ *and* $\Delta^{-A} = \Delta \setminus \{A\}$ *if* $A \in Ab$. *Let* $\texttt{ABF}^A = (\mathcal{L}, \mathcal{R} \cup \{\to A\}, Ab^{-A}, \overline{}, \mathbb{V}, \leq, v')$ *with* $v'(C) = v(C)$ *for every* $C \in Ab \setminus \{A\}$.

Definition 14. *Let* $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$, $\mathsf{sem} \in \{\mathsf{grou}, \mathsf{pref}, \mathsf{stab}\}$, $\texttt{x} \in \{\texttt{r}, \texttt{d}\}$, *and* $A, B \in \mathcal{L}$. *In Table 3 we define several of the well-known postulates frequently discussed in non-monotonic logic.*

Before presenting our results, we show some relations between the properties defined above. First, we comment shortly on the relationship between cumulative and rational ABFs. Thereafter we explain the relationship of Ab-monotony with Monotony and Rational Monotony.

For most formal systems of defeasible reasoning, Rational Monotony can be seen as a special case of Cautious Monotony. In the case of assumption-based argumentation, however, violations of consistency may give rise to ABFs which violate Cautious Monotony while not violating Rational Monotony as demonstrated in Example 17 (for $\texttt{d}$-preferred and $\texttt{d}$-stable semantics).

Example 17. Let $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where:

$$Ab = \{s, r, p, t', t\} \quad \mathcal{L} = Ab \cup \overline{Ab} \quad \mathcal{R} = \{\, s, r \to \overline{p};\;\; t', t \to \overline{s}\,\}$$
$$\mathbb{V} = \{1, 2, 3, 4\} \quad v(t') = v(t) = 1 \quad v(s) = 2 \quad v(p) = 3 \quad v(r) = 4$$

We have $\texttt{d}\text{-}\mathsf{pref}(\texttt{ABF}) = \texttt{d}\text{-}\mathsf{stab}(\texttt{ABF}) = \texttt{d}\text{-}\mathsf{grou}(\texttt{ABF}) = \{Ab\}$. Notice that it holds that $\texttt{ABF} \mid\!\sim^{\texttt{d}}_{\mathsf{sem}} \overline{s}, \overline{p}$ and $\texttt{ABF} \not\mid\!\sim^{\texttt{d}}_{\mathsf{sem}} \overline{r}, \overline{t'}, \overline{t}$. Also, $\texttt{ABF}^A \mid\!\sim^{\texttt{d}}_{\mathsf{sem}} B$ if $\texttt{ABF} \not\mid\!\sim^{\texttt{d}}_{\mathsf{sem}} \overline{A}$ for all $A \in Ab$ and all $B \in \mathcal{L}$. So RM holds for $\texttt{ABF}$.

However, neither CM nor CM-Ab holds. Take, for instance, $\texttt{ABF}^s \not\mid\!\sim^{\texttt{d}}_{\mathsf{sem}} p$ while $\texttt{ABF} \mid\!\sim^{\texttt{d}}_{\mathsf{sem}} p$ and $\texttt{ABF} \mid\!\sim^{\texttt{d}}_{\mathsf{sem}} s$.

However, when an $\texttt{ABF}$ satisfies $\texttt{f}$-consistency for a given semantics, the expected relationship between Cautious and Rational Monotony is preserved, as shown in the following fact.

Definition 15 ($\texttt{f}$-consistency)**.** *Where* $\texttt{x} \in \{\texttt{r}, \texttt{d}\}$ *and* $\mathsf{sem} \in \{\mathsf{grou}, \mathsf{stab}, \mathsf{pref}\}$, $\texttt{ABF}$ *satisfies* $\texttt{f}$*-consistency under* $\texttt{x}\text{-}\mathsf{sem}$ *iff for every* $\Delta \in \texttt{x}\text{-}\mathsf{sem}(\texttt{ABF})$, Δ *is* $\texttt{f}$*-conflict-free.*

Fact 8. *Where* $\texttt{x} \in \{\texttt{r}, \texttt{d}\}$ *and* $\mathsf{sem} \in \{\mathsf{grou}, \mathsf{stab}, \mathsf{pref}\}$, *if* $\texttt{ABF}$ *is* $\texttt{f}$*-consistent under* $\texttt{x}\text{-}\mathsf{sem}$ *and it satisfies RM (relative to* $\mid\!\sim^{\mathsf{sem}}_{\texttt{x}}$*) then it satisfies CM-Ab (relative to* $\mid\!\sim^{\mathsf{sem}}_{\texttt{x}}$*).*[22]

[22] The proof of this fact can be found in Appendix G.

Cautious Cut (CC)				
If $\mathtt{ABF} \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} A$ and	$\mathtt{ABF}^A \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$	then	$\mathtt{ABF} \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$.	*CC-ent*
	$\Delta \in \mathsf{x\text{-}sem}(\mathtt{ABF})$	then	$\Delta^{-A} \in \mathsf{x\text{-}sem}(\mathtt{ABF}^A)$.	*CC-sem*
Monotony (M)				
If	$\mathtt{ABF} \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$	then	$\mathtt{ABF}^A \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$.	*M-ent*
If	$\Delta^{-A} \in \mathsf{x\text{-}sem}(\mathtt{ABF}^A)$	then	$\Delta \in \mathsf{x\text{-}sem}(\mathtt{ABF})$.	*M-sem*
Ab-Monotony (M-Ab) [where $A \in Ab$ and $\overline{A} \notin \mathrm{Cn}_{\mathcal{R}}(\emptyset)$]				
If	$\mathtt{ABF} \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$	then	$\mathtt{ABF}^A \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$.	*M-Ab-ent*
If	$\Delta^{-A} \in \mathsf{x\text{-}sem}(\mathtt{ABF}^A)$	then	$\Delta \in \mathsf{x\text{-}sem}(\mathtt{ABF})$.	*M-Ab-sem*
Cautious Monotony (CM)				
If $\mathtt{ABF} \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} A$ and	$\mathtt{ABF} \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$	then	$\mathtt{ABF}^A \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$.	*CM-ent*
	$\Delta^{-A} \in \mathsf{x\text{-}sem}(\mathtt{ABF}^A)$	then	$\Delta \in \mathsf{x\text{-}sem}(\mathtt{ABF})$.	*CM-sem*
Cautious Ab-Monotony (CM-Ab) [where $A \in Ab$]				
If $\mathtt{ABF} \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} A$ and	$\mathtt{ABF} \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$	then	$\mathtt{ABF}^A \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$.	*CM-Ab-ent*
	$\Delta^{-A} \in \mathsf{x\text{-}sem}(\mathtt{ABF}^A)$	then	$\Delta \in \mathsf{x\text{-}sem}(\mathtt{ABF})$.	*CM-Ab-sem*
Cumulativity iff CC and CM				
Rational Monotony (RM) [where $A \in Ab$]				
If $\mathtt{ABF} \not\mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} \overline{A}$ and	$\mathtt{ABF} \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$	then	$\mathtt{ABF}^A \mathrel{\mid\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$.	*RM-ent*
	$\Delta^{-A} \in \mathsf{x\text{-}sem}(\mathtt{ABF}^A)$	then	$\Delta \in \mathsf{x\text{-}sem}(\mathtt{ABF})$.	*RM-sem*

Table 3: Properties of $\mathtt{ABF}$ (relative to $\mathsf{x\text{-}sem}$), Definition 14. Each property consists of two necessary and jointly sufficient requirements, one concerning the entailment one concerning the semantic selection. Also, A and B are universally quantified over $\mathcal{L}$ (where A satisfies some further restrictions in (M-Ab), (CM-Ab) and (RM)).

From Fact 8 it follows that any $\mathtt{ABF}$ that satisfies RM satisfies CM-Ab for any $\mathtt{r}$-semantics. This can be seen by observing that for any $\mathtt{r}$-semantics, $\mathtt{f}$-consistency is always satisfied.

Fact 9. *Any* $\mathtt{ABF}$ *is* $\mathtt{f}$*-consistent under* $\mathtt{r}\text{-}\mathsf{sem}$ *for any* $\mathsf{sem} \in \{\mathsf{grou}, \mathsf{pref}, \mathsf{stab}\}$.

To see this notice that Δ $\mathtt{f}$-defeats A iff Δ $\mathtt{d}$-defeats A or A $\mathtt{r}$-defeats Δ (cf. Fact 2).

The following fact thus follows immediately from Fact 8 and Fact 9.

Fact 10. *Where* $\mathsf{sem} \in \{\mathsf{grou}, \mathsf{stab}, \mathsf{pref}\}$*, if* $\mathtt{ABF}$ *satisfies RM (relative to* $\mid\!\sim^{\mathsf{sem}}_{\mathtt{r}}$*) then it satisfies CM-Ab (relative to* $\mid\!\sim^{\mathsf{sem}}_{\mathtt{r}}$*).*

For $\mathtt{d}$-defeats, by Theorem 1 and Fact 8, closure of an $\mathtt{ABF}$ under contraposition is a sufficient condition for the expected relation between RM and CM-Ab to hold:

Fact 11. *Where* $\mathsf{sem} \in \{\mathsf{grou}, \mathsf{stab}, \mathsf{pref}\}$*, if* $\mathtt{ABF}$ *is closed under contraposition and satisfies RM (relative to* $\mid\!\sim^{\mathsf{sem}}_{\mathtt{d}}$*) then it satisfies CM-Ab (relative to* $\mid\!\sim^{\mathsf{sem}}_{\mathtt{d}}$*).*

While RM as defined above expresses the robustness of the $\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}}$-entailment under the addition of assumptions A whose contraries $\overline{A}$ are not derivable from a given $\mathtt{ABF}$, one could go a step further and also demand robustness of the $\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}}$-entailment under the addition of contraries of assumptions $\overline{A}$ where A is not derivable from a given $\mathtt{ABF}$. The following example shows that this stronger variant of RM does not hold, not even for well-behaved non-prioritized ABFs.

Example 18. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where $\mathcal{R} = \{\overline{q', p};\ \overline{q', q};\ p, q \to s;\ q' \to s\}$ and

$$Ab = \{p, q, q'\} \quad \mathcal{L} = Ab \cup \overline{Ab} \cup \{s\}$$
$$\mathbb{V} = \{1\} \quad v(p) = v(q) = v(q') = 1$$

For an illustration see Figure 3. Note that $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} s$ and $\mathtt{ABF} \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} q$ (where $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}\}$). To verify this notice that the $\mathtt{x}$-stable (resp. $\mathtt{x}$-preferred) sets are $\{p, q\}$ and $\{q'\}$. When moving to $\mathtt{ABF}^{\overline{q}}$ we have: $\mathtt{ABF}^{\overline{q}} \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} s$ since the $\mathtt{x}$-stable (resp. $\mathtt{x}$-preferred) sets are $\{p\}$ and $\{q'\}$ and furthermore $\{p\} \nvdash_{\mathcal{R} \cup \{\to \overline{q}\}} s$.

(a) Defeat Diagram for $\mathtt{ABF}$ (b) Defeat Diagram for $\mathtt{ABF}^{\overline{q}}$

Figure 3: A fragment of the defeat diagram for Example 18.

The property *Ab-monotony* is not to be confused with Monotony. The following example shows that Ab-monotony does *not* imply Monotony. In other words, there is an important difference between adding assumptions as a fact and adding *any* formula as a fact.

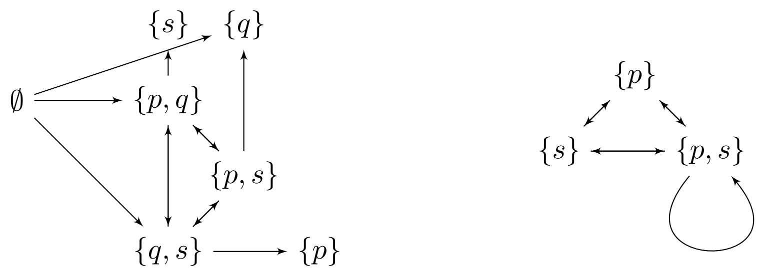
Example 19 (Ex. 16 cont.)**.** We again consider Ex. 16. We have already established that it violates Monotony. Note though that it does not violate Ab-Monotony since $\mathtt{ABF}$ and $\mathtt{ABF}^{p}$ have the same $\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}}$-consequences for any $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$ and $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}, \mathsf{grou}\}$.

In the definition of M-Ab, the restriction $\overline{A} \notin \mathrm{Cn}_{\mathcal{R}}(\emptyset)$ is necessary, otherwise counter-examples are readily found, such as the following example.

Example 20. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where (for $\alpha \in \{p, q, s\}$) (see Fig. 4):

$$\begin{array}{lll} Ab = \{p, q, s\} & \mathcal{L} = \{p, q, s, p', q', s'\} & \mathcal{R} = \{\rightarrow q';\ q, s \rightarrow p';\ p, s \rightarrow q';\ p, q \rightarrow s'\} \\ \mathbb{V} = \{1\} & v(p) = v(q) = v(s) = 1 & \overline{\alpha} = \alpha' \text{ for any } \alpha \in Ab \end{array}$$

We have $\mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}) = \{\{p, s\}\}$ while $\mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}^q) = \{\{p\}, \{s\}\}$ (for any $\mathsf{sem} \in \{\mathsf{stab}, \mathsf{pref}\}$ and any $\mathtt{x} \in \{\mathtt{r}, \mathtt{d}, \mathtt{f}\}$). Note for this that $\{q, s\}$ $\mathtt{x}$-defeats p in $\mathtt{ABF}$ and is $\mathtt{x}$-defeated by $\emptyset$ while $\{s\}$ $\mathtt{x}$-defeats p in $\mathtt{ABF}^q$. Similarly, $\{q, p\}$ $\mathtt{x}$-defeats s in $\mathtt{ABF}$ and is $\mathtt{x}$-defeated by $\emptyset$ while $\{p\}$ $\mathtt{x}$-defeats s in $\mathtt{ABF}^q$. In view of this $\{p, s\}$ is conflict-free in the context of $\mathtt{ABF}$ but not in the context of $\mathtt{ABF}^q$ and thus $\{p, s\}$ is neither in $\mathtt{x}\text{-}\mathsf{pref}(\mathtt{ABF}^q)$ nor in $\mathtt{x}\text{-}\mathsf{stab}(\mathtt{ABF}^q)$.



Defeat Diagram for $\mathtt{ABF}$ Defeat Diagram for $\mathtt{ABF}^q$

Figure 4: A fragment of the defeat diagram for Example 20.

To conclude this preliminary discussion of the properties that will be studied in this section, we note that it is trivial to see that M-Ab implies CM-Ab and RM for assumptions $B \in Ab$ for which $\overline{B} \notin \mathrm{Cn}_{\mathcal{R}}(\emptyset)$. For other assumptions this need not be so, as the following example illustrates.

Example 21. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where $\mathcal{R} = \{\ \rightarrow \overline{q}\} \cup \{q \rightarrow \overline{s_i};\ s_i \rightarrow \overline{s_i};\ s_i \rightarrow \overline{p} \mid i \in \{1, 2\}\}$ and

$$\begin{array}{rr} Ab = \{p, q, s_1, s_2\} & \mathcal{L} = Ab \cup \overline{Ab} \\ \mathbb{V} = \{1\} & v(p) = v(q) = v(s_1) = v(s_2) = 1 \end{array}$$

A partial defeat diagram for this $\mathtt{ABF}$ can be found in Figure 5. Note that there is no d-stable extension of $\mathtt{ABF}$ and therefore, for instance, $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathsf{d}} q$ and $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathsf{d}} s_1$. However, $\mathtt{ABF}^q$ has one stable extension, namely $\{p\}$ and therefore $\mathtt{ABF}^q \not\mid\!\sim^{\mathsf{stab}}_{\mathsf{d}} s_1$. This shows that CM-Ab does not hold for $\mathtt{ABF}$ and $\mathsf{d}\text{-}\mathsf{stab}$.

The situation is different for M-Ab. Here we note that for all $B \in Ab$ for which $\overline{B} \notin \mathrm{Cn}_{\mathcal{R}}(\emptyset)$, namely $B \in \{s_1, s_2, p\}$, also $\mathtt{ABF}^B$ has no stable extensions. Therefore, trivially, M-Ab holds for ABF and d-stab.

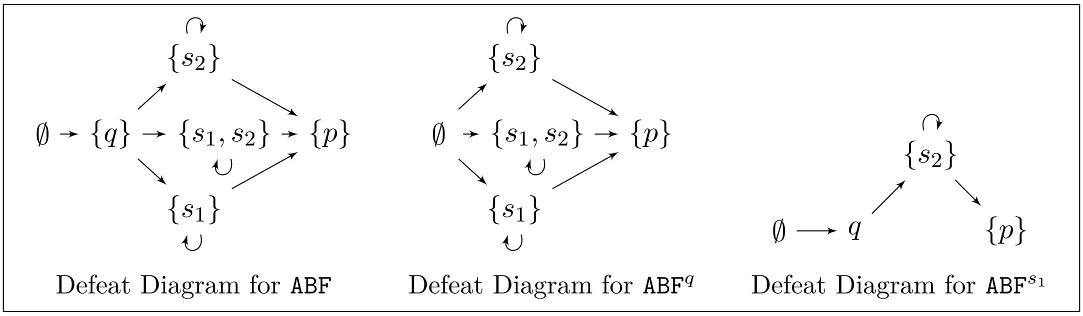


Figure 5: A fragment of the defeat diagram for Example 21.

Definition 16. $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *is* assumption-consistent *iff for all* $B \in Ab$, $\overline{B} \notin \mathrm{Cn}_{\mathcal{R}}(\emptyset)$.

Fact 12. *If* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *is assumption-consistent, then M-Ab implies both CM-Ab and RM.*

In the following, we again proceed in two steps: first we present positive results (Theorems 7–12, which are all proven in Appendix G) and then we provide negative results for specific subclasses of frameworks in view of counter-examples (Examples 22–32). Table 4 summarizes the results of this section. Cautious Cut and Monotony

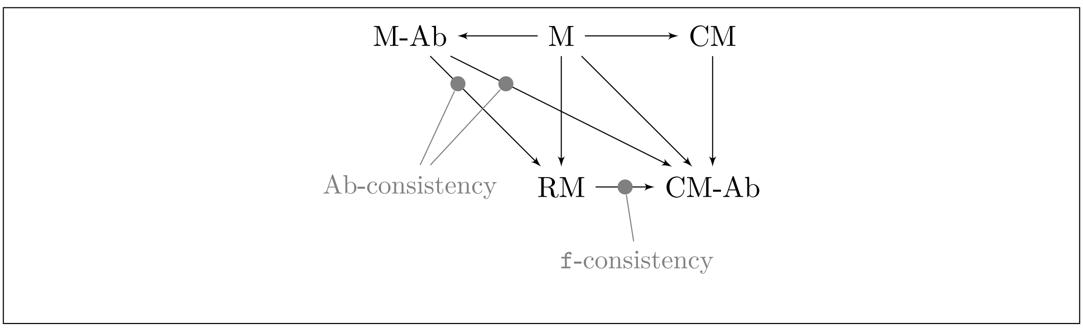


Figure 6: Schematic representation of relations between the properties studied in this section. $\mathsf{Prop}_1 \rightarrow \mathsf{Prop}_2$ means that any ABF that satisfies Prop_1 also satisfies Prop_2. An arrow with a circle on it means that the relation holds only when the ABF satisfies the specified condition (e.g. M-Ab implies RM when Ab-consistency is satisfied).

	no restr.	sane	contr.	well-behav.	wb., $\leq = \mathbb{V}^2$
M, x-sem					Ex. 16
CC, d-grou		Ex. 22	Thm. 8		
CC, d-pref/stab		Ex. 22	Thm. 9		
CC, r-grou				Ex. 26	Thm. 12
CC, r-pref		Ex. 23	Ex. 24	Thm. 10	
CC, r-stab		Ex. 23	Thm. 7		
CM, d-grou		Ex. 22	Thm. 8		
CM, d-stab/pref		Ex. 22	Ex. 30	Thm. 10	
CM, r-grou				Ex. 27	Thm. 12
CM, r-stab/pref		Ex. 23	Ex. 30	Thm. 10	
CM-Ab, d-grou		Ex. 31	Thm. 8		
CM-Ab, r-grou				Ex. 32	Thm. 12
CM-Ab, x-stab/pref		Ex. 25	Ex. 30	Thm. 10	
RM/M-Ab x-grou				Ex. 29	Thm. 12
RM/M-Ab, x-pref/stab				Ex. 28	Thm. 11

Table 4: Properties for Nonmonotonic Reasoning of ABA^{d} and ABA^{r}.

are investigated for ABFs in general, ABFs closed under contraposition and well-behaved ABFs. For Rational Monotony, we furthermore investigate the behaviour of ABFs for which every assumption has the same priority. Again, an empty cell to the left of a counter-example means that the example holds for the cell under consideration as well, whereas an empty cell to the right of a positive result means that the result applies to the cell under consideration as well.

We first note that the property of being closed under contraposition, sanity and well-behavedness of an ABA-framework are invariant under enhancements (for the proof see Appendix G.1):

Fact 13. *Where* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *is closed under contraposition [resp. sane, resp. well-behaved], then also* $\mathtt{ABF}^B = (\mathcal{L}, \mathcal{R} \cup \{\rightarrow B\}, Ab \setminus \{B\}, \overline{}, \mathbb{V}, \leq, v)$ *is closed under contraposition [resp. sane, resp. well-behaved].*

Theorem 7. *Any* $\mathtt{ABF}$ *closed under contraposition satisfies Cautious Cut for* $\mid\!\sim^{\mathsf{stab}}_{\mathtt{r}}$.

Theorem 8. *Any* $\mathtt{ABF}$ *closed under contraposition is cumulative for* $\mid\!\sim^{\mathsf{grou}}_{\mathtt{d}}$.

Theorem 9. *Where* $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}\}$*, any* $\mathtt{ABF}$ *closed under contraposition satisfies Cautious Cut for* $\mid\!\sim^{\mathsf{sem}}_{\mathtt{d}}$.

Theorem 10. *Where* $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}\}$ *and* $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$*, any well-behaved* $\mathtt{ABF}$ *is cumulative for* $\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}}$.

Theorem 11. *Where* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *is well-behaved,* $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$, $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}\}$*,* $\leq\, = \mathbb{V}^2$*,* $\mathtt{ABF}$ *is Ab-monotonic for* $\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}}$.

Theorem 12. *Where* $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$*,* $\leq\, = \mathbb{V}^2$*, and* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *is well-behaved,* $\mathtt{ABF}$ *is Ab-monotonic for* $\mid\!\sim^{\mathsf{grou}}_{\mathtt{x}}$.

We now move to our negative results by giving counter-examples for the properties from nonmonotonic reasoning relative to specific classes of frameworks (see Table 4 for an overview).

Example 22. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where (see Figure 7):

$$Ab = \{s, t, r, p\} \quad \mathcal{L} = Ab \cup \overline{Ab} \cup \{x\} \quad \mathcal{R} = \{\ s \to x;\ r, x \to \overline{p};\ p \to \overline{t}\ \}$$
$$\mathbb{V} = \{1, 2, 3, 4\} \quad v(t) = 1 \quad v(s) = 2 \quad v(p) = 3 \quad v(r) = 4$$

Notice that $\mathtt{ABF}$ is sane. The $\mathtt{d}$-grounded, unique $\mathtt{d}$-preferred, and unique $\mathtt{d}$-stable extension is $\{s, r, p\}$. Note for this that although $\{s, r\} \vdash_{\mathcal{R}} \overline{p}$, we have $\{s, r\} < p$ and so $\{s, r\}$ does not $\mathtt{d}$-defeat p. Thus, x follows. Now if we move to $\mathtt{ABF}^x$, r $\mathtt{d}$-defeats p and t is $\mathtt{d}$-defended. The $\mathtt{d}$-grounded and unique $\mathtt{d}$-preferred and $\mathtt{d}$-stable extension is now $\{s, r, t\}$. So, where $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{grou}, \mathsf{stab}\}$, while $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} p$ we have $\mathtt{ABF}^x \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} p$ and while $\mathtt{ABF}^x \mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} t$ we have $\mathtt{ABF} \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} t$.

$$\{p\} \longrightarrow \{t\}$$

(a) Defeat Diagram for $\mathtt{ABF}$

$$\{r\} \longrightarrow \{p\} \longrightarrow \{t\}$$

(b) Defeat Diagram for $\mathtt{ABF}^x$

Figure 7: A fragment of the defeat diagrams for Example 22.

Example 23. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where (see Figure 8):

$$Ab = \{p, t, s\} \quad \mathcal{L} = Ab \cup \overline{Ab} \quad \mathcal{R} = \{\ s \to \overline{p};\ s \to \overline{t}\ \}$$
$$\mathbb{V} = \{1, 2, 3\} \quad v(t) = 1 \quad v(s) = 2 \quad v(p) = 3$$

Note first that $\mathtt{ABF}$ is sane. The set $\{p, t\}$ is $\mathtt{r}$-stable, $\mathtt{r}$-preferred and $\mathtt{r}$-grounded. Note for this that there is no $\mathtt{r}$-defeat on p and p $\mathtt{r}$-defeats s and in this way defends

t. Moving to $\texttt{ABF}^p$, the set $\{t\}$ cannot be defended from the $\texttt{d}$-defeat by $\{s\}$. The only $\texttt{r}$-stable, $\texttt{r}$-preferred and $\texttt{r}$-grounded extension is now $\{s\}$, which is undefeated. Where $\mathsf{sem} \in \{\mathsf{stab}, \mathsf{pref}, \mathsf{grou}\}$, we have, for instance, $\texttt{ABF} \mid\!\sim^{\mathsf{sem}}_{\texttt{r}} t$ while $\texttt{ABF}^p \not\mid\!\sim^{\mathsf{sem}}_{\texttt{r}} t$ and $\texttt{ABF} \not\mid\!\sim^{\mathsf{sem}}_{\texttt{r}} s$ while $\texttt{ABF}^p \mid\!\sim^{\mathsf{sem}}_{\texttt{r}} s$.

$$\begin{tikzcd} \{t\} & \{s\} \arrow[l] & \{p\} \arrow[l, dashed] \end{tikzcd}$$

Defeat Diagram for $\texttt{ABF}$

$$\begin{tikzcd} \{t\} & \{s\} \arrow[l] \end{tikzcd}$$

Defeat Diagram for $\texttt{ABF}^p$

Figure 8: A fragment of the defeat diagram for Example 23.

$$\begin{tikzcd}
 & \{t_1\} \arrow[loop above] \arrow[d, shift left] & \{t_2\} \arrow[loop above] \arrow[dl, shift left] & & \\
\{r\} \arrow[r, dashed] & \{t_1, t_2\} \arrow[u, shift left] \arrow[ur, shift left] \arrow[loop below] \arrow[r] & \{s\} \arrow[r, shift left] & \{t\} \arrow[l, shift left]
\end{tikzcd}$$

(a) Defeat Diagram for $\texttt{ABF}$

$$\begin{tikzcd}
\{t_1\} \arrow[loop above] \arrow[d, shift left] & \{t_2\} \arrow[loop above] \arrow[dl, shift left] & & \\
\{t_1, t_2\} \arrow[u, shift left] \arrow[ur, shift left] \arrow[loop below] \arrow[r] & \{s\} \arrow[r, shift left] & \{t\} \arrow[l, shift left]
\end{tikzcd}$$

(b) Defeat Diagram for $\texttt{ABF}^r$

Figure 9: A fragment of the defeat diagram for Example 24.

Example 24. Let $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where: $Ab = \{r, t_1, t_2, t, s\}$, $\mathcal{L} = Ab \cup \overline{Ab}$, $\mathbb{V} = \{1, 2, 3\}$ and $v(s) = v(t) = 1$, $v(t_1) = v(t_2) = 2$ and $v(r) = 3$. Let $\mathcal{R}$ be the closure under contraposition[23] of

$$\mathcal{R}' = \{\, t_1 \to \overline{t_1};\;\; t_2 \to \overline{t_2};\;\; t_1, t_2 \to \overline{r};\;\; t_1, t_2 \to \overline{s};\;\; t \to \overline{s} \,\}$$

See Figure 9 for an illustration. We have two $\texttt{r}$-preferred extensions: $\{r, s\}$ and $\{r, t\}$. However, in $\texttt{ABF}^r$ s cannot be anymore defended from the attack by $\{t_1, t_2\}$ via r. So, the only $\texttt{r}$-preferred extension is now $\{t\}$.

Example 25. Let $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \leq, v)$ where: $Ab = \{u, t, s\}$, $\mathcal{L} = Ab \cup \overline{Ab}$, $\mathbb{V} = \{1, 2\}$, $v(u) = 1$, $v(t) = v(s) = 2$, and $\mathcal{R} = \{u, t \to \overline{s};\ s \to \overline{t}\}$. Where $\texttt{x} \in \{\texttt{r}, \texttt{d}\}$, $\{s, u\}$ is the only $\texttt{x}$-stable and $\texttt{x}$-preferred extension. Note that t cannot be defended against the attack by s. For $\texttt{ABF}^u$ we have two $\texttt{x}$-preferred and $\texttt{x}$-stable extensions, namely $\{s\}$ and $\{t\}$. Since now $\{t\} \vdash_{\mathcal{R} \cup \{\to u\}} \overline{s}$, t can defend itself against the attack by s. Altogether s ceases to be a consequence.

[23]We let the closure under contraposition $\mathcal{R}$ of a set of rules $\mathcal{R}'$ be the smallest set that contains $\mathcal{R}'$ and is such that $\vdash_{\mathcal{R}}$ satisfies contraposition.

Example 26. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where (see Figure 10)

$$Ab = \{x, p, q, s, s', s'', r\} \quad \mathcal{L} = Ab \cup \overline{Ab} \cup \{t\} \quad \mathbb{V} = \{1, 2, 3, 4, 5\}$$
$$v(x) = 1 \quad v(s') = v(s'') = 2 \quad v(s) = 3 \quad v(p) = v(q) = 4 \quad v(r) = 5$$
$$\mathcal{R} = \{\ \overline{s', s'', x};\ \overline{p, q, r};\ \overline{p, q, s};\ s \to t;\ s', s'', t \to \overline{r};\ r, s', t \to \overline{s''};\ r, s'', t \longrightarrow \overline{s'}\ \}.$$

Note that $\mathtt{ABF}$ is well-behaved. The $\mathtt{r}$-grounded set of $\mathtt{ABF}$ is $\{r, s\}$. Note that $\{s\} \vdash_{\mathcal{R}} t$. However, since $\{s', s''\} \vdash_{\mathcal{R} \cup \{\to t\}} \overline{r}$, r $\mathtt{r}$-defeats $\{s', s''\}$, reinstating x. Thus the $\mathtt{r}$-grounded extension of $\mathtt{ABF}^t$ is $\{r, s, x\}$. So, $\mathtt{ABF} \not\mathrel{|\!\sim}^{\mathsf{grou}}_{\mathtt{r}} x$ while $\mathtt{ABF}^t \mathrel{|\!\sim}^{\mathsf{grou}}_{\mathtt{r}} x$.

$$\begin{tikzcd} \{s', s''\} \arrow[r] & \{x\} & \\ \{r\} \arrow[r, dashed] & \{p, q\} \arrow[r] & \{s\} \end{tikzcd}$$

(a) Defeat Diagram for $\mathtt{ABF}$

$$\begin{tikzcd} \{s', s''\} \arrow[r] & \{x\} & \\ \{r\} \arrow[u, dashed] \arrow[r, dashed] & \{p, q\} \arrow[r] & \{s\} \end{tikzcd}$$

(b) Defeat Diagram for $\mathtt{ABF}^B$

Figure 10: A fragment of the defeat diagram for Example 26.

Example 27. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where (see Figure 11)

$$Ab = \{p, q, s, s', s'', r\} \quad \mathcal{L} = Ab \cup \overline{Ab} \cup \{t\} \quad \mathbb{V} = \{1, 2, 3, 4, 5\}$$
$$v(x) = 1 \quad v(s') = v(s'') = 2 \quad v(s) = 3 \quad v(p) = v(q) = 4 \quad v(r) = 5$$

and $\mathcal{R} =$

$$\{\overline{p, q, r};\ \overline{p, q, s};\ \overline{s, s', s'', r};\ s \to t;\ s', s'', t \to \overline{x};\ x, s', t \to \overline{s''};\ x, s'', t \to \overline{s'}\ \}$$

Notice that $\mathtt{ABF}$ is well-behaved. The $\mathtt{r}$-grounded set of $\mathtt{ABF}$ is $\{r, s, x\}$. Note that $\{s\} \vdash_{\mathcal{R}} t$. However, since $\{s', s''\} \vdash_{\mathcal{R} \cup \{\to t\}} \overline{x}$, $\{s', s''\}$ $\mathtt{d}$-defeats x and this time $\{r\}$ cannot reinstate x. Thus the $\mathtt{r}$-grounded extension of $\mathtt{ABF}^t$ is $\{r, s\}$.

$$\begin{tikzcd} \{s, s', s''\} \arrow[r] & \{x\} & \\ \{r\} \arrow[u, dashed] \arrow[r, dashed] & \{p, q\} \arrow[r] & \{s\} \end{tikzcd}$$

(a) Defeat Diagram for $\mathtt{ABF}$

$$\begin{tikzcd} \{s, s', s''\} & \{x\} & \{s', s''\} \arrow[l] \\ \{r\} \arrow[u, dashed] \arrow[r, dashed] & \{p, q\} \arrow[r] & \{s\} \end{tikzcd}$$

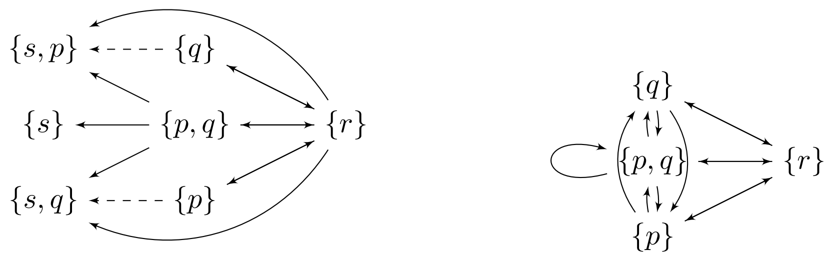
(b) Defeat Diagram for $\mathtt{ABF}^B$

Figure 11: A fragment of the defeat diagram for Example 27.

Example 28. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where (see Figure 12)

$$Ab = \{p, q, s, r\} \quad \mathcal{L} = Ab \cup \overline{Ab} \cup \{x\} \quad \mathbb{V} = \{1, 2\}$$
$$v(s) = 1 \quad v(p) = v(q) = v(r) = 2$$
$$\mathcal{R} = \{\overline{s, p, q};\ \ \overline{r, p};\ \ \overline{r, q};\ \ p, q \to x;\ \ r \to x\}$$

Observe that $\mathtt{ABF}$ is well-behaved. Note that, for $\mathtt{x} \in \{\mathtt{r}, \mathtt{d}\}$, $\{p, q\}$ and $\{s, r\}$ are the only $\mathtt{x}$-stable and $\mathtt{x}$-preferred extensions of $\mathtt{ABF}$. This means that, where $\mathsf{sem} \in \{\mathsf{stab}, \mathsf{pref}\}$, $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} x$ and $\mathtt{ABF} \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} \overline{s}$ (in view of $\{r, s\}$ being $\mathtt{x}$-stable resp. $\mathtt{x}$-preferred). However, if we add $\rightarrow s$ to $\mathcal{R}$, $\{p\} \vdash_{\mathcal{R} \cup \{\rightarrow s\}} \overline{q}$ and $\{q\} \vdash_{\mathcal{R} \cup \{\rightarrow s\}} \overline{p}$. This means that there are three $\mathtt{x}$-stable resp. $\mathtt{x}$-preferred extensions: $\{r\}$, $\{p\}$ and $\{q\}$. Note that since $\{p\}$ and $\{q\}$ do not allow to derive x, $\mathtt{ABF}^s \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} x$.



(a) Defeat Diagram for $\mathtt{ABF}$ (b) Defeat Diagram for $\mathtt{ABF}^s$

Figure 12: A fragment of the defeat diagram for Example 28.

Example 29. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where (see Figure 13)

$$Ab = \{t, s, p, q\} \quad \mathcal{L} = Ab \cup \overline{Ab} \quad \mathcal{R} = \{\ \overline{t, p, s};\ \ \overline{p, q}\ \} \quad \mathbb{V} = \{1, 2\}$$
$$v(t) = 1 \quad v(s) = v(p) = v(q) = 2$$

Notice that $\mathtt{ABF}$ is well-behaved. Then, for $\mathtt{x} \in \{\mathtt{r}, \mathtt{d}\}$, $\{s\}$ is the $\mathtt{x}$-grounded extension of $\mathtt{ABF}$ and $\mathtt{ABF} \not\mid\!\sim^{\mathsf{grou}}_{\mathtt{x}} \overline{t}$. However, when moving to $\mathtt{ABF}^t$, the $\mathtt{x}$-grounded set is $\emptyset$ since now s is $\mathtt{d}$-defeated by t and vice versa.

(a) Defeat Diagram for $\mathtt{ABF}$ (b) Defeat Diagram for $\mathtt{ABF}^t$

Figure 13: A fragment of the defeat diagram for Example 29.

Example 30. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where (see Figure 14)

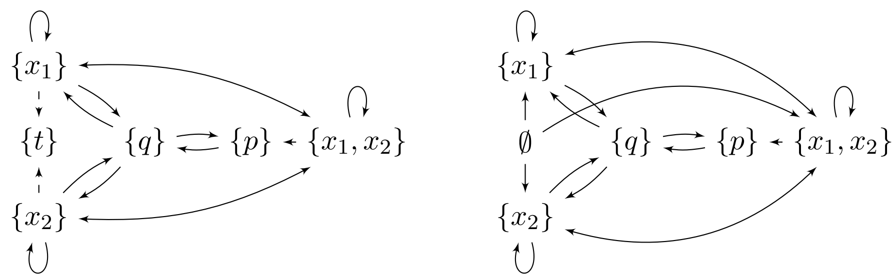
$$Ab = \{q, p, t, x_1, x_2\} \quad \mathcal{L} = Ab \cup \overline{Ab}$$
$$\mathbb{V} = \{1, 2\} \quad v(t) = 1 \quad v(x_1) = v(x_2) = v(q) = v(p) = 2$$

and $\mathcal{R}$ is the closure under contraposition of:

$$\{\ \overline{q, p};\ \ \overline{q, x_1};\ \ \overline{q, x_2};\ \ \overline{x_1, x_2, p};\ \ \overline{t, x_1};\ \ \overline{t, x_2};\ \ x_1 \rightarrow \overline{x_1};\ \ x_2 \rightarrow \overline{x_2}\ \}$$

Notice that $\mathtt{ABF}$ is closed under contraposition. The only $\mathtt{d}$-stable, $\mathtt{r}$-stable, $\mathtt{d}$-preferred, and $\mathtt{r}$-preferred extension is $\{q, t\}$. Thus, $\mathtt{ABF} \mathrel{|\!\sim}^{\mathsf{sem}}_{\mathtt{x}} t$ where $\mathsf{sem} \in \{\mathsf{stab}, \mathsf{pref}\}$ and $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$. Note that p cannot only be defended from the attack by $\{x_1, x_2\}$ by q, but $\{p, q\}$ is not conflict-free.

Now if we move to $\mathtt{ABF}^t$ and add $\rightarrow t$ to $\mathcal{R}$, also $\{p\}$ is $\mathtt{d}$-stable, $\mathtt{r}$-stable, $\mathtt{d}$-preferred and $\mathtt{r}$-preferred (in addition to $\{q\}$). The reason is that now $\emptyset \vdash_{\mathcal{R}\cup\{\rightarrow t\}} \overline{x_1}$ and $\emptyset \vdash_{\mathcal{R}\cup\{\rightarrow t\}} \overline{x_2}$ and $\{p\} \vdash_{\mathcal{R}\cup\{\rightarrow t\}} \overline{q}$. Thus, $\mathtt{ABF}^t \not\mathrel{|\!\sim}^{\mathsf{sem}}_{\mathtt{x}} q$.



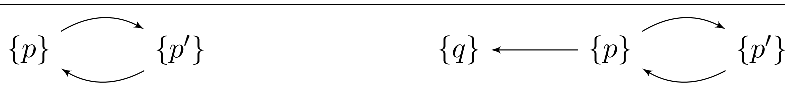
(a) Defeat Diagram for $\mathtt{ABF}$ (b) Defeat Diagram for $\mathtt{ABF}^t$

Figure 14: A fragment of the defeat diagram for Example 30.

Example 31. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where (see Figure 15)

$$Ab = \{p, q, p', s\} \quad \mathcal{L} = Ab \cup \overline{Ab} \quad \mathcal{R} = \{\ p, s \rightarrow \overline{q};\ \ p \rightarrow \overline{p'};\ \ p' \rightarrow \overline{p}\ \}$$
$$\mathbb{V} = \{1, 2\} \quad v(s) = 1 \quad v(p) = v(p') = v(q) = 2$$

Note that, for $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$, $\{s, q\}$ is the (unique) $\mathtt{x}$-grounded extension. When we move to $\mathtt{ABF}^s$ we have $p \vdash_{\mathcal{R}\cup\{\rightarrow s\}} \overline{q}$ and so the (unique) $\mathtt{x}$-grounded extension of $\mathtt{ABF}^s$ is $\emptyset$. So, we have $\mathtt{ABF} \mathrel{|\!\sim}^{\mathsf{grou}}_{\mathtt{x}} q$ while $\mathtt{ABF}^s \not\mathrel{|\!\sim}^{\mathsf{grou}}_{\mathtt{x}} q$.



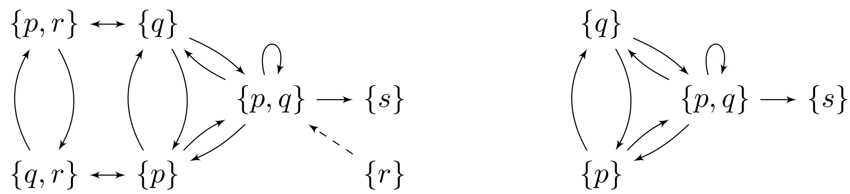
(a) Defeat Diagram for $\mathtt{ABF}$ (b) Defeat Diagram for $\mathtt{ABF}^s$

Figure 15: A fragment of the defeat diagram for Example 31.

Example 32. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where (see Figure 16)

$$Ab = \{p, q, s, r\} \quad \mathcal{L} = Ab \cup \overline{Ab} \quad \mathcal{R} = \{\ \overline{p, q};\ \ \overline{p, q, s};\ \ \overline{p, q, r}\ \}$$
$$\mathbb{V} = \{1, 2, 3\} \quad v(s) = 1 \quad v(p) = v(q) = 2 \quad v(r) = 3$$

Note that $\{r, s\}$ is the $\mathtt{r}$-grounded extension of $\mathtt{ABF}$ since r is un-defeated and it defends s from the attack by $\{p, q\}$ since r $\mathtt{r}$-defeats $\{p, q\}$. Consider now $\mathtt{ABF}^r$. Now $r \notin Ab \setminus \{r\}$ and so it does not $\mathtt{r}$-defeat $\{p, q\}$ anymore which now leaves s $\mathtt{r}$-undefended. Thus, the unique $\mathtt{r}$-grounded extension is now $\emptyset$. So we have $\mathtt{ABF} \mathrel{|\!\sim}^{\mathsf{grou}}_{\mathtt{r}} s$ while $\mathtt{ABF}^r \mathrel{|\!\not\sim}^{\mathsf{grou}}_{\mathtt{r}} s$.



(a) Defeat Diagram for $\mathtt{ABF}$ (b) Defeat Diagram for $\mathtt{ABF}^r$

Figure 16: A fragment of the defeat diagram for Example 32.

6 Connection with Preferred Subtheories

We already saw in Section 4 that for well-behaved ABFs, $\mathtt{r}$-preferred, $\mathtt{r}$-stable, $\mathtt{d}$-preferred and $\mathtt{d}$-stable extensions collapse. It turns out we can say even more about this class of extensions: they coincide with the preferred subtheories [10] of their respective ABFs.

We first adapt the definition of preferred subtheories for ABFs. For this it is convenient to suppose that $\mathbb{V}$ is a (finite) initial sequence of $\mathbb{N}$ with $\leq$ being the canonical order.[24] Preferred subtheories are obtained by selecting the *lexicographically most preferred* or *$\prec$-maximal* sets among the *maximally consistent subsets of* $\mathtt{ABF}$, denoted $\mathsf{MCS}(\mathtt{ABF})$. The latter are the maximal (w.r.t. set inclusion) sets of assumptions that contain no inconsistent set of assumptions, i.e. that contain no set of assumptions Δ such that $\Delta \setminus \{A\} \vdash_{\mathcal{R}} \overline{A}$ for some $A \in \Delta$. The set of inconsistent sets of assumptions for a given $\mathtt{ABF}$ is denoted by $\mathsf{IS}(\mathtt{ABF})$.

Definition 17. *Where* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$,

- $\mathsf{IS}(\mathtt{ABF})$ *is the set of all* $\Delta \subseteq Ab$ *such that* $\Delta \setminus \{A\} \vdash_{\mathcal{R}} \overline{A}$ *for some* $A \in \Delta$.[25]

[24] Since we assume that $(\mathbb{V}, \leq)$ is a total order, this assumption does not result in any loss of generality.

[25] Notice that this definition of inconsistent sets of assumptions is only adequate when considering *sane* $\mathtt{ABF}$s, as we presuppose in this section. When $\mathtt{ABF}$s contain paradoxical assumptions, it might be the case that $\Delta \vdash_{\mathcal{R}} \overline{A}$ for some $A \in \Delta$ while there is no $\Delta' \subseteq \Delta \setminus \{A\}$ for which $\Delta' \vdash_{\mathcal{R}} \overline{A}$ and accordingly, this will have to be taken into account when generalizing the definition of $\mathsf{IS}(\mathtt{ABF})$ to such $\mathtt{ABF}$s.

- $\mathsf{CS}(\mathtt{ABF})$ *is the set of all* $\Delta \subseteq Ab$ *such that for no* $\Theta \in \mathsf{IS}(\mathtt{ABF})$*,* $\Theta \subseteq \Delta$*.*
- $\mathsf{MCS}(\mathtt{ABF})$ *is the set of all* $\Delta \in \mathsf{CS}(\mathtt{ABF})$ *that are* $\subseteq$*-maximal.*
- *Where* $\Delta \subseteq Ab$ *and* $i \in \mathbb{N}$*,* $\pi_i(\Delta) = \{A \in \Delta \mid v(A) = i\}$*.*
- $\prec \subseteq \wp(Ab) \times \wp(Ab)$ *is defined as:* $\Delta \prec \Theta$ *iff there is an* $i \geq 1$ *such that* $\pi_j(\Delta) = \pi_j(\Theta)$ *for every* $j > i$ *and* $\pi_i(\Delta) \subset \pi_i(\Theta)$*.*
- $\mathsf{MCS}_{\prec}(\mathtt{ABF}) = \max_{\prec}(\mathsf{MCS}(\mathtt{ABF}))$[26]

To illustrate the above definition, we give an Example.

Example 33. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ with $\mathcal{R} = \{\ \overline{p, q};\ \ \overline{r, q, s}\ \}$ and $Ab = \{p, q, r, s\}$. Furthermore let $\mathbb{V} = \{1, 2, 3\}$ and $v(s) = 3$, $v(r) = 2$ and $v(p) = v(q) = 1$. We have $\mathsf{IS}(\mathtt{ABF}) = \{\Theta \subseteq Ab \mid \Theta \supseteq \{p, q\}$ or $\Theta \supseteq \{r, q, s\}\}$. This means that

$$\mathsf{MCS}(\mathtt{ABF}) = \{\{p, r, s\}, \{q, r\}, \{q, s\}\}.$$

Since $\pi_3(\{q, r\}) = \emptyset$ and $\pi_3(\{p, r, s\}) = \pi_3(\{q, s\}) = \{s\}$, $\{q, r\} \prec \{p, r, s\}$ and $\{q, r\} \prec \{q, s\}$. Since $\pi_2(\{p, r, s\}) = \{r\}$ whereas $\pi_2(\{q, s\}) = \emptyset$, $\{q, s\} \prec \{p, r, s\}$. This means that

$$\mathsf{MCS}_{\prec}(\mathtt{ABF}) = \{\{p, r, s\}\}.$$

Theorem 13. *For any well-behaved* $\mathtt{ABF}$ *we have:*

$$\mathsf{MCS}_{\prec}(\mathtt{ABF}) = \mathtt{r}\text{-}\mathsf{pref}(\mathtt{ABF}) = \mathtt{d}\text{-}\mathsf{pref}(\mathtt{ABF}) = \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF}) = \mathtt{d}\text{-}\mathsf{stab}(\mathtt{ABF})$$

This theorem is proven in Appendix D.

7 Related Work

Properties known from non-monotonic logic have also been studied in [18]. Their definitions of (restricted forms of) monotony- and cut-properties are extension-based. For our discussion it will suffice to give one example for skeptical consequence in a non-prioritized setting. Given an ABA framework $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{})$, a sem-extension Δ, an $A \in \mathrm{Cn}_{\mathcal{R}}(\Delta) \setminus Ab$ [resp. $A \in \mathrm{Cn}_{\mathcal{R}}(\Delta) \cap Ab$], and let $\mathtt{ABF}' = (\mathcal{L}, \mathcal{R} \cup \{\top \to A\}, Ab, \overline{})$ [resp. $\mathtt{ABF}' = (\mathcal{L}, \mathcal{R} \cup \{\top \to A\}, Ab \setminus \{A\}, \overline{})$]:

> *Skeptical strict [resp. asm] mon* is satisfied iff for all sem-extensions Δ' of $\mathtt{ABF}'$, $\mathrm{Cn}_{\mathcal{R}}(\Delta) \subseteq \mathrm{Cn}_{\mathcal{R}}(\Delta')$.

[26]Since we assume $\mathbb{V}$ to be finite, $\max_{\prec}(\mathsf{MCS}(\mathtt{ABF}))$ will never be empty.

At first sight this may seem to amount to a semantical counter-part of cautious monotony as defined above in Definition 14 for ABA-induced non-monotonic entailment relations. However, the more natural semantic counter-part for the skeptical versions are (see e.g., [4, 2] where such versions are used):

Strict version: Every sem-extension Δ' of $\mathtt{ABF}'$ is also a sem-extension of $\mathtt{ABF}$.

asm version: For every sem-extension Δ' of $\mathtt{ABF}'$, $\Delta' \setminus \{A\}$ is a sem-extension of $\mathtt{ABF}$.

The strict semantic versions are the ones we prove in this paper (in addition to the properties of the nonmonotonic entailment relations).

In fact, many typical non-monotonic logics which are (skeptically) cumulative don't satisfy *skeptical mon* as defined in [18]. Take as an example the inevitable consequence from [35]: Given a set of formulas Σ, A is an *inevitable consequence* of Σ iff it follows from all maximal consistent subsets of Σ. The natural corresponding notion to an "extension" is thus a maximal consistent subset. Take, for instance, $\Sigma = \{p, \neg p\}$. The maximal consistent subsets are $E_1 = \{p\}$ and $E_2 = \{\neg p\}$. Now, clearly $p \vee \neg p$ follows from E_1. So consider $\Sigma' = \Sigma \cup \{p \vee \neg p\}$. The maximal consistent subsets are $E_1 \cup \{p \vee \neg p\}$ and $E_2 \cup \{p \vee \neg p\}$. However, $\mathrm{Cn}(E_2 \cup \{p \vee \neg p\}) \not\subseteq \mathrm{Cn}(E_1)$.

We can observe the same discrepancy for ABA frameworks. Take the well-behaved framework $\mathtt{ABF}$ (without preferences) consisting of $Ab = \{p, q\}$, $\mathcal{R} = \{p \to s, q \to s\}$ and where $\overline{p} = q$ and $\overline{q} = p$. We have two preferred/stable extensions, $\{p\}$ and $\{q\}$ and $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} s$ for $\mathtt{x} \in \{\mathtt{r}, \mathtt{d}\}$ and $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}\}$. Moving to $\mathtt{ABF}^s$ we still have the same extensions. However, skeptical mon as defined in [18] is not satisfied since, for instance, $\mathrm{Cn}(\{p\}) \not\subseteq \mathsf{Cn}(\{q\})$. Recall that with Theorem 10, $\mathtt{ABF}$ is cumulative.

Altogether, it seems the skeptical mon property is most sensible in the context of single-extension semantics.

In sum, the properties studied in [18] don't correspond to (skeptical) Cautious Monotony and Cautious Cut as usually defined in nonmonotonic logic. To the best of our knowledge, the current paper provides the first study of this type ABA-systems.

$\mathrm{ABA}^{\mathsf{r}}$ was the first approach to assumption-based argumentation where priorities determine defeat relations [16]. In [23], additionally $\mathrm{ABA}^{\mathsf{d}}$ was introduced and translations between $\mathrm{ABA}^{\mathsf{d}}$ and $\mathrm{ABA}^{\mathsf{r}}$ were presented. Although similar in spirit, the present paper is more general in several respects: we investigate properties for both $\mathrm{ABA}^{\mathsf{d}}$ and $\mathrm{ABA}^{\mathsf{r}}$ and in this way offer a systematic comparison between these two frameworks. We have shown that for well-behaved frameworks $\mathrm{ABA}^{\mathsf{d}}$ and $\mathrm{ABA}^{\mathsf{r}}$ give rise to the same stable and preferred extensions.

Finally, we investigate properties for flat ABFs, ABFs closed under contraposition and well-behaved ABFs, whereas [18] investigate nonmonotonic inference properties for frameworks under so called *weak contraposition*, a variant of contraposition:[27]

Definition 18 (Weak Contraposition, [1, 18])**.** *For all* $\Delta \cup \{A\} \subseteq Ab$*, if* $\Delta \vdash_{\mathcal{R}} \overline{A}$ *and* $\Delta < A$*, then there is a* $B \in \min(\Delta \cap Ab)$ *and a* $\Delta' \subseteq (\Delta \setminus \{B\}) \cup \{A\}$ *for which* $\Delta' \vdash_{\mathcal{R}} \overline{B}$*.*

Many of the results shown in this paper for contrapositive ABFs do not hold for weakly contrapositive ABFs. For instance, the correspondence between preferred subtheories and $\mathtt{d}$-stable, $\mathtt{r}$-stable, $\mathtt{d}$-preferred and $\mathtt{r}$-preferred extensions for well-behaved ABFs in Theorem 13 does not hold anymore for weakly well-behaved ABFs (those that are weakly contrapositive and sane (Def. 11)). Similarly, for weakly well-behaved ABFs it is not in general the case that $\mathsf{r\text{-}pref}(\mathtt{ABF}) \subseteq \mathsf{d\text{-}pref}(\mathtt{ABF})$ or vice versa (in contrast to the bi-conditionals of Theorem 6). Both these claims are demonstrated in the following example.

Example 34. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ with $Ab = \{p, p', p'', r, r', s\}$ where $v(r') = 1$, $v(s) = v(p) = v(p') = v(p'') = 2$ and $v(r) = 3$. Furthermore, let

$$\mathcal{R} = \{\ p \to \overline{p'};\ \ p' \to \overline{p''};\ \ p'' \to \overline{p};\ \ p, s \to \overline{r};\ \ p, s \to \overline{r'};\ \ r, p \to \overline{s}\ \}.$$

See Figure 17 for an illustration. Note that this $\mathtt{ABF}$ is weakly well-behaved. $\{r, r'\}$ is $\mathtt{r}$-preferred in this $\mathtt{ABF}$ but not $\mathtt{d}$-preferred. The only $\mathtt{d}$-preferred extension is $\{r\}$. Furthermore, note that $\{r, r', p\} \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$. Finally, observe that there are no $\mathtt{r}$-stable or $\mathtt{d}$-stable extensions.

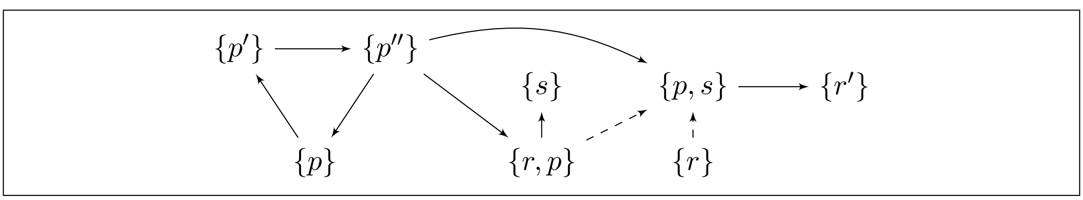


Figure 17: A fragment of the defeat diagram for Example 34. Full lines represent $\mathtt{d}$-defeats whereas dashed lines represent proper $\mathtt{r}$-defeats.

Also, CM-Ab and CM do not hold for weakly well-behaved ABFs and for preferred and stable semantics, in contrast to Theorem 10 where CM (and thus CM-Ab) is shown for well-behaved ABFs. We give a counter-example.

[27] We also notice that [18] don't restrict their study to *total* preorders.

Example 35. We give an example of a sane ABF closed under weak contraposition. Let $Ab = \{s, p, q, x_1, x_2, x'_1, x''_1, x'_2, x''_2\}$, $v(q) = 1$, $v(x_1) = v(x'_1) = v(x''_1) = v(x_2) = v(x'_2) = v(x''_2) = v(s) = v(p) = 2$, and $\mathcal{R}$ be

$$\{\ \overline{s,p};\ \overline{s,x_1};\ \overline{s,x'_1};\ \overline{s,x''_1};\ \overline{s,x_2};\ \overline{s,x'_2};\ \overline{s,x''_2};\ \overline{x_1,x_2,p};\ \overline{q,x_1};\ \overline{q,x'_1};\ \overline{q,x''_1};\ \overline{q,x_2};\ \overline{q,x'_2};\ \overline{q,x''_2};\ x_1 \to \overline{x'_1};\ x'_1 \to \overline{x''_1}\ x''_1 \to \overline{x_1};\ x_2 \to \overline{x'_2}\ x'_2 \to \overline{x''_2};\ x''_2 \to \overline{x_2}\ \}.$$

See Figure 18 for an illustration. The only d-stable, r-stable, d-preferred, and r-preferred extension is $\{s, q\}$. Thus, ABF $\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} q$ where $\mathsf{sem} \in \{\mathsf{stab}, \mathsf{pref}\}$ and $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$. Note that p cannot be defended from the attack by $\{x_1, x_2\}$, except by s with which it is not conflict-free.

Now if we move to $\mathtt{ABF}^q$ and add $\to q$ to $\mathcal{R}$, also $\{p\}$ is d-stable, r-stable, d-preferred and r-preferred (in addition to $\{s\}$). The reason is that now $\emptyset \vdash_{\mathcal{R} \cup \{\to q\}} \overline{x_1}$ and $\emptyset \vdash_{\mathcal{R} \cup \{\to q\}} \overline{x_2}$ and $\{p\} \vdash_{\mathcal{R} \cup \{\to q\}} \overline{s}$. Thus, $\mathtt{ABF}^q \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} s$.

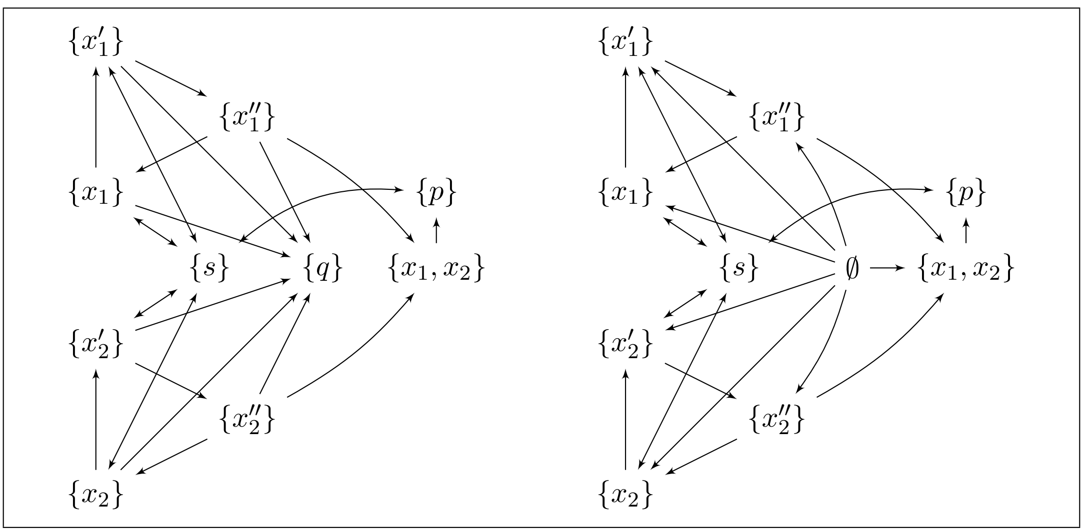


Figure 18: A fragment of the defeat diagram for Example 35 (left side for ABF and right side for $\mathtt{ABF}^q$). Full lines represent d-defeats whereas dashed lines represent proper r-defeats.

Nevertheless, we can also report on some novel positive results for weak contraposition. For instance, the Fundamental Lemma holds for weakly contrapositive ABFs and r-defeat generalizing Theorems 4 and 1. This is shown in Appendix E (Lemma 10).[28] Moreover, one can generalize Theorem 5 for weakly contrapositive

[28]To the best of our knowledge the Fundamental Lemma has only been shown for contrapositive

ABFs (see Appendix F, Theorem 16) which, for instance, shows that the $\mathtt{d}$-stable and $\mathtt{r}$-stable extensions of a weakly contrapositive ABFs coincide.

8 Discussion

The results of Section 6 shed further light on the connection between ABA and other formalisms for defeasible reasoning. Previously works have compared ABA to logic programming [12], as well as ASPIC, nonmonotonic logics with preferential semantics [24], autoepistemic logic [9] and default logic [9].

We moreover note that the coincidence between $\mathtt{x}$-preferred and $\mathtt{x}$-stable semantics (for any $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$) shown in Theorem 6 is also interesting from a computational point of view. It means that it is sufficient to establish membership of assumptions in an admissible set of assumptions in order to show that an assumption is part of a stable set of assumptions [19]. In general, the latter is harder to show, since it is necessary to show for every assumption that it is either part of the stable set of assumptions or defeated by the set of assumptions under consideration. On the other hand, proof theories for preferred sets of assumptions do not need take into account all sets of assumptions. In view of such computational considerations, [19] showed that preferred subtheories (based on classical logic) coincide with stable and preferred extensions in deductive argumentation based on classical logic. Theorem 6 can be seen as a generalization of this result since we allow for deductive argumentation based on *any* well-behaved rule base, in addition to classical logic.

We do not believe that our results give conclusive evidence in favour of either $\text{ABA}^{\mathtt{d}}$ and $\text{ABA}^{\mathtt{r}}$ especially since there seems to be a trade-off between consistency and the satisfaction of the Fundamental Lemma (see Section 3). Furthermore, we have shown that $\text{ABA}^{\mathtt{d}}$ and $\text{ABA}^{\mathtt{r}}$ coincide when restricting attention to *well-behaved* ABFs (see Section 4). For such ABFs, when making use of the preferred or stable semantics, it seems computationally more efficient to make use of $\text{ABA}^{\mathtt{d}}$, since adding $\mathtt{r}$-defeats does not change the consequences of the ABF under consideration. Furthermore, this class ABFs avoids the trade-off between consistency and the satisfaction of the Fundamental Lemma.

With regards to properties for non-monotonic reasoning, we notice that two wide classes of consequence relations are *cumulative*: $\mid\!\sim^{\mathtt{d}}_{\mathsf{grou}}$ for ABFs that are closed under contraposition and $\mid\!\sim^{\mathtt{x}}_{\mathsf{sem}}$ for any $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$ and $\mathsf{sem} \in \{\mathsf{stab}, \mathsf{pref}\}$ for any ABF that is well-behaved. Furthermore, when priorities over the assumptions are not taken into account, *any* $\mathtt{x}$*-semantics* is rational for *well-behaved* ABF*s*. Even though we

ABFs and $\mathtt{r}$-defeat before in publication [18]. In the unpublished manuscript [17], however, one finds a proof of the Fundamental Lemma for weakly contrapositive ABFs.

believe these results offer valuable insights into both the behaviour of consequence relations based on assumption-based argumentation and the relation between such consequence relations and other formalisms for defeasible reasoning, we do not believe that these results offer conclusive reasons to say that one approach is "better" or "worse" than the other. Indeed, even though the properties for cumulative inference relations have been claimed to "constitute a basic set of principles that any reasonable account of defaults must obey" [22], the generality of this claim has been put into doubt by Alexander Bochman [6, 7, 8], who posits a distinction between *explanatory* and *preferential* reasoning, where only for the latter cumulativity is feasible. Furthermore, some of the properties considered in Section 5 are not outside of controversy, such as rational monotony (cf. e.g. [37]). In sum, we submit that the feasibility of the postulates for nonmonotonic reasoning depends on the precise context of application. Once the feasibility of the postulates for nonmonotonic reasoning has been decided for a given application context, our result offer an indication of which formalism is the most apt.

Mainly for reasons of space we have focused in this contribution on skeptical consequence relations. Frequently, in the context of nonmonotonic logics also a credulous notion of consequence is considered: A is credulously entailed by an ABF (according to a semantics sem) iff it follows from at least one sem-extensions. We nevertheless issue a word of warning. Our positive and negative results obtained for skeptical consequence do in many cases not apply to credulous entailment.

The following Example 36 illustrates that in contradistinction to $\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}}$ for $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$ and $\mathsf{sem} \in \{\mathsf{stab}, \mathsf{pref}\}$, Cautious Monotony does not hold for credulous entailments based on well-behaved $\mathtt{ABF}$s:

Example 36. Let $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ where:

$$Ab = \{p, q\} \quad \mathcal{L} = Ab \cup \overline{Ab} \quad \mathbb{V} = \{1\} \quad \mathcal{R} = \{\ \overline{p, q}\ \}$$
$$v(p) = v(q) = 1$$

For any $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$, $\mathtt{ABF}$ has two $\mathtt{x}$-preferred extensions which are also $\mathtt{x}$-stable: $\{p\}$ and $\{q\}$. Since q is contained in at least one extension, q is credulously derivable. If we move to $\mathtt{ABF}^q$, $\emptyset \vdash_{\mathcal{R} \cup \{\to q\}} \overline{p}$, which means that there is only one $\mathtt{x}$-preferred (and $\mathtt{x}$-stable) set of assumptions: $\emptyset$.

$$\begin{tikzcd} \{p\} \arrow[r, bend left] & \{q\} \arrow[l, bend left] \end{tikzcd}$$

(a) Defeat Diagram for $\mathtt{ABF}$

$$\begin{tikzcd} \{p\} & \emptyset \arrow[l] \end{tikzcd}$$

(b) Defeat Diagram for $\mathtt{ABF}^q$

Figure 19: A fragment of the defeat diagram for Example 36.

In view of such examples, we leave the investigation of the generalization of the results in this paper to credulous consequence relations for future work. Furthermore, we plan to generalize our results to non-flat ABFs and preorders that are not necessarily total in future work.

References

[1] Ziyi Bao, Kristijonas Čyras, and Francesca Toni. Abaplus: Attack reversal in abstract and structured argumentation with preferences. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 420–437. Springer, 2017.

[2] Diderik Batens. A universal logic approach to adaptive logics. *Logica universalis*, 1(1):221–242, 2007.

[3] Mathieu Beirlaen, Jesse Heyninck, Pere Pardo, and Christian Straßer. Argument strength in formal argumentation. *Journal of Applied Logics-IfCoLog Journal of Logics and their Applications*, 5(3):629–675, 2018.

[4] S. Benferhat, D. Dubois, and H. Prade. Some syntactic approaches to the handling of inconsistent knowledge bases: A comparative study part 1: The flat case. *Studia Logica*, 58(1):17–45, 1997.

[5] Philippe Besnard and Anthony Hunter. Argumentation based on classical logic. In Iyad Rahwan, Guillermo R Simari, and Johan van Benthem, editors, *Argumentation in Artificial Intelligence*, pages 133–152. Springer, 2009.

[6] Alexander Bochman. *Explanatory nonmonotonic reasoning*, volume 4. World scientific, 2005.

[7] Alexander Bochman. Two paradigms of nonmonotonic reasoning. In *ISAIM*. Citeseer, 2006.

[8] Alexander Bochman. *A logical theory of nonmonotonic inference and belief change*. Springer Science & Business Media, 2013.

[9] Andrei Bondarenko, Phan Minh Dung, Robert A Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1):63–101, 1997.

[10] Gerhard Brewka. Preferred subtheories: An extended logical framework for default reasoning. In *IJCAI*, volume 89, pages 1043–1048, 1989.

[11] Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

[12] Martin Caminada and Claudia Schulz. On the equivalence between assumption-based argumentation and logic programming. *Journal of Artificial Intelligence Research*, 60:779–825, 2017.

[13] Martinus Wigbertus Antonius Caminada, Sanjay Modgil, and Nir Oren. Preferences and unrestricted rebut. *Computational Models of Argument*, 2014.

[14] Martinus Wigbertus Antonius Caminada, Samy Sá, Joao Alcântara, and Wolfgang Dvořák. On the difference between assumption-based argumentation and abstract ar-

gumentation. *IfCoLog Journal of Logics and their Applications*, 2015.

[15] Kristijonas Čyras, Xiuyi Fan, Claudia Schulz, and Francesca Toni. Assumption-based argumentation: Disputes, explanations, preferences. *Handbook of Formal Argumentation*, 2018.

[16] Kristijonas Čyras and Francesca Toni. ABA+: assumption-based argumentation with preferences. In *KR*, pages 553–556, 2016.

[17] Kristijonas Čyras and Francesca Toni. ABA+: assumption-based argumentation with preferences. *arXiv preprint arXiv:1610.03024*, 2016.

[18] Kristijonas Čyras and Francesca Toni. Properties of ABA+ for non-monotonic reasoning. *arXiv preprint arXiv:1603.08714*, 2016.

[19] M. D'Agostino and S. Modgil. A study of argumentative characterisations of preferred subtheories. In *Proc. IJCAI'18*, pages 1788–1794, 2018.

[20] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–358, 1995.

[21] Dov M Gabbay. Theoretical foundations for non-monotonic reasoning in expert systems. In *Logics and models of concurrent systems*, pages 439–457. Springer, 1985.

[22] Hector Geffner and Judea Pearl. Conditional entailment: Bridging two approaches to default reasoning. *Artificial Intelligence*, 53(2-3):209–244, 1992.

[23] Jesse Heyninck, Pere Pardo, and Christian Straßer. Assumption-based approaches to reasoning with priorities. In *AI*3 2017.*, pages 58–72, 2017.

[24] Jesse Heyninck and Christian Straßer. Relations between assumption-based approaches in nonmonotonic logic and formal argumentation. In *16th International Workshop on Non-Monotonic Reasoning*, page 65, 2016.

[25] Jesse Heyninck and Christian Straßer. Revisiting unrestricted rebut and preferences in structured argumentation. In *IJCAI*, pages 1088–1092, 2017.

[26] Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *AI*, 44(1):167–207, 1990.

[27] Tuomo Lehtonen, Johannes P Wallner, and Matti Järvisalo. From structured to abstract argumentation: Assumption-based acceptance via af reasoning. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 57–68. Springer, 2017.

[28] Witold Łukaszewicz. Considerations on default logic: an alternative approach 1. *Computational intelligence*, 4(1):1–16, 1988.

[29] David Makinson. *Bridges from classical to nonmonotonic logic*. College Publications, 2005.

[30] Wiktor Marek and Mirosław Truszczyński. Autoepistemic logic. *Journal of the ACM (JACM)*, 38(3):587–618, 1991.

[31] Sanjay Modgil and Henry Prakken. A general account of argumentation with preferences. *Artif. Intell.*, 195:361–397, 2013.

[32] Sanjay Modgil and Henry Prakken. The aspic+ framework for structured argumenta-

tion: a tutorial. *A&C*, 5(1):31–62, 2014.

[33] Henry Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.

[34] Raymond Reiter. A logic for default reasoning. *AI*, 13(1):81–132, 1980.

[35] Nicholas Rescher and Ruth Manor. On inference from inconsistent premises. *Theory and Decision*, 1:179–217, 1970.

[36] Claudia Schulz and Francesca Toni. Labellings for assumption-based and abstract argumentation. *International Journal of Approximate Reasoning*, 84:110–149, 2017.

[37] Robert Stalnaker. What is a nonmonotonic consequence relation? *Fundamenta Informaticae*, 21(1, 2):7–21, 1994.

[38] Francesca Toni. Assumption-based argumentation for closed and consistent defeasible reasoning. In *Annual Conference of the Japanese Society for Artificial Intelligence*, pages 390–402. Springer, 2007.

[39] Francesca Toni. A tutorial on assumption-based argumentation. *Argument and Computation*, 5(1):89–117, 2014.

Appendices

The order of appearance of results in these appendices does not necessarily mirror the order of the statements of these results in the paper but is primarily based on logical dependency.[29]

Appendix A On the Requirement of Relevance in $\mathcal{R}$-deductions

The reader may have noticed that in this paper we have based $\mathtt{x}$-defeat (where $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$) on a notion of $\mathcal{R}$-deduction that demands minimality in its support (see Definition 6).[30] In the literature often a monotonic deducability relation is used as the basis for the definition of defeat (see e.g. [9, 16, 36]). In this appendix we show that both approaches lead to the same result in terms of semantic selections whence our results also apply to the presentation of ABA based on a monotonic deducability relation (see Corollary 3 and Theorem 14 below). In order to show this

[29]Whenever results in the Appendix apply to $\mathtt{d}$-defeat based frameworks, they also apply to $\mathtt{f}$-defeat based frameworks (since any $\mathtt{d}$-based framework with the trivial ordering $v(A) = v(A')$ for all $A, A' \in Ab$ gives rise to the same semantic selections as the corresponding $\mathtt{f}$-based framework). In such cases we sometimes omit the discussion of $\mathtt{f}$-based frameworks for the sake of brevity.

[30]Minimality of support is also required in other prominent approaches to argumentation such as [5].

we first define $\mathtt{x}$-defeat*:

Definition 19 (Attack*, defeat*, reverse defeat*)**.** *Given* $\Delta \cup \{A\} \subseteq Ab$*,*

- Δ attacks* A *iff* $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta)$*.*
- Δ attacks* Θ *iff* Δ *attacks** *some* $A \in \Theta$*.*
- Δ $\mathtt{d}$-defeats* A *iff some* $\Delta' \subseteq \Delta$ *attacks** A *for which* $\Delta' \not< A$*.*
- Δ $\mathtt{d}$-defeats* Θ *iff* Δ $\mathtt{d}$*-defeats** *some* $A \in \Theta$*.*
- Δ $\mathtt{r}$-defeats* Θ *iff* Δ $\mathtt{d}$*-defeats** Θ*, or for some* $\Theta' \subseteq \Theta$ *and* $A \in \Delta$*,* Θ' *attacks** A *with* $\Theta' < A$

We first show that the notions of $\mathtt{f}$- and $\mathtt{d}$-defeat are extensionally equivalent to $\mathtt{f}$- and $\mathtt{d}$-defeat*.

Proposition 1. *Where* $\mathtt{x} \in \{\mathtt{f}, \mathtt{d}\}$ *and* $\Delta \cup \{A\} \subseteq Ab$*,* Δ $\mathtt{x}$*-defeats* A *iff* Δ $\mathtt{x}$*-defeats** A*.*

Proof. ($\Rightarrow$) Suppose Δ $\mathtt{f}$-defeats [resp. $\mathtt{d}$-defeats] A. Thus, there is a $\Delta' \subseteq \Delta$ for which $\Delta' \vdash_{\mathcal{R}} \overline{A}$ [and $\Delta' \not< A$]. Thus, $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta')$ [and $\Delta' \not< A$] and hence Δ $\mathtt{f}$-defeats* [$\mathtt{d}$-defeats*] A.

($\Leftarrow$) Suppose Δ $\mathtt{f}$-defeats* [resp. $\mathtt{d}$-defeats*] A. Thus, there is a $\Delta' \subseteq \Delta$ for which $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta')$ [and $\Delta' \not< A$]. Hence, there is a $\Delta'' \subseteq \Delta'$ for which $\Delta'' \vdash_{\mathcal{R}} \overline{A}$ [and $\Delta'' \not< A$ (since $v(B) \leq v(C)$ for all $B \in \min(\Delta')$ and all $C \in \min(\Delta'')$)]. Thus, Δ'' and so Δ $\mathtt{f}$-defeats [resp. $\mathtt{d}$-defeats] A. □

We can prove one direction also for $\mathtt{r}$-defeat.

Proposition 2. *Where* $\Delta, \Theta \subseteq Ab$*, if* Δ $\mathtt{r}$*-defeats* Θ *then* Δ $\mathtt{r}$*-defeats** Θ*.*

Proof. Suppose Δ $\mathtt{r}$-defeats Θ. If Δ $\mathtt{d}$-defeats Θ then by Proposition 1 it also $\mathtt{d}$-defeats* and thus also $\mathtt{r}$-defeats* Θ. Else, there is a $\Theta' \subseteq \Theta$ and a $B \in \Delta$ for which Θ' $\mathtt{f}$-defeats B and $\Theta' < B$. By Proposition 1, Θ' $\mathtt{f}$-defeats* B. Thus, Δ $\mathtt{r}$-defeats* Θ. □

However, $\mathtt{r}$-defeats and $\mathtt{r}$-defeats* do not always coincide as the following example shows.

Example 37. Let $Ab = \{p, q, r\}$, $\mathcal{L} = Ab \cup \overline{Ab}$, $\mathbb{V} = \{1, 2\}$, $v(q) = 1$, $v(p) = v(r) = 2$, and $\mathcal{R} = \{p \to \overline{r}\}$. Note that while r $\mathtt{r}$-defeats* $\{p, q\}$ since $\overline{r} \in \mathrm{Cn}_{\mathcal{R}}(\{p, q\})$ and $\{p, q\} < r$, it is not the case that r $\mathtt{r}$-defeats $\{p, q\}$. The reason is that $\{p, q\}$ does not attack r *with*[31] $\{p, q\}$ (since $\{p, q\} \nvdash_{\mathcal{R}} \overline{r}$ in view of the nonmonotonicity of $\vdash_{\mathcal{R}}$) but rather with $\{p\}$ where $\{p\} \not< r$.

Can we say more about cases where some Δ $\mathtt{r}$-defeats* Θ but it does not amount to $\mathtt{r}$-defeat? It is interesting to notice that such defeats always give rise to inverse $\mathtt{d}$-defeat. In fact, there will be a $\Theta' \subset \Theta$ that $\mathtt{d}$-defeats Δ and that is not $\mathtt{r}$-defeated* by Δ (see Proposition 3 below). In view of this such $\mathtt{r}$-defeats* are really redundant since in the overall attack dynamics the right-hand side (e.g. Θ) 'will win over' the left hand side (e.g. Δ) anyway (see Theorem 14 below).

Proposition 3. *Where $\Delta, \Theta \subseteq Ab$, if Δ $\mathtt{r}$-defeats* Θ, then Δ $\mathtt{r}$-defeats Θ or*

1. *Θ $\mathtt{d}$-defeats and so also $\mathtt{r}$-defeats Δ, and*
2. *there is a $\Theta' \subset \Theta$ that $\mathtt{d}$-defeats Δ while Δ does not $\mathtt{r}$-defeat* Θ'.*

Proof. Suppose Δ $\mathtt{r}$-defeats* Θ while it doesn't $\mathtt{r}$-defeat Θ. We show first item 1. By Proposition 1, Δ does not $\mathtt{d}$-defeat* Θ since otherwise it also $\mathtt{d}$-defeats and so $\mathtt{r}$-defeats Θ. Hence, there is a $\Theta' \subseteq \Theta$ and a $B \in \Delta$ for which Θ' $\mathtt{f}$-defeats* B and $\Theta' < B$. By Proposition 1, Θ' $\mathtt{f}$-defeats B. Note that Θ' does not $\mathtt{f}$-defeat B with Θ' since otherwise Δ $\mathtt{r}$-defeats Θ. Thus, there is a $\Theta'' \subset \Theta'$ for which $\Theta'' \vdash_{\mathcal{R}} \overline{B}$. Note that $\Theta'' \not< B$ since otherwise Δ $\mathtt{r}$-defeats Θ. Thus, Θ $\mathtt{d}$-defeats Δ in B.

We now move to item 2. Having established item 1 we know there is a $\subseteq$-minimal $\Theta' \subseteq \Theta$ that $\mathtt{d}$-defeats Δ. Assume for a contradiction that Δ $\mathtt{r}$-defeats* Θ'. In view of Proposition 1, Δ does not $\mathtt{d}$-defeat* Θ' since then Δ $\mathtt{r}$-defeats Θ. So, there is a $\Theta'' \subseteq \Theta'$ and a $C \in \Delta$ for which Θ'' $\mathtt{f}$-defeats* C and $\Theta'' < C$. By Proposition 1, Θ'' $\mathtt{f}$-defeats C. Hence, there is a $\Theta''' \subseteq \Theta''$ for which $\Theta''' \vdash_{\mathcal{R}} \overline{C}$. In view of the fact that Δ does not $\mathtt{r}$-defeat Θ, $\Theta''' \not< C$. Thus, since $\Theta'' < C$ and $\Theta'' \subseteq \Theta'$, $\Theta''' \subset \Theta'$. Since Θ''' $\mathtt{d}$-defeats Δ this contradicts the $\subseteq$-minimality of Θ'. $\square$

In order to show that $\mathtt{x}$-defeat and $\mathtt{x}$-defeat* give rise to the same semantic selections we first define the semantics based on $\mathtt{x}$-defeat*. This is perfectly analogous to Definition 7.

Definition 20 (Argumentation semantics* [9])**.** *Given some sets $\Delta, \Delta' \subseteq Ab$, we define for each $\mathtt{x} \in \{\mathtt{f}, \mathtt{d}, \mathtt{r}\}$:*

- *Δ is* $\mathtt{x}$-conflict-free* *iff it does not $\mathtt{x}$-defeat* itself.*

[31] Recall the last item of Definition 6.

- Δ x-defends* Δ' *iff for any* $\Delta'' \subseteq Ab$ *that* x*-defeats** Δ', Δ x*-defeats** Δ''.
- Δ *is* x-admissible* *iff* Δ *is* x*-conflict-free** *and* Δ x*-defends** *itself.*
- Δ *is* x-complete* *iff* Δ *is* x*-admissible** *and* Δ *contains every* $\Delta' \subseteq Ab$ *it* x*-defends**.
- Δ *is* x-preferred* *iff* Δ *is* $\subseteq$*-maximally* x*-admissible**.
- Δ *is* x-grounded* *iff* Δ *is* $\subseteq$*-minimally* x*-complete**.
- Δ *is* x-stable* *iff* Δ *is* x*-conflict-free** *and* Δ x*-defeats** *every* $A \in Ab \setminus \Delta$.

x*-conflict-free**, x*-naive**, x*-admissible**, x*-complete**, x*-grounded**, x*-preferred** *resp.* x*-stable** *will be denoted by* x-cf*, x-adm*, x-comp*, x-grou*, x-pref*, x-stab*. *For any semantics* $\mathsf{sem}^\star \in \{\mathsf{cf}^\star, \mathsf{adm}^\star, \mathsf{comp}^\star, \mathsf{grou}^\star, \mathsf{pref}^\star, \mathsf{stab}^\star\}$, x-sem*(ABF) *is defined as the sets of assumptions that are* x-sem*, *as defined above.*

The following corollary is an immediate consequence of Proposition 1.

Corollary 3. *Where* $\mathsf{sem} \in \{\mathsf{cf}, \mathsf{adm}, \mathsf{comp}, \mathsf{grou}, \mathsf{pref}, \mathsf{stab}\}$ *and* $\mathtt{x} \in \{\mathtt{f}, \mathtt{d}\}$, $\Delta \in$ x-sem*(ABF) *iff* $\Delta \in$ x-sem(ABF).

With the help of Propositions 2 and 3 we now prove the same for r-based semantics.

Theorem 14. *Where* $\mathsf{sem} \in \{\mathsf{cf}, \mathsf{adm}, \mathsf{comp}, \mathsf{grou}, \mathsf{pref}, \mathsf{stab}\}$, $\Delta \in$ r-sem*(ABF) *iff* $\Delta \in$ r-sem(ABF).

Proof. [r-cf] Suppose $\Delta \subseteq Ab$ and $\Delta', \Delta'' \subseteq \Delta$. If Δ' r-defeats* Δ'' then by Proposition 3, Δ' r-defeats Δ'' or Δ'' r-defeats Δ'. If Δ' r-defeats Δ'', by Proposition 2, Δ' r-defeats* Δ''.

[r-adm] Suppose $\Delta \in$ r$-$adm*(ABF). By our previous item, $\Delta \in$ r$-$cf(ABF). Suppose now that $\Theta \subseteq Ab$ r-defeats Δ. By Proposition 2, it r-defeats* Δ. Thus, Δ r-defeats* Θ. Assume for a contradiction that Δ does not r-defeat Θ. By Proposition 3, there is a $\Theta' \subset \Theta$ that d-defeats Δ while Δ does not r-defeat* Θ'. Since by Proposition 1, Θ' also d-defeats* Δ this is a contradiction to Δ being r-admissible*. So Δ r-defeats Θ.

Suppose now that $\Delta \in$ r$-$adm(ABF). By our previous item $\Delta \in$ r$-$cf*(ABF). Suppose $\Theta \subseteq Ab$ r-defeats* Δ. If Θ also r-defeats Δ then Δ r-defeats Θ and thus by Proposition 2 Δ r-defeats* Θ. Else, by Proposition 3, there is a $\Theta' \subset \Theta$ that d-defeats (and thus also r-defeats) Δ and Δ does not r-defeat* Θ'. But then by Proposition 2 Δ also does not r-defeat Θ' which contradicts Δ being admissible.

[$\mathtt{r}$-comp] Suppose $\Delta \in \mathtt{r}-\mathsf{comp}(\mathtt{ABF})$. By our previous item, $\Delta \in \mathtt{r}-\mathsf{adm}^\star(\mathtt{ABF})$. Suppose Δ $\mathtt{r}$-defends* $\Theta \subseteq Ab$. Assume $\Lambda \subseteq Ab$ $\mathtt{r}$-defeats Θ. By Proposition 2, Λ $\mathtt{r}$-defeats* Θ. Hence, Δ $\mathtt{r}$-defeats* Λ. Assume for a contradiction that Δ does not $\mathtt{r}$-defeats Λ. By Proposition 3 there is a $\Lambda' \subset \Lambda$ that $\mathtt{r}$-defeats* Δ and which is not $\mathtt{r}$-defeated* by Δ. This contradicts $\Delta \in \mathtt{r}-\mathsf{adm}(\mathtt{ABF})$. Thus, Δ $\mathtt{r}$-defeats Λ. Hence, Θ is defended by Δ and so $\Theta \subseteq \Delta$.

Suppose now that $\Delta \in \mathtt{r}-\mathsf{comp}^\star(\mathtt{ABF})$. By our previous item, $\Delta \in \mathtt{r}-\mathsf{adm}(\mathtt{ABF})$. Suppose Δ $\mathtt{r}$-defends $\Theta \subseteq Ab$. Let $C \in \Theta$ be arbitrary. Suppose $\Lambda \subseteq Ab$ $\mathtt{r}$-defeats* C. By Proposition 3, either Λ also $\mathtt{r}$-defeats C or $\emptyset$ $\mathtt{d}$-defeats Δ and so by Proposition 1 $\emptyset$ also $\mathtt{d}$-defeats* Δ. The latter is impossible since $\Delta \in \mathtt{r}-\mathsf{adm}^\star(\mathtt{ABF})$. Hence, since Λ $\mathtt{r}$-defeats C, Δ $\mathtt{r}$-defeats Λ and so by Proposition 2 it also $\mathtt{r}$-defeats* Λ. Hence, Δ $\mathtt{r}$-defends* C and so $C \in \Delta$. Since C was arbitrary in Θ, $\Theta \subseteq \Delta$.

The claims for $\mathsf{sem} \in \{\mathsf{grou}, \mathsf{pref}\}$ follow with the previous item.

[$\mathtt{r}$-stab] Suppose $\Delta \in \mathtt{r}-\mathsf{stab}(\mathtt{ABF})$. By our first item, $\Delta \in \mathtt{r}-\mathsf{cf}^\star(\mathtt{ABF})$. Let $A \in Ab \setminus \Delta$. Thus, Δ $\mathtt{r}$-defeats A. By Proposition 2 it also $\mathtt{r}$-defeats* A. Thus, $\Delta \in \mathtt{r}-\mathsf{stab}^\star(\mathtt{ABF})$.

Suppose now that $\Delta \in \mathtt{r}-\mathsf{stab}^\star(\mathtt{ABF})$. By our first item, $\Delta \in \mathtt{r}-\mathsf{cf}(\mathtt{ABF})$. Let $A \in Ab \setminus \Delta$. Thus, Δ $\mathtt{r}$-defeats* A. Assume for a contradiction that Δ does not $\mathtt{r}$-defeat A. Then by Proposition 3, $\emptyset$ $\mathtt{d}$-defeats* Δ which contradicts that $\Delta \in \mathtt{r}-\mathsf{cf}^\star(\mathtt{ABF})$. Thus, $\Delta \in \mathtt{r}-\mathsf{stab}(\mathtt{ABF})$. □

Appendix B Proofs for Section 2

Fact 1. Where $\mathtt{x} \in \{\mathtt{f}, \mathtt{d}, \mathtt{r}\}$, $\Delta', \Theta \subseteq Ab$ and $\Delta \subseteq \Delta'$:

- if Δ $\mathtt{x}$-defeats Θ then Δ' $\mathtt{x}$-defeats Θ
- if Θ $\mathtt{x}$-defeats Δ then Θ $\mathtt{x}$-defeats Δ'.

Proof. For the case where $\mathtt{x} \in \{\mathtt{f}, \mathtt{d}\}$ both claims hold trivially. For item 1 and $\mathtt{x} = \mathtt{r}$, suppose that Δ $\mathtt{r}$-defeats Θ but it does not $\mathtt{d}$-defeat Θ (since if Θ would $\mathtt{d}$-defeat Θ, this case would reduce to the case $\mathtt{x} = \mathtt{d}$). This means that there is some $\Theta' \subseteq \Theta$ and an $A \in \Delta$ such that $\Theta' \vdash_{\mathcal{R}} \overline{A}$ and $\Theta' < A$. Since $A \in \Delta'$, Δ' also $\mathtt{r}$-defeats Θ. For item 2 and $\mathtt{x} = \mathtt{r}$ suppose Θ $\mathtt{x}$-defeats Δ but does not $\mathtt{d}$-defeat it. Thus, there is an $A \in \Theta$ and a $\Delta'' \subseteq \Delta$ such that $\Delta'' \vdash_{\mathcal{R}} \overline{A}$ and $\Delta'' < A$. Since $\Delta'' \subseteq \Delta'$, Θ also $\mathtt{r}$-defeats Δ'. □

Fact 2. If $\Delta \vdash_{\mathcal{R}} \overline{A}$ then either A $\mathtt{r}$-defeats Δ or Δ $\mathtt{r}$-defeats A.

Proof. Suppose that $\Delta \vdash_{\mathcal{R}} \overline{A}$. If $\Delta \not< A$ then Δ $\mathtt{r}$-defeats A. Otherwise A $\mathtt{r}$-defeats Δ. □

Appendix C Proofs for Section 3

C.1 Proof of Theorem 1

Fact 14. *Where $A \in Ab$, $\Delta \subseteq Ab$ and $\Delta' \subset \Delta$, if $\Delta \not< A$ then $\Delta' \not< A$.*

Proof. Suppose $\Delta \not< A$. Thus, there is no $B \in \min(\Delta)$ such that $B < A$. Thus, there is no $B \in \min(\Delta')$ such that $B < A$. □

Fact 15. *Where $\Delta \subseteq Ab$ and $A, B \in Ab$, for all $B \in \min(\Delta)$ for which $B < A$ and for all $\Theta \subseteq (\Delta \setminus \{B\}) \cup \{A\}$, $\Theta \not< B$.*

Proof. We show that $(\Delta \setminus \{B\}) \cup \{A\} \not< B$. The rest follows by Fact 14. Suppose there is a $B \in \min(\Delta)$ for which $B < A$. Suppose first that $\Delta \setminus \{B\} = \emptyset$. Then $(\Delta \setminus \{B\}) \cup \{A\} = \{A\}$ and since $\{A\} \not< B$ also $(\Delta \setminus \{B\}) \cup \{A\} \not< B$. Suppose now that $\Delta \setminus \{B\} \neq \emptyset$. Let $C \in \min(\Delta \setminus \{B\})$. Then $C \not< B$. In view of the totality of $\leq$ we have two cases: $C \leq A$ or $A < C$. In the first case $C \in \min((\Delta \setminus \{B\}) \cup \{A\})$ and thus $(\Delta \setminus \{B\}) \cup \{A\} \not< B$. In the second case $A \in \min((\Delta \setminus \{B\}) \cup \{A\})$ and thus $(\Delta \setminus \{B\}) \cup \{A\} \not< B$. □

Fact 16. *If $\Delta < A$ then there is a $B \in \min(\Delta)$ such that $\Theta \not< B$ for all $\Theta \subseteq (\Delta \setminus \{B\}) \cup \{A\}$.*

Proof. Suppose $\Delta < A$. Thus, there is a $B \in \min(\Delta)$ for which $B < A$. The rest follows with Fact 15. □

Lemma 3. *If $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$ is closed under contraposition then:*

$$\Delta \textit{ is } \mathtt{d}\textit{-conflict-free} \quad \textit{iff} \quad \Delta \textit{ is } \mathtt{f}\textit{-conflict-free.}$$

Proof. ($\Leftarrow$) This direction is trivial. ($\Rightarrow$) Suppose now for a contradiction that Δ is $\mathtt{d}$-conflict-free in $\mathtt{ABF}$ yet there is are $A \in \Delta$ and $\Delta' \subseteq \Delta$ for which $\Delta' \vdash_{\mathcal{R}} \overline{A}$. Since Δ is $\mathtt{d}$-conflict-free, this means $\Delta' < A$. Take $B \in \min(\Delta')$. By contraposition, there is a $\Delta^\star \subseteq \{A\} \cup (\Delta' \setminus \{B\})$ for which $\Delta^\star \vdash_{\mathcal{R}} \overline{B}$. By Fact 15, $\Delta^\star \not< B$. But then $\Delta^\star \subseteq \Delta$ $\mathtt{d}$-$<$-defeats $B \in \Delta$, in contradiction with Δ being $\mathtt{d}$-conflict-free. □

Theorem 1. Where $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$ is closed under contraposition. For any $\Delta \subseteq Ab$, if Δ is $\mathtt{d}$-conflict-free then there is no $A \in \Delta$, for which $\overline{A} \in \mathrm{Cn}_{\mathcal{R}}(\Delta)$.

Proof. Follows directly from Lemma 3. □

C.2 Proof of Theorem 3

Theorem 3. For any $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$, if $\Delta \in \texttt{d-adm}(\texttt{ABF})$ $\texttt{d}$-defends $A \in Ab$ then $\Delta \cup \{A\}$ is $\texttt{d}$-admissible.

Proof. We first show that Δ (and so in view of Fact 1 also $\Delta \cup \{A\}$) $\texttt{d}$-defends $\Delta \cup \{A\}$. Suppose that Θ $\texttt{d}$-defeats $\Delta \cup \{A\}$. If the attack is in Δ, Δ $\texttt{d}$-defeats Θ since Δ is $\texttt{d}$-admissible, while if the attack is in A, also Δ $\texttt{d}$-defeats Θ since it $\texttt{d}$-defends A.

Now we show that $\Delta \cup \{A\}$ is $\texttt{d}$-conflict-free. Assume for a contradiction that some $\Theta \subseteq \Delta \cup \{A\}$ $\texttt{d}$-defeats $\Delta \cup \{A\}$. Thus, as shown in the first paragraph, Δ $\texttt{d}$-defeats Θ in some C. Since $C \in \Delta \cup \{A\}$ and Δ $\texttt{d}$-defends $\Delta \cup \{A\}$, Δ $\texttt{d}$-defeats Δ. This is a contradiction to the $\texttt{d}$-conflict-freeness of Δ. Altogether, we have shown that $\Delta \cup \{A\}$ is $\texttt{d}$-admissible. $\square$

C.3 Proof of Corollary 1

The following two facts follow immediately from Fact 1.

Fact 17. *If Δ $\texttt{r}$-defends Θ then Δ $\texttt{r}$-defends $\Theta' \subseteq \Theta$.*

Fact 18. *If Δ $\texttt{r}$-defends B then $\Delta \cup \Delta'$ $\texttt{r}$-defends B.*

Lemma 1. If $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ satisfies the Fundamental Lemma, then for all $\Delta, \Theta \subseteq Ab$, if Δ is $\texttt{r}$-admissible and it $\texttt{r}$-defends Θ, then $\Delta \cup \Theta$ is $\texttt{r}$-admissible as well.

Proof. Let $\Theta = \{B_1, \ldots, B_n\}$. We show inductively that $\Delta_i = \Delta \cup \{B_1, \ldots, B_i\} \in \texttt{r-adm}(\texttt{ABF})$. The base case follows from the Fundamental Lemma. Suppose now that $\Delta_i \in \texttt{r-adm}(\texttt{ABF})$. Since Δ $\texttt{r}$-defends B_{i+1}, also Δ_i $\texttt{r}$-defends B_{i+1} (by Facts 17 and 18). By the Fundamental Lemma, $\Delta_{i+1} \in \texttt{r-adm}(\texttt{ABF})$. $\square$

C.3.1 Iteratively Constructing the Grounded Extension

As noted above, one of the assurances the Fundamental Lemma gives is the fact that the grounded extension can be build up in an iterative way.

Definition 21. *Given $\texttt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ and $\texttt{x} \in \{\texttt{d}, \texttt{r}\}$ we define:*

- $\texttt{x}\text{-}\mathsf{grou}_0(\texttt{ABF})$ *is the union of all sets $\Theta \subseteq Ab$ that are not $\texttt{x}$-defeated by any $\Delta \subseteq Ab$.*
- $\texttt{x}\text{-}\mathsf{grou}_{i+1}(\texttt{ABF})$ *is the union of all sets $\texttt{x}$-defended by $\texttt{x}\text{-}\mathsf{grou}_i(\texttt{ABF})$.*

Remark 7. *Notice that in Definition 7,* $\mathtt{x}$-grou *is defined as the* set *of all* $\subseteq$*-minimally* $\mathtt{x}$*-complete sets of assumptions, i.e. it is a set of sets. If there is a unique grounded extension, it is sensible to consider the set of assumptions* $\bigcup \mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF})$ *instead of the set containing a single set of assumptions* $\mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF})$*.*

Theorem 15. *For any* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$*,*

1. *if* $\mathtt{ABF}$ *is closed under contraposition and where* $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$*,* $\bigcup \mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF}) = \bigcup_{i\geq 1} \mathtt{x}\text{-}\mathsf{grou}_i(\mathtt{ABF}) = \bigcap \mathsf{comp}(\mathtt{ABF}) \in \mathsf{comp}(\mathtt{ABF})$*;*

2. $\bigcup \mathtt{d}\text{-}\mathsf{grou}(\mathtt{ABF}) = \bigcup_{i\geq 1} \mathtt{d}\text{-}\mathsf{grou}_i(\mathtt{ABF}) = \bigcap \mathsf{comp}(\mathtt{ABF}) \in \mathsf{comp}(\mathtt{ABF})$*.*

Proof. Consider the sequence $\langle F^n(\emptyset)\rangle_{n\geq 1} = \langle \mathtt{x}\text{-}\mathsf{grou}_i(\mathtt{ABF})\rangle_{i\geq 0}$ where $F(\Theta)$ denotes the set of all assumptions that are $\mathtt{x}$-defended by Θ. I.e. $F^1(\emptyset) = \mathtt{x}\text{-}\mathsf{grou}_0(\mathtt{ABF})$, $F^2(\emptyset) = \mathtt{x}\text{-}\mathsf{grou}_1(\mathtt{ABF})$, and so on. By the Fundamental Lemma (Theorem 3 and Corollary 1) this is a $\subseteq$-monotonic sequence of admissible subsets of Ab. As such there is a fixed-point which is identical to $\bigcup_{i\geq 0} \mathtt{x}\text{-}\mathsf{grou}_i(\mathtt{ABF})$. (Note that the fixed-point is reached after finitely many iterations since Ab is finite.) Notice that it holds that $\bigcup_{i\geq 0} \mathtt{x}\text{-}\mathsf{grou}_i(\mathtt{ABF}) \in \mathsf{comp}(\mathtt{ABF})$ since it is admissible and contains all assumptions it defends.

Let $\Delta \in \mathsf{comp}(\mathtt{ABF})$. We show that each $\mathtt{x}\text{-}\mathsf{grou}_i(\mathtt{ABF}) \subseteq \Delta$. For $i = 0$ this holds since $F(\emptyset)$ is defended by $\emptyset$ and so by Δ in view of Fact 1. Consider $i+1$. Then $F^i(\emptyset)$ defends every $A \in F^{i+1}(\emptyset) \setminus F^i(\emptyset)$ and since by the inductive hypothesis $F^i(\emptyset) \subseteq \Delta$ also Δ defends every such A in view of Fact 1. Thus, $F^{i+1}(\emptyset) \subseteq \Delta$.

Hence since $\bigcup_{i\geq 0} \mathtt{x}\text{-}\mathsf{grou}_i(\mathtt{ABF}) \in \mathsf{comp}(\mathtt{ABF})$ and $\bigcup_{i\geq 0} \mathtt{x}\text{-}\mathsf{grou}_i(\mathtt{ABF}) \subseteq \Delta$ for every $\Delta \in \mathsf{comp}(\mathtt{ABF})$, $\bigcup_{i\geq 0} \mathtt{x}\text{-}\mathsf{grou}_i(\mathtt{ABF}) = \bigcap \mathsf{comp}(\mathtt{ABF})$. □

Appendix D Proofs for Section 6

In the following we suppose that $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$ is well-behaved (Definition 12), that $\mathbb{V}$ is a finite initial sequence of $\mathbb{N}$, $\leq$ the canonical order on $\mathbb{N}$ and $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$.

For readability we recall Definition 17:

Definition 17. Where $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$,

- $\mathsf{IS}(\mathtt{ABF})$ is the set of all $\Delta \subseteq Ab$ such that $\Delta \setminus \{A\} \vdash_{\mathcal{R}} \overline{A}$ for some $A \in \Delta$;
- $\mathsf{CS}(\mathtt{ABF})$ is the set of all $\Delta \subseteq Ab$ such that for no $\Theta \in \mathsf{IS}(\mathtt{ABF})$, $\Theta \subseteq \Delta$;
- $\mathsf{MCS}(\mathtt{ABF})$ is the set of all $\Delta \in \mathsf{CS}(\mathtt{ABF})$ that are $\subseteq$-maximal;

- Where $\Delta \subseteq Ab$ and $i \in \mathbb{N}$, $\pi_i(\Delta) = \{A \in \Delta \mid v(A) = i\}$;
- $\prec \subseteq \wp(Ab) \times \wp(Ab)$ is defined as: $\Delta \prec \Theta$ iff there is an $i \geq 1$ such that $\pi_j(\Delta) = \pi_j(\Theta)$ for every $j > i$ and $\pi_i(\Delta) \subset \pi_i(\Theta)$;
- $\mathsf{MCS}_{\prec}(\mathtt{ABF}) = \max_{\prec}(\mathsf{MCS}(\mathtt{ABF}))$.

In the following we use $\Delta \preceq \Theta$ to denote that $\Delta \prec \Theta$ or $\Delta = \Theta$.

The proof of the following fact is straight-forward and similar to the proof of Fact 24 below.

Fact 19. *$\prec$ is transitive.*

Fact 20. *Where $\Delta, \Theta \in \mathsf{CS}(\mathtt{ABF})$ and $\Delta \subset \Theta$, $\Delta \prec \Theta$.*

Proof. Suppose $\Delta \subset \Theta$. Since Ab is finite there is an $l \in \mathbb{N}$ for which $\pi_l(\Theta) \supset \pi_l(\Delta)$ and $\pi_k(\Theta) = \pi_k(\Delta)$ for all $k > l$. □

Lemma 4. *Where $\Delta \in \mathsf{CS}(\mathtt{ABF})$ there is a $\Theta \in \mathsf{MCS}(\mathtt{ABF})$ such that $\Delta \subseteq \Theta$.*

Proof. This follows immediately in view of the finiteness of Ab. One can construct the maximal consistent superset of Δ via the usual Lindenbaum construction which we now illustrate. Let $A_1, A_2, \ldots$ be an enumeration of Ab. We construct $\Theta = \bigcup_{i \geq 1} \Delta_i$ as follows:

- $\Delta_0 = \Delta$
- $\Delta_{i+1} = \begin{cases} \Delta_i \cup \{A_{i+1}\} & \text{if } \Delta_i \cup \{A_{i+1}\} \in \mathsf{CS}(\mathtt{ABF}) \\ \Delta_i & \text{else} \end{cases}$

By the construction, for each i, $\Delta_i \in \mathsf{CS}(\mathtt{ABF})$. Suppose $\Theta \notin \mathsf{CS}(\mathtt{ABF})$. Thus, there is a $\Theta' \subseteq \Theta$ for which $\Theta' \in \mathsf{IS}(\mathtt{ABF})$. Thus, there is a minimal j for which $\Theta' \subseteq \Delta_j$. Thus, $\Delta_j \notin \mathsf{CS}(\mathtt{ABF})$ which is a contradiction. □

Fact 21. *Where $\Theta \in \mathsf{MCS}(\mathtt{ABF})$ there is a $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$ such that $\Delta = \Theta$ or $\Theta \prec \Delta$.*

Proof. This follows immediately with the finiteness of Ab. □

Fact 22. *Where $\mathtt{x} \in \{\mathtt{d}, \mathtt{x}\}$, if $\Delta \in \mathsf{IS}(\mathtt{ABF})$ then $\Delta \nsubseteq \Theta$ for all $\mathtt{x}$-conflict-free $\Theta \subseteq Ab$.*

Proof. Suppose Θ is $\mathtt{x}$-conflict-free and $\Delta \in \mathsf{IS}(\mathtt{ABF})$. Assume for a contradiction that $\Delta \subseteq \Theta$. Thus, there is an $A \in \Delta$ for which $\Delta \setminus \{A\} \vdash_{\mathcal{R}} \overline{A}$. We have two cases: (a) $\Delta \setminus \{A\} < A$ or (b) $\Delta \setminus \{A\} \not< A$. In case (b), Δ $\mathtt{d}$-defeats A which contradicts the $\mathtt{x}$-conflict-freeness of Θ. Suppose case (a). If $\mathtt{x} = \mathtt{r}$, A $\mathtt{r}$-defeats $\Delta \setminus \{A\}$ which contradicts the $\mathtt{r}$-conflict-freeness of Δ. Otherwise (i.e. $\mathtt{x} = \mathtt{d}$), by contraposition, for an arbitrary $B \in \min(\Delta)$ there is a $\Delta' \subseteq \Delta \setminus \{B\}$ such that $\Delta' \vdash_{\mathcal{R}} \overline{B}$. Moreover, by Fact 15, $\Delta' \not< B$ and thus Δ' $\mathtt{d}$-defeats B which contradicts the $\mathtt{d}$-conflict-freeness of Δ. □

Fact 23. *Where* $\mathtt{x} \in \{\mathtt{d}, \mathtt{x}\}$, *if* $\Theta \subseteq Ab$ *is not* $\mathtt{x}$*-conflict-free, there is a* $\Delta \in \mathsf{IS}(\mathtt{ABF})$ *for which* $\Delta \subseteq \Theta$.

Proof. Suppose Θ is not $\mathtt{x}$-conflict-free. Thus, there is a $\subseteq$-minimal $\Theta' \subseteq \Theta$ and an $A \in \Theta$ for which $\Theta' \vdash_{\mathcal{R}} \overline{A}$. Since $\mathtt{ABF}$ is sane, $A \notin \Theta'$ and consequently $(\Theta' \cup \{A\}) \setminus \{A\} \vdash_{\mathcal{R}} \overline{A}$, i.e. , $\Theta' \cup \{A\} \in \mathsf{IS}(\mathtt{ABF})$. □

Lemma 5. $\mathtt{d}\text{-}\mathsf{stab}(\mathtt{ABF}) \subseteq \mathsf{MCS}_{\prec}(\mathtt{ABF})$.

Proof. Suppose $\Delta \in \mathtt{d}\text{-}\mathsf{stab}(\mathtt{ABF})$. By Fact 22, there is no $\Theta \in \mathsf{IS}(\mathtt{ABF})$ such that $\Theta \subseteq \Delta$. Thus $\Delta \in \mathsf{CS}(\mathtt{ABF})$.

By Lemma 4 there is a $\Theta \in \mathsf{MCS}(\mathtt{ABF})$ for which $\Delta \subseteq \Theta$. Assume $\Delta \subset \Theta$ and thus there is a $B \in \Theta \setminus \Delta$. Since Δ is stable, Δ $\mathtt{d}$-defeats B. By sanity, $\Delta' \cup \{B\} \in \mathsf{IS}(\mathtt{ABF})$ for some $\Delta' \subseteq \Delta$ which contradicts $\Theta \in \mathsf{CS}(\mathtt{ABF})$. Thus, $\Delta \in \mathsf{MCS}(\mathtt{ABF})$.

Now suppose there is a $\Theta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$ for which $\Delta \prec \Theta$. Thus, there is an $i \geq 1$ such that for all $j > i$, $\pi_j(\Theta) = \pi_j(\Delta)$ and $\pi_i(\Theta) \supset \pi_i(\Delta)$. Let $B \in \pi_i(\Theta) \setminus \pi_i(\Delta)$. Since Δ is $\mathtt{d}$-stable, Δ $\mathtt{d}$-defeats B. Thus, there is a $\Delta' \subseteq \Delta$ such that $\Delta' \vdash_{\mathcal{R}} \overline{B}$, $\Delta' \not< B$ and, by sanity there is a $\Delta'' \subseteq \Delta' \setminus \{B\}$ for which $\Delta'' \vdash_{\mathcal{R}} \overline{B}$ and hence $\Delta'' \cup \{B\} \in \mathsf{IS}(\mathtt{ABF})$. By Fact 14, $\Delta'' \not< B$. Hence, for all $C \in \Delta''$, $C \geq B$. Thus, $\Delta'' \subseteq \Theta$. We reached a contradiction to $\Theta \in \mathsf{CS}(\mathtt{ABF})$. □

Lemma 6. *Where* $\mathtt{x} \in \{\mathtt{d}, \mathtt{x}\}$, $\mathsf{MCS}_{\prec}(\mathtt{ABF}) \subseteq \mathtt{x}\text{-}\mathsf{stab}(\mathtt{ABF})$.

Proof. Suppose $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$ and $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$.

Conflict-freeness. Suppose Δ is not $\mathtt{x}$-conflict-free. By Fact 23 there is a $\Delta' \subseteq \Delta$ for which $\Delta' \in \mathsf{IS}(\mathtt{ABF})$ in contradiction to $\Delta \in \mathsf{CS}(\mathtt{ABF})$.

Stability. Assume for a contradiction that there is an $A \in Ab \setminus \Delta$ such that Δ does not $\mathtt{x}$-defeat A. We know that $\Delta \cup \{A\} \notin \mathsf{CS}(\mathtt{ABF})$ is not consistent since otherwise $\Delta \notin \mathsf{MCS}(\mathtt{ABF})$.

Let $\Delta_1, \Delta_2, \ldots$ be a list of all subsets of Δ for which $\Delta_i \cup \{A\} \in \mathsf{IS}(\mathtt{ABF})$. Since Δ does not $\mathtt{x}$-defeat A, for all these Δ_i, $\min(\Delta_i) < A$. To see this note that for each i, $\Delta_i \notin \mathsf{IS}(\mathtt{ABF})$ and hence there is $B_i \in \Delta_i \cup \{A\}$ for which $(\Delta_i \cup \{A\}) \setminus \{B_i\} \vdash \overline{B_i}$. By contraposition, there is a $\Delta'_i \subseteq \Delta_i$ for which $\Delta'_i \vdash_{\mathcal{R}} \overline{A}$. Since Δ_i does not $\mathtt{x}$-defeat A, $\Delta'_i < A$ and so $\Delta_i < A$ by Fact 14.

Let Λ be a set that contains at least one member B_i of each $\min(\Delta_i)$. Let $\Theta = (\Delta \setminus \Lambda) \cup \{A\}$. Then $\Delta \prec \Theta$. To see this, notice that $v(B) < v(A)$ for every $B \in \Lambda$ and thus (†) $\pi_i(\Delta) = \pi_i(\Theta)$ for every $i > v(A)$. Since furthermore $\pi_{v(A)}(\Theta) = \pi_{v(A)}(\Delta) \cup \{A\}$ and $A \notin \Delta$, we see that $\pi_{v(A)}(\Theta) \supset \pi_{v(A)}(\Delta)$.

We now assume for a contradiction that $\Theta \notin \mathsf{CS}(\mathtt{ABF})$. Thus, there is a $\Omega \subseteq \Theta$ such that $\Omega \in \mathsf{IS}(\mathtt{ABF})$. Since $\Delta \in \mathsf{CS}(\mathtt{ABF})$, $\Omega \not\subseteq \Delta$ and hence $A \in \Omega$. So, $\Omega \setminus \{A\} = \Delta_i$ for some i. But this is impossible since $B_i \notin \Theta$. Thus, $\Theta \in \mathsf{CS}(\mathtt{ABF})$.

By Lemma 4, there is a $\Theta' \supseteq \Theta$ such that $\Theta' \in \mathsf{MCS}(\mathtt{ABF})$. Since, by Facts 19 and 20, $\Delta \prec \Theta'$, we have reached a contradiction to our main assumption. Thus, $\Delta \in \mathtt{x}\text{-}\mathsf{stab}(\mathtt{ABF})$. □

Lemma 7. $\mathtt{d}\text{-}\mathsf{stab}(\mathtt{ABF}) = \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF})$.

Proof. Suppose first that $\Delta \in \mathtt{d}\text{-}\mathsf{stab}(\mathtt{ABF})$. By Lemma 5, $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$. By Lemma 6, this means that $\Delta \in \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF})$.

Suppose now that $\Delta \in \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF})$. Assume that $A \in Ab \setminus \Delta$. We know that Δ $\mathtt{r}$-defeats A. Thus, either Δ $\mathtt{d}$-defeats A or there is a $B \in \Delta$ for which $A \vdash_{\mathcal{R}} \overline{B}$ and $A < B$. In the latter case, by contraposition, $\{B\} \vdash_{\mathcal{R}} \overline{A}$ or $\emptyset \vdash_{\mathcal{R}} \overline{A}$ and consequently, Δ $\mathtt{d}$-defeats A. Thus, $\Delta \in \mathtt{d}\text{-}\mathsf{stab}(\mathtt{ABF})$. □

In the next proof we will use the following order $\sqsubset \subseteq \wp(Ab) \times \wp(Ab)$:

Definition 22. $\Delta \sqsubset \Theta$ *iff there is an* $i \geq 1$ *such that* $\pi_i(\Delta) \subset \pi_i(\Theta)$ *and for all* $1 \leq j < i$, $\pi_j(\Delta) = \pi_j(\Theta)$.

Fact 24. *(i)* $\sqsubset$ *is transitive, (ii)* $\sqsubset$ *is* $\subset$*-monotonic.*

Proof. Ad (i). Suppose $\Delta_1 \sqsubset \Delta_2$ and $\Delta_2 \sqsubset \Delta_3$. Thus,

1. there is an i_1 for which (i) $\pi_{i_1}(\Delta_1) \subset \pi_{i_1}(\Delta_2)$ and (ii) for all $1 \leq j < i_1$, $\pi_j(\Delta_1) = \pi_j(\Delta_2)$,
2. there is an i_2 for which (i) $\pi_{i_2}(\Delta_2) \subset \pi_{i_2}(\Delta_3)$ and (ii) for all $1 \leq j < i_2$, $\pi_j(\Delta_2) = \pi_j(\Delta_3)$.

We distinguish two cases: (a) $i_1 < i_2$ and (b) $i_1 \geq i_2$.

Ad (a). In view of 1.i and 2.ii, $\pi_{i_1}(\Delta_1) \subset \pi_{i_1}(\Delta_2) = \pi_{i_1}(\Delta_3)$ and in view of 1.ii and 2.ii, for all $1 \leq j < i_1$, $\pi_{i_1}(\Delta_1) = \pi_{i_1}(\Delta_2) = \pi_{i_1}(\Delta_3)$. Hence, $\Delta_1 \sqsubset \Delta_3$.

Ad (b). In view of 1 and 2.i, $\pi_{i_2}(\Delta_1) \subseteq \pi_{i_2}(\Delta_2) \subset \pi_{i_2}(\Delta_3)$ and, in view of 1.ii and 2.ii, for all $1 \leq j < i_2$, $\pi_{i_2}(\Delta_1) = \pi_{i_2}(\Delta_2) = \pi_{i_3}(\Delta_3)$. Hence, $\Delta_1 \sqsubset \Delta_3$. □

Fact 25. *If $\Delta \subseteq Ab$ and $M \in \min(\Delta)$, $C \in Ab$, $C > M$, and $\Delta' \subseteq (\Delta \setminus \{M\}) \cup \{C\}$, then $\Delta' \sqsubset \Delta$.*

Proof. Where $i = v(M)$, since $M < C$ and $\Delta' \subseteq (\Delta \setminus \{M\}) \cup \{C\}$, we have $\pi_i(\Delta) \subseteq \pi_i((\Delta \setminus \{M\}) \cup \{C\}) \subset \pi_i(\Delta)$ and for all $1 \leq j < i$, $\pi_j(\Delta') \subseteq \pi_j(\Delta)$. Clearly, then $\Delta' \sqsubset \Delta$. $\square$

Lemma 8. $\mathtt{x}\text{-}\mathsf{pref}(\mathtt{ABF}) \subseteq \mathsf{MCS}_{\prec}(\mathtt{ABF})$

Proof. Suppose $\Delta \in \mathtt{x}\text{-}\mathsf{pref}(\mathtt{ABF})$. If $\Delta \notin \mathsf{CS}(\mathtt{ABF})$ then there is a $\Theta \in \mathsf{IS}(\mathtt{ABF})$ for which $\Theta \subseteq \Delta$ and hence by Fact 22, $\Delta \notin \mathtt{x}\text{-}\mathsf{adm}(\mathtt{ABF})$ which is a contradiction. Thus, $\Delta \in \mathsf{CS}(\mathtt{ABF})$. By Lemma 4, there is a $\Theta' \in \mathsf{MCS}(\mathtt{ABF})$ for which $\Delta \subseteq \Theta'$.

By Fact 20, $\Delta \preceq \Theta'$. Since Ab is finite, there is a $\Theta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$ such that $\Theta' \preceq \Theta$. Assume for a contradiction that $\Delta \prec \Theta$. Thus, there is an $i \geq 0$ for which $\pi_i(\Delta) \subset \pi_i(\Theta)$ and $\pi_j(\Delta) = \pi_j(\Theta)$ for all $j > i$. Since Δ is $\mathtt{x}$-preferred, $\Delta \cup \pi_i(\Theta)$ is not $\mathtt{x}$-admissible. We now show that in fact $\Omega = \Delta \cup \pi_i(\Theta)$ is $\mathtt{x}$-admissible which is a contradiction and thus our assumption is false. Thus, $\Delta \not\prec \Theta$ which implies that $\Delta = \Theta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$.

We first show that Ω is $\mathtt{x}$-conflict-free. Assume the opposite. Then there are $\Delta' \subseteq \Omega$ and $C \in \Omega$ for which $\Delta' \vdash_{\mathcal{R}} \overline{C}$. (†) Without loss of generality we assume that Δ' is $\sqsubset$-minimal with this property, i.e., we let Δ' and C be such that there are no $\Delta'' \subseteq \Omega$ and $C' \in \Omega$ for which $\Delta'' \vdash_{\mathcal{R}} \overline{C'}$ and $\Delta'' \sqsubset \Delta'$. We have two cases: (1) $\Delta' \geq C$ or (2) $\Delta' < C$.

> Suppose (2) and let $M \in \min(\Delta')$. By contraposition, there is a $\Delta'' \subseteq (\Delta' \setminus \{M\}) \cup \{C\}$ for which $\Delta'' \vdash_{\mathcal{R}} \overline{M}$. By Fact 25, $\Delta'' \sqsubset \Delta'$ which contradicts (†).
>
> Suppose (1). We distinguish two cases: (a) $C \notin \Delta$ and (b) $C \in \Delta$.
>
>> In case (a), $C \in \pi_i(\Theta) \setminus \Delta$ and since $\Delta' \geq C$, $\Delta' \subseteq \bigcup_{j \geq i} \pi_j(\Delta) \cup \pi_i(\Theta) = \bigcup_{j \geq i} \pi_j(\Theta) \subseteq \Theta$. But since then $\Delta' \subseteq \Theta$ $\mathtt{d}$-defeats $C \in \Theta$ this contradicts the $\mathtt{x}$-conflict-freeness of Θ (note that by Fact 23 Θ is $\mathtt{x}$-conflict-free since $\Theta \in \mathsf{CS}(\mathtt{ABF})$).
>>
>> Suppose (b). Since Δ is $\mathtt{x}$-admissible and Δ' $\mathtt{d}$-defeats $C \in \Delta$, Δ $\mathtt{x}$-defeats Δ'. We have two cases: (i) Δ $\mathtt{d}$-defeats Δ' and (ii) not (i) and Δ $\mathtt{r}$-defeats Δ'.
>>
>>> Suppose (i). Thus, since Δ is $\mathtt{x}$-conflict-free, there are $\Lambda \subseteq \Delta$ and a $E \in \Delta' \setminus \Delta$ for which $\Lambda \vdash_{\mathcal{R}} \overline{E}$ and $\Lambda \geq E$. Thus, $E \in \pi_i(\Theta) \setminus \Delta$ and since $\Lambda \geq E$, $\Lambda \subseteq \bigcup_{j \geq i} \pi_j(\Delta) \cup \pi_i(\Theta) = \bigcup_{j \geq i} \pi_j(\Theta) \subseteq \Theta$. Again, this contradicts the $\mathtt{x}$-conflict-freeness of Θ.

Suppose (ii). Thus, there is a $T \in \Delta$ and a $\Delta'' \subseteq \Delta'$ for which $\Delta'' \vdash_{\mathcal{R}} \overline{T}$ and $\Delta'' < T$. Let $M \in \min(\Delta'')$. By contraposition, there is a $\Delta^\star \subseteq (\Delta'' \setminus \{M\}) \cup \{T\}$ for which $\Delta^\star \vdash_{\mathcal{R}} \overline{M}$. Since by Fact 25, $\Delta^\star \sqsubset \Delta'$, this is a contradiction to (†).

Altogether this shows that Ω is $\mathtt{x}$-conflict-free.

In order to show that Ω is $\mathtt{x}$-admissible, it remains to be shown that Θ defends itself against all attacks. We assume for a contradiction Λ $\mathtt{x}$-defeats Ω and ($\star$) Ω does not $\mathtt{x}$-defeat Λ. Since Δ is $\mathtt{x}$-admissible, we have two cases:

1. Λ $\mathtt{d}$-defeats some $C \in \pi_i(\Theta) \setminus \Delta$, or
2. some $C \in \Lambda$ $\mathtt{r}$-defeats some $\Delta' \subseteq \Omega$ where $\Delta' \setminus \Delta \neq \emptyset$ and so $\Delta' \vdash_{\mathcal{R}} \overline{C}$ and $\Delta' < C$.

In the remainder we show that both cases lead to a contradiction.

Suppose case 1. Then, $\Lambda \geq C$. Since by Lemma 6, Θ is $\mathtt{x}$-admissible, Θ $\mathtt{x}$-defeats Λ. Suppose first that Θ $\mathtt{d}$-defeats Λ in some $L \in \Lambda$. Thus, there is a $\Theta' \subseteq \Theta$ for which $\Theta' \vdash_{\mathcal{R}} \overline{L}$ and $\Theta' \geq L$. Note that $v(L) \geq i$ as $\Lambda \geq C$ and $v(C) = i$. So $\Theta' \subseteq \Omega$ and hence Ω $\mathtt{d}$-defeats Λ – a contradiction to ($\star$). Suppose now that some $T \in \Theta$ $\mathtt{r}$-defeats $\Lambda' \subseteq \Lambda$. So $\Lambda' \vdash_{\mathcal{R}} \overline{T}$ and $\Lambda' < T$. Thus, $v(T) > i$ and so $T \in \Omega$. So Ω $\mathtt{r}$-defeats Λ – again a contradiction to ($\star$).

Suppose case 2. Without loss of generality, we suppose that Δ' is $\sqsubset$-minimal in Ω with the property of being $\mathtt{r}$-defeated by C. Let $M \in \min(\Delta')$. So $M < C$. We distinguish two cases: (a) $v(M) \geq i$ and (b) $v(M) < i$.

Case a. If $v(M) \geq i$, $\Delta' \subseteq \bigcup_{j \geq i} \pi_j(\Delta) \cup \pi_i(\Theta) \subseteq \Theta \cap \Omega$. Since by Lemma 6, Θ is $\mathtt{x}$-admissible, Θ $\mathtt{x}$-defeats C. So, there is a $\Theta' \subseteq \Theta$ for which $\Theta' \vdash_{\mathcal{R}} \overline{C}$ and $\Theta' \not< C$. Since $v(C) > i$, $\Theta' \subseteq \Omega$ and so Ω defends itself against the attack by Λ which is a contradiction to ($\star$).

Case b. Suppose now that $v(M) < i$ and so $M \in \Delta$. Since $\Delta' \vdash_{\mathcal{R}} \overline{C}$, by contraposition there is a $\Delta^\star \subseteq (\Delta' \setminus \{M\}) \cup \{C\}$ for which $\Delta^\star \vdash_{\mathcal{R}} \overline{M}$. By Fact 25, $\Delta^\star \sqsubset \Delta'$. Also, since $\Delta^\star \not< M$, $\Delta^\star$ $\mathtt{x}$-defeats Δ in M. Thus, Δ $\mathtt{x}$-defeats $\Delta^\star$.

Suppose first that Δ $\mathtt{d}$-defeats $\Delta^\star$. Thus, there is a $\Delta^\dagger \subseteq \Delta$ for which $\Delta^\dagger \vdash_{\mathcal{R}} \overline{D}$ for some $D \in \Delta^\star$ and $\Delta^\dagger \geq D$. By the conflict-freeness of Ω, $D = C$. This is a contradiction to ($\star$).

> Suppose now that Δ $\mathtt{r}$-defeats $\Delta^\star$. Thus, there is a $D \in \Delta$ and a $\Delta^\dagger \subseteq \Delta^\star$ for which $\Delta^\dagger \vdash_{\mathcal{R}} \overline{D}$ and $\Delta^\dagger < D$. Again, by the conflict-freeness of Ω, $C \in \Delta^\dagger$. So, $M < D$. By contraposition, there is a $\Delta^\ddagger \subseteq (\Delta^\dagger \setminus \{C\}) \cup \{D\}$ for which $\Delta^\ddagger \vdash_{\mathcal{R}} \overline{C}$. Since $\Delta^\ddagger \subseteq \Omega$ and Ω does not defeat C, $\Delta^\ddagger < C$ and so $\Delta^\ddagger$ is $\mathtt{r}$-defeated by C. Note that by the $\sqsubset$-minimality of Δ' it cannot be the case that $\Delta^\dagger \sqsubset \Delta'$. We will contradict this this now to conclude the proof. Since $\Delta^\dagger \subseteq \Delta^\star \subseteq (\Delta' \setminus \{M\}) \cup \{C\}$, $\Delta^\dagger \setminus \{C\} \subseteq \Delta' \setminus \{M\}$. Since $\Delta^\ddagger \subseteq (\Delta^\dagger \setminus \{C\}) \cup \{D\}$, $\Delta^\ddagger \subseteq (\Delta' \setminus \{M\}) \cup \{D\}$. Since $D > M$ and $M \in \min(\Delta')$, by Fact 25, $\Delta^\ddagger \sqsubset \Delta'$. But this is a contradiction to the $\sqsubset$-minimality of Δ'.

Altogether this shows that Ω is $\mathtt{x}$-admissible which completes our proof by reductio. □

Lemma 9. $\mathtt{x}\text{-}\mathsf{pref}(\mathtt{ABF}) = \mathsf{MCS}_{\prec}(\mathtt{ABF})$

Proof. In Lemma 8 we have shown that $\mathtt{x}\text{-}\mathsf{pref}(\mathtt{ABF}) \subseteq \mathsf{MCS}_{\prec}(\mathtt{ABF})$.

We now show that $\mathtt{x}\text{-}\mathsf{pref}(\mathtt{ABF}) \supseteq \mathsf{MCS}_{\prec}(\mathtt{ABF})$. Suppose that $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$. By Lemma 6, $\Delta \in \mathtt{x}\text{-}\mathsf{stab}(\mathtt{ABF})$ and hence $\Delta \in \mathtt{x}\text{-}\mathsf{pref}(\mathtt{ABF})$. □

Theorem 13. For any well-behaved ABF we have:

$$\mathsf{MCS}_{\prec}(\mathtt{ABF}) = \mathtt{r}\text{-}\mathsf{pref}(\mathtt{ABF}) = \mathtt{d}\text{-}\mathsf{pref}(\mathtt{ABF}) = \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF}) = \mathtt{d}\text{-}\mathsf{stab}(\mathtt{ABF})$$

Proof. This follows by Lemma 5, 6, 7 and Lemma 9. □

Appendix E The Fundamental Lemma for Weak Contraposition

We will again use the relation $\sqsubset$ from Definition 22 in the following proofs. Recall also Definition 18 of Weak Contraposition.

Lemma 10 (Fundamental Lemma for Weak Contraposition and $\mathtt{r}$-defeat)**.** *If* ABF *is weakly contrapositive, Δ is $\mathtt{r}$-admissible and $\mathtt{r}$-defends $A \in Ab$, then $\Delta \cup \{A\}$ is $\mathtt{r}$-admissible.*

Proof. We show that if the $\mathtt{r}$-admissible Δ $\mathtt{r}$-defends $A \in Ab \setminus \Delta$ then $\Delta' = \Delta \cup \{A\}$ is $\mathtt{r}$-admissible. We proceed in two steps, first showing that Δ' is $\mathtt{r}$-conflict-free and then that Δ' $\mathtt{r}$-defends itself.

Conflict-freeness. Suppose Δ' is not $\mathtt{r}$-conflict-free. Thus, there are $\Lambda \subseteq \Delta'$ and $B \in \Delta'$ such that either (a) Λ $\mathtt{d}$-defeats B or (b) B $\mathtt{r}$-defeats Λ.

We first note that ($\star$), Δ does not d-defeat A since otherwise Δ also defeats Δ to defend A. This, however, contradicts the r-conflict-freeness of Δ.

Suppose now (b) and so $\Lambda \vdash \overline{B}$ where $\Lambda < B$. By weak contraposition, there is a $D \in \min(\Lambda)$ and a $\Lambda'' \subseteq (\Lambda \setminus \{D\}) \cup \{B\}$ for which $\Lambda'' \vdash \overline{D}$. Since in view of Fact 15 $\Lambda'' \not< D$, Λ'' d-defeats D and we are in case (a). We will show that case (a) is impossible which in turn shows that both (a) and (b) are impossible and thus implies that Δ' is r-conflict-free.

So, suppose (a) and thus $\Lambda \vdash \overline{B}$ where $\Lambda \geq B$. We know that $\Lambda \cup \{B\} \nsubseteq \Delta$ since Δ is r-conflict-free.

- Suppose $B = A$. Then some $\Omega \subseteq \Delta$ r-defeats Λ since Δ r-defends A.

- Suppose $B \neq A$. Then $B \in \Delta$ and hence some $\Omega \subseteq \Delta$ r-defeats Λ since Δ is r-admissible.

In sum, so far we have shown (†) that whenever (a) or (b) for some $\Lambda \cup \{B\} \subseteq \Delta'$, there is a $\Omega \subseteq \Delta$ that r-defeats Λ. Also, $A \in \Lambda$ by the r-conflict-freeness of Δ. So, suppose Ω r-defeats Λ. We now distinguish the two types of defeat.

- ($\star\star$) Assume first for a contradiction that Ω d-defeats Λ. Then there is a $L \in \Lambda$ for which $\Omega \vdash \overline{L}$ and $\Omega \geq L$. By the r-conflict-freeness of Λ, $L = A$. By ($\star$) this is impossible.

- So, Ω r-defeats Λ which means that there is a $O \in \Omega$ and a $\Lambda' \subseteq \Lambda$ for which $\Lambda' \vdash \overline{O}$ and $\Lambda' < O$. Since $O \in$ ÂtΔ and by the conflict-freeness of Δ, $A \in \Lambda'$.

  Without loss of generality we now assume that Λ' is $\sqsubset$-minimal with the property of being a subset of Δ' that is r-defeated by Δ. (Since *Ab* is finite such a set exists.)

  By weak contraposition, there is an $L' \in \min(\Lambda')$ and a $\Lambda'' \subseteq (\Lambda' \setminus \{L'\}) \cup \{O\}$ for which $\Lambda'' \vdash \overline{L'}$. So L' is d-defeated by Λ''. This is case (a).

  Thus, by (†) there is a $\Omega' \subseteq \Delta$ that r-defeats Λ''. As shown in ($\star\star$), we are not dealing with a d-defeat.

  Since by Fact 25, $\Lambda'' \sqsubset \Lambda'$, we have a contradiction with the $\sqsubset$-minimality of Λ'.

This completes our proof by reductio showing that Δ' is r-conflict-free.

Admissibility. Suppose now some $\Theta \subseteq Ab$ r-defeats Δ'. Consider first the case of a d-defeat. Thus there is a $B \in \Delta'$ and $\Theta \vdash \overline{B}$ while $\Theta \geq B$. Since Δ is r-admissible and it r-defends A, Δ r-defeats Θ.

Consider now that the attack is a proper $\mathtt{r}$-defeat and thus there is a $T \in \Theta$ and a $\Lambda \subseteq \Delta'$ for which $\Lambda \vdash \overline{T}$ and $\Lambda < T$. By weak contradiction there is an $L \in \min(\Lambda)$ and a $\Lambda' \subseteq (\Lambda \setminus \{L\}) \cup \{T\}$ for which $\Lambda' \vdash \overline{L}$. Thus, Λ' $\mathtt{d}$-defeats L. Hence, Δ' $\mathtt{r}$-defeats Λ'.

If it $\mathtt{d}$-defeats Λ', it $\mathtt{d}$-defeats T by the $\mathtt{r}$-conflict-freeness of Δ'. So consider the case where it $\mathtt{r}$-defeats Λ'.

So, we know that there is a $\Lambda' \subseteq \Delta' \cup \{T\}$ that is $\mathtt{r}$-defeated by Δ'. Without loss of generality we suppose that Λ' is a $\sqsubset$-minimal subset of $\Delta' \cup \{T\}$ that is $\mathtt{r}$-defeated by Δ'. (Again, since *Ab* is finite such a set exists.) We again show that Δ' $\mathtt{d}$-defeats T.

Note that there is a $K \in \Delta'$ such that $\Lambda' \vdash \overline{K}$ and $\Lambda' < K$. By weak contraposition there is a $L' \in \min(\Lambda')$ and a $\Lambda'' \subseteq (\Lambda' \setminus \{L'\}) \cup \{K\}$ for which $\Lambda'' \vdash \overline{L'}$. By Fact 15, $\Lambda'' \not< L'$.

- If $L' = T$, Δ' $\mathtt{d}$-defeats Θ in T via Λ''.
- Else, $L' \in \Delta'$ and hence Δ' $\mathtt{r}$-defeats Λ''. If it is a $\mathtt{d}$-defeat then the defeat is in T (due to the $\mathtt{r}$-conflict-freeness of Δ'). Else, by Fact 25, we have a contradiction to the $\sqsubset$-minimality of Λ' since $\Lambda'' \sqsubset \Lambda'$.

Altogether we have shown that Δ' $\mathtt{d}$-defeats Θ. $\square$

The next corollary follows in view of Lemma 10 and Lemma 1.

Corollary 4. *For any* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *that is closed under weak contraposition, if* $\Delta \in \mathtt{r}\text{-}\mathtt{adm}(\mathtt{ABF})$ $\mathtt{r}$*-defends* $\Theta \subseteq Ab$ *then* $\Delta \cup \Theta$ *is* $\mathtt{r}$*-admissible.*

Appendix F Proofs for Section 4

F.1 Proof of Theorem 5

Theorem 16. *Let* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ *be closed under weak contraposition. Then,*

(1) every $\mathtt{d}$*-conflict-free set is* $\mathtt{r}$*-conflict-free and vice versa;*

(2) every $\mathtt{d}$*-stable set is* $\mathtt{r}$*-stable and vice versa;*

(3) every $\mathtt{d}$*-admissible set is an* $\mathtt{r}$*-admissible set;*

(4) every $\mathtt{d}$*-complete set is a subset of an* $\mathtt{r}$*-complete set.*

Proof. Ad 1. Suppose that Δ is not $\mathtt{r}$-conflict-free. Then there are $\Delta' \subseteq \Delta$ and $A \in \Delta$ for which $\Delta' \vdash_{\mathcal{R}} \overline{A}$ and (a) $\Delta' < A$ or (b) $\Delta' \geq A$. Suppose (a). By weak contraposition there are $B \in \min(\Delta')$ and $\Delta'' \subseteq (\Delta' \setminus \{B\}) \cup \{A\}$ such that $\Delta'' \vdash_{\mathcal{R}} \overline{B}$. By Fact 15, $\Delta'' \not< B$ and so $\Delta'' \subseteq \Delta$ $\mathtt{d}$-defeats $B \in \Delta$. Also in case (b), $\Delta' \subseteq \Delta$ $\mathtt{d}$-defeats Δ. So, Δ is not $\mathtt{d}$-conflict-free.

Suppose Δ is $\mathtt{r}$-conflict-free. Therefore Δ does not $\mathtt{r}$-defeat any $\Delta' \subseteq \Delta$. Thus, Δ does not $\mathtt{d}$-defeat any $A \in \Delta$ and so Δ is $\mathtt{d}$-conflict-free.

Ad 2. Suppose first that Δ is a $\mathtt{d}$-stable set, i.e. Δ $\mathtt{d}$-defeats every $A \in Ab \setminus \Delta$. Then clearly also Δ $\mathtt{r}$-defeats every $A \in Ab \setminus \Delta$. Also, by Item 1 Δ is $\mathtt{r}$-conflict-free since it is $\mathtt{d}$-conflict-free. Thus, Δ is $\mathtt{r}$-stable.

Suppose now that Δ is an $\mathtt{r}$-stable set and suppose for a contradiction that Δ is not $\mathtt{d}$-stable. By Item 1, Δ is $\mathtt{d}$-conflict-free. Consequently there is an $A \in Ab \setminus \Delta$ such that Δ $\mathtt{r}$-defeats A but does not $\mathtt{d}$-defeat A. I.e., $\{A\} \vdash_{\mathcal{R}} \overline{C}$ for some $C \in \Delta$ such that $C > A$. By weak contraposition, $\{C\} \vdash_{\mathcal{R}} \overline{A}$ or $\emptyset \vdash_{\mathcal{R}} \overline{A}$. But then Δ does $\mathtt{d}$-defeat A which is a contradiction.

Ad 3. Suppose first that $\Delta' \subseteq Ab$ is $\mathtt{d}$-admissible and suppose that some $\Theta \subseteq Ab$ $\mathtt{r}$-defeats Δ'. If Θ also $\mathtt{d}$-defeats Δ', Δ' $\mathtt{d}$-defeats Θ since it is $\mathtt{d}$-admissible. This implies that Δ' also $\mathtt{r}$-defeats Θ. Suppose now that there is a $\Delta \subseteq \Delta'$ such that $\Delta \vdash_{\mathcal{R}} \overline{A}$ and $\Delta < A$ for some $A \in \Theta$. By weak contraposition there is a $B \in \min(\Delta)$ and a $\Lambda \subseteq (\Delta \setminus \{B\}) \cup \{A\}$ for which $\Lambda \vdash_{\mathcal{R}} \overline{B}$. Note that by Fact 15 $\Lambda \not< B$ and so Λ $\mathtt{d}$-defeats B. Since Δ' is $\mathtt{d}$-admissible, Δ' $\mathtt{d}$-defeats Λ. Since $\mathtt{d}$-admissibility implies $\mathtt{d}$-conflict-freeness, Δ' defeats Λ in A. Consequently, Δ' also $\mathtt{r}$-defeats Θ.

Ad 4. Suppose that Δ is $\mathtt{d}$-complete. By (3), Δ is $\mathtt{r}$-admissible. For any $\Theta \subseteq Ab$, let $F(\Theta) \subseteq Ab$ be the set of all assumptions $\mathtt{r}$-defended by Θ. Consider the sequence $\langle F^n(\Delta)\rangle_{0 \leq n}$ where $F^0(\Delta) = \Delta$. By Corollary 4, this is a $\subseteq$-monotonic sequence of admissible subsets of Ab. Since Ab is finite there is fixed point $F^n(\Delta) \supseteq \Delta$ which is $\mathtt{r}$-complete. $\square$

Theorem 5 is a direct consequence of Theorem 16.

Appendix G Proofs for Section 5

Definition 23. *Where* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ $\Delta \subseteq Ab$ *and* $B \in \mathcal{L}$, *let* $\Delta^{-B} = \Delta^{+B} = \Delta$ *if* $B \notin Ab$, *and* $\Delta^{-B} = \Delta \setminus \{B\}$ *and* $\Delta^{+B} = \Delta \cup \{B\}$ *else.*

Notice in particular that this means that $Ab^{-B} = Ab \setminus \{B\}$.

G.1 Proof of Facts 8 and 13

Fact 8. Where $\mathsf{sem} \in \{\mathsf{grou}, \mathsf{pref}, \mathsf{stab}\}$, $\mathtt{x} \in \{\mathtt{r}, \mathtt{d}\}$ and every $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$ is $\mathtt{f}$-consistent, if $\mathtt{ABF}$ satisfies RM (relative to $\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}}$) then it satisfies CM-Ab (relative to $\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}}$).

Proof. Where $B \in Ab$ and $A \in \mathcal{L}$, suppose that $\mathtt{ABF}$ satisfies RM. Suppose now that $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} A$ and $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} B$. This means that for every $\mathtt{x}$-sem-extension Δ, $\Delta \vdash_{\mathcal{R}} A$ and $\Delta \vdash_{\mathcal{R}} B$. Since, by the $\mathtt{f}$-consistency of Δ, $\Delta \nvdash_{\mathcal{R}} \overline{B}$ and so $\mathtt{ABF} \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} \overline{B}$. By RM, $\mathtt{ABF}^B \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} A$. ☐

Fact 13. *Where* $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \leq, v)$ *is closed under contraposition [resp. sane, resp. well-behaved], then also* $\mathtt{ABF}^B = (\mathcal{L}, \mathcal{R} \cup \{\to B\}, Ab^{-B}, \overline{}, \leq, v)$ *is closed under contraposition [resp. sane, resp. well-behaved].*

Proof. Suppose $\mathtt{ABF}$ is closed under contraposition. Suppose $\Delta \vdash_{\mathcal{R}\cup\{\to B\}} \overline{C}$ where $\Delta \subseteq \mathcal{L}$, $A, C \in Ab$ and $A \in \Delta$. Then either (a) $\Delta \vdash_{\mathcal{R}} \overline{C}$ or (b) $\Delta \cup \{B\} \vdash_{\mathcal{R}} \overline{C}$. In the former case there is a $\Theta \subseteq \{C\} \cup (\Delta \setminus \{A\})$ for which $\Theta \vdash_{\mathcal{R}} \overline{A}$ and in the second case there is a $\Theta' \subseteq \{C\} \cup ((\Delta \cup \{B\}) \setminus \{A\})$ for which $\Theta' \vdash_{\mathcal{R}} \overline{A}$ since $\mathtt{ABF}$ is closed under contraposition. Thus, in any case there is a $\Lambda \subseteq \{C\} \cup (\Delta \setminus \{A\})$ for which $\Lambda \vdash_{\mathcal{R}\cup\{\to B\}} \overline{A}$. Thus, $\mathtt{ABF}^B$ is closed under contraposition.

Suppose $\mathtt{ABF}$ is sane. Where $\Delta \subseteq \mathcal{L}$ and $A \in Ab$ assume that $\Delta \vdash_{\mathcal{R}\cup\{\to B\}} \overline{A}$. Thus, $\Delta \cup \{B\} \vdash_{\mathcal{R}} \overline{A}$. By the sanity of $\mathtt{ABF}$, $(\Delta \cup \{B\}) \setminus \{A\} \vdash_{\mathcal{R}} \overline{A}$ and hence $\Delta \setminus \{A\} \vdash_{\mathcal{R}\cup\{\to B\}} \overline{A}$. Thus, $\mathtt{ABF}^B$ is sane.

Suppose $\mathtt{ABF}$ is well-behaved. Thus, it is sane and closed under contraposition. By the previous two cases, also $\mathtt{ABF}^B$ is sane and closed under contraposition. Thus, $\mathtt{ABF}^B$ is well-behaved. ☐

G.2 Proof of Theorem 7

Fact 26. *If* $\Delta' \vdash_{\mathcal{R}} \overline{A}$ *and* $\Delta' \cup \{A\} \subseteq \Delta$ *then* Δ *is not* $\mathtt{r}$*-conflict-free.*

Proof. We have two cases: (1) $\Delta' < A$ and (2) $\Delta' \not< A$. In the first case A $\mathtt{r}$-defeats Δ'. In the second case Δ' $\mathtt{r}$-defeats A. Thus, Δ is not $\mathtt{r}$-conflict-free. ☐

In the remainder of this subsection we assume that $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ is closed under contraposition and we let $\mathtt{ABF}^B = (\mathcal{L}, \mathcal{R} \cup \{\to B\}, Ab^{-B}, \overline{}, \mathbb{V}, \leq, v)$.

Fact 27. *If* Δ $\mathtt{x}$*-defeats* $A \in Ab^{-B}$ *in* $\mathtt{ABF}$ *then also* Δ^{-B} $\mathtt{x}$*-defeats* A *in* $\mathtt{ABF}^B$*.*

Proof. Suppose Δ $\mathtt{d}$-defeats $A \in Ab^{-B}$ in $\mathtt{ABF}$. Thus, $\Delta \vdash_{\mathcal{R}} \overline{A}$ and $\Delta \not< A$. Thus, there is a $\Delta' \subseteq \Delta^{-B}$ for which $\Delta' \vdash_{\mathcal{R}\cup\{\to B\}} \overline{A}$. By Fact 14, $\Delta' \not< A$ and hence Δ^{-B} $\mathtt{d}$-defeats A in $\mathtt{ABF}^B$.

Suppose now Δ $\mathtt{r}$-defeats $A \in Ab^{-B}$ in $\mathtt{ABF}$ but does not $\mathtt{d}$-defeat it. Thus, there is a $C \in \Delta$ for which $A \vdash_{\mathcal{R}} \overline{C}$ and $A < C$. If $C \neq B$ then also $A \vdash_{\mathcal{R}\cup\{\to B\}} \overline{C}$ and thus also Δ $\mathtt{r}$-defeats A in $\mathtt{ABF}^B$. Now suppose $B = C$. By contraposition, $B \vdash_{\mathcal{R}} \overline{A}$ or $\emptyset \vdash_{\mathcal{R}} \overline{A}$ and thus $\emptyset \vdash_{\mathcal{R}\cup\{\to B\}} \overline{A}$. Hence, A is also defeated by Δ^{-B} in $\mathtt{ABF}^B$. ☐

Fact 28. *Where $\Delta \cup \Delta_B \subseteq Ab$ and $A, B \in \mathcal{L}$, if $\Delta_B \vdash_{\mathcal{R}} B$ and $\Delta^{-B} \vdash_{\mathcal{R}\cup\{\to B\}} A$ then $\Delta^{-B} \vdash_{\mathcal{R}} A$ or $\Delta^{-B} \cup \Delta_B \vdash_{\mathcal{R}} A$.*

Proof. If $\Delta^{-B} \vdash_{\mathcal{R}\cup\{\to B\}} A$ but $\Delta^{-B} \nvdash_{\mathcal{R}} A$ then the rule $\to B$ was used in the derivation D underlying $\Delta^{-B} \vdash_{\mathcal{R}\cup\{\to B\}} A$. A derivation of A from $\Delta^{-B} \cup \Delta_B$ with the rules $\mathcal{R}$ is then obtained by replacing the instance(s) of $\to B$ in D by the derivation underlying $\Delta_B \vdash_{\mathcal{R}} B$. ☐

Lemma 11 (CC-semantic, $\mathtt{r}$-stab)**.** *Where $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathtt{r}} B$, $\Delta \in \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF})$ implies $\Delta^{-B} \in \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF}^B)$.*

Proof. Let $\Delta \in \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF})$. We show that $\Delta^{-B} \in \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF}^B)$ by showing $\mathtt{r}$-conflict-freeness and $\mathtt{r}$-stability. Since $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathtt{r}} B$ there is a $\Delta_B \subseteq \Delta$ for which $\Delta_B \vdash_{\mathcal{R}} B$.

$\mathtt{r}$-*conflict-freeness.* Assume for a contradiction that Δ^{-B} is not $\mathtt{r}$-conflict-free in $\mathtt{ABF}^B$. Thus, there is a $\Delta' \subseteq \Delta^{-B}$ and a $A \in \Delta^{-B}$ for which $\Delta' \vdash_{\mathcal{R}\cup\{\to B\}} \overline{A}$. Thus, by Fact 26, if $\Delta' \vdash_{\mathcal{R}} \overline{A}$ we have reached a contradiction since Δ is $\mathtt{r}$-conflict-free in $\mathtt{ABF}$. By Fact 28, $\Delta' \cup \Delta_B \vdash_{\mathcal{R}} \overline{A}$. We have again reached a contradiction since Δ is then not $\mathtt{r}$-conflict-free in $\mathtt{ABF}$ by Fact 26.

$\mathtt{r}$-*stability.* Let $A \in Ab^{-B} \setminus \Delta^{-B}$. Thus, $A \in Ab \setminus \Delta$. Since $\Delta \in \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF})$, Δ $\mathtt{r}$-defeats A in $\mathtt{ABF}$. By Fact 27, Δ^{-B} $\mathtt{r}$-defeats A in $\mathtt{ABF}^B$. ☐

Theorem 7. Any $\mathtt{ABF}$ closed under contraposition satisfies Cautious Cut for $\mid\!\sim^{\mathsf{stab}}_{\mathtt{r}}$.

Proof. In view of Lemma 11 we only need to show CC-entailment for $\mathtt{x}\text{-}\mathsf{stab}$. Suppose $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathtt{r}} B$ and $\mathtt{ABF}^B \mid\!\sim^{\mathsf{stab}}_{\mathtt{r}} A$. Let $\Delta \in \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF})$. Thus, there is a $\Delta_B \subseteq \Delta$ such that $\Delta_B \vdash_{\mathcal{R}} B$. By Lemma 11, $\Delta^{-B} \in \mathtt{r}\text{-}\mathsf{stab}(\mathtt{ABF}^B)$. Thus, there is a $\Delta' \subseteq \Delta^{-B}$ for which $\Delta' \vdash_{\mathcal{R}\cup\{\to B\}} A$. By Fact 28, $\Delta' \vdash_{\mathcal{R}} A$ or $\Delta' \cup \Delta_B \vdash_{\mathcal{R}} A$. Since $\Delta' \cup \Delta_B \subseteq \Delta$ this suffices to show that $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathtt{r}} A$. ☐

G.3 Proof of Theorem 8

In this section, suppose some given $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ is closed under contraposition and $\mathtt{ABF} \mid\!\sim^{\mathsf{grou}}_{\mathsf{d}} B$. Thus, there is a $\Delta_B \subseteq \bigcap \mathsf{d\text{-}grou}(\mathtt{ABF})$ for which $\Delta_B \vdash_{\mathcal{R}} B$. Slightly abusing notation we will also write $\mathsf{d\text{-}grou}(\mathtt{ABF})$ instead of $\bigcap \mathsf{d\text{-}grou}(\mathtt{ABF})$. Recall that by Theorem 15 there is an inductive definition of $\mathsf{d\text{-}grou}(\mathtt{ABF})$. Finally, we let $\mathtt{ABF}^B = (\mathcal{L}, \mathcal{R} \cup \{\rightarrow B\}, Ab^{-B}, \overline{}, \mathbb{V}, \leq, v)$.

Fact 29. *Where $\Theta \cup \{A\} \subseteq Ab$, if $\Theta \not< A$ and $\Theta \cup \Delta < A$, then there is a $C \in (\min(\Delta) \cap \min(\Delta \cup \Theta)) \setminus \Theta$ for which $C < A$.*

Proof. Suppose $\Theta \not< A$ and $\Theta \cup \Delta < A$. Since $\Theta \not< A$, for all $C \in \min(\Theta)$, $C \not< A$. Since $\Theta \cup \Delta < A$, there is a $D \in \min(\Theta \cup \Delta)$ for which $D < A$. Thus, $D \in (\min(\Delta) \cap \min(\Delta \cap \Theta)) \setminus \Theta$. ☐

Fact 30. *Where $\Delta \cup \Theta \cup \{A\} \subseteq Ab$, if $\Theta \cup \Delta < A$, $\Theta \not< A$, and $\Theta \cup \Delta \vdash_{\mathcal{R}} \overline{A}$, there are $B \in \min(\Delta)$ and $\Lambda \subseteq ((\Theta \cup \Delta) \setminus \{B\}) \cup \{A\}$ for which $\Lambda \vdash_{\mathcal{R}} \overline{B}$ and $\Lambda \not< B$.*

Proof. Suppose $\Theta \cup \Delta < A$, $\Theta \not< A$, and $\Theta \cup \Delta \vdash_{\mathcal{R}} \overline{A}$. Since $\Theta \cup \Delta < A$ and $\Theta \not< A$, $\min(\Theta \cup \Delta) = \min(\Delta)$. Let $B \in \min(\Delta)$. By contraposition and since $\Theta \cup \Delta \vdash_{\mathcal{R}} \overline{A}$, there is a $\Lambda \subseteq ((\Theta \cup \Delta) \setminus \{B\}) \cup \{A\}$ for which $\Lambda \vdash_{\mathcal{R}} \overline{B}$. By Fact 15 and since $B \in \min(\Theta \cup \Delta)$, $\Lambda \not< B$. ☐

Lemma 12. *Where $\Delta \in \mathsf{d\text{-}comp}(\mathtt{ABF})$ and $C \in Ab$,*

1. *if $\Delta' \subseteq \Delta$ such that $\Delta' \cup \Delta_B \vdash_{\mathcal{R}} \overline{C}$, $\Delta' \not< C$ and $\Delta' \cup \Delta_B < C$, then Δ d-defeats C;*
2. *if $D \in \Delta$, $\Theta \not< D$, $\Theta \cup \Delta_B \vdash_{\mathcal{R}} \overline{D}$ and $\Theta \cup \Delta_B < D$ then Δ d-defeats Θ.*

Proof. Ad 1. By Fact 29 there is a $D \in (\min(\Delta_B) \cap \min(\Delta_B \cup \Delta')) \setminus \Delta'$ such that $D < C$. By contraposition, there is a $\Theta \subseteq ((\Delta' \cup \Delta_B) \setminus \{D\}) \cup \{C\}$ such that $\Theta \vdash_{\mathcal{R}} \overline{D}$. By Fact 15, $\Theta \not< D$. Thus, Θ d-defeats Δ and since Δ is d-complete, Δ d-defeats Θ in some F. Since $(\Delta' \cup \Delta_B) \subseteq \Delta$ and Δ does not attack Δ, $F \notin \Delta' \cup \Delta_B$, $F = C$.

Ad 2. Since $\Theta \not< D$, $\Theta \cup \Delta_B < D$, and by Fact 29, there is an $E \in (\min(\Delta_B) \cap \min(\Theta \cup \Delta_B)) \setminus \Theta$ for which $E < D$. By contraposition, there is a $\Lambda \subseteq ((\Delta_B \cup \Theta) \setminus \{E\}) \cup \{D\}$ for which $\Lambda \vdash_{\mathcal{R}} \overline{E}$. By Fact 15, $\Lambda \not< E$. Thus, Λ d-defeats Δ and since Δ is d-complete, Δ d-defeats Λ in some F. Since Δ does not d-defeat Δ and $\Delta_B \cup \{D\} \subseteq \Delta$, $F \in \Theta$. ☐

Lemma 13. *Where $\Delta \in \mathsf{d\text{-}comp}(\mathtt{ABF})$, if Δ^{-B} d-defeats $C \in Ab^{-B}$ in $\mathtt{ABF}^B$ then Δ d-defeats C in $\mathtt{ABF}$.*

Proof. Suppose Δ^{-B} d-defeats C in $\mathtt{ABF}^B$. Then there is a $\Delta' \subsetneq \Delta^{-B}$ for which $\Delta' \vdash_{\mathcal{R}\cup\{\to B\}} \overline{C}$ and $\Delta' \not< C$. In case $\Delta' \vdash_{\mathcal{R}} \overline{C}$, $\Delta' \subseteq \Delta$ and thus Δ d-defeats C in $\mathtt{ABF}$. Else, by Fact 28, $\Delta' \cup \Delta_B \vdash_{\mathcal{R}} \overline{C}$. If $\Delta' \cup \Delta_B \not< C$, Δ d-defeats C in $\mathtt{ABF}$. Else, by Lemma 12 (item 1), Δ d-defeats C in $\mathtt{ABF}$. □

Lemma 14. *If* $\Delta \in \mathsf{d\text{-}comp}(\mathtt{ABF})$ *then* $\Delta^{-B} \in \mathsf{d\text{-}comp}(\mathtt{ABF}^B)$.

Proof. Suppose $\Delta \in \mathsf{d\text{-}comp}(\mathtt{ABF})$.

d-*conflict-freeness.* Suppose Δ^{-B} d-defeats some $D \in \Delta^{-B}$ in $\mathtt{ABF}^B$. By Lemma 13, Δ d-defeats D in $\mathtt{ABF}$ in contradiction to $\Delta \in \mathsf{d\text{-}comp}(\mathtt{ABF})$.

d-*Admissibility.* Suppose some $\Theta \subseteq Ab^{-B}$ d-defeats Δ^{-B} in D in $\mathtt{ABF}^B$. Thus, $\Theta \vdash_{\mathcal{R}\cup\{\to B\}} \overline{D}$ and $\Theta \not< D$. If $\Theta \vdash_{\mathcal{R}} \overline{D}$ then Δ d-defeats Θ in $\mathtt{ABF}$ since $\Delta \in \mathsf{d\text{-}comp}(\mathtt{ABF})$. If $\Theta \not\vdash_{\mathcal{R}} \overline{D}$, by Fact 28, $\Theta \cup \Delta_B \vdash_{\mathcal{R}} \overline{D}$. We distinguish two cases: (i) $\Theta \cup \Delta_B \not< D$ and (ii) $\Theta \cup \Delta_B < D$. In case (i), Δ d-defeats $\Theta \cup \Delta_B$ in some T in $\mathtt{ABF}$ since $\Delta \in \mathsf{d\text{-}comp}(\mathtt{ABF})$. Since $\Delta_B \subseteq \Delta$, $T \in \Theta$. In case (ii), by Lemma 12 (item 2), Δ d-defeats Θ in $\mathtt{ABF}$. In any case, by Fact 27, also Δ^{-B} d-defeats Θ in $\mathtt{ABF}^B$.

d-*completeness.* Suppose Δ^{-B} d-defends some $A \in Ab^{-B}$ in $\mathtt{ABF}^B$. Suppose further that some $\Theta \subseteq Ab^{-B}$ d-defeats A in $\mathtt{ABF}^B$. Thus, Δ^{-B} d-defeats some $T \in \Theta$ in $\mathtt{ABF}^B$. By Lemma 13, Δ d-defeats T in $\mathtt{ABF}$. Hence, Δ d-defends A in $\mathtt{ABF}$ and since $\Delta \in \mathsf{d\text{-}comp}(\mathtt{ABF})$ also $A \in \Delta$ and hence $A \in \Delta^{-B}$. □

Lemma 15. *If* $\mathsf{d\text{-}grou}(\mathtt{ABF})$ d-*defeats some* $T \in Ab$ *then there is a* $\Delta \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$ *that* d-*defeats* T *and for which* $T < \Delta$ *(and equivalently* $v(T) < v(\Delta)$*).*

Proof. Suppose $\Lambda \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$ d-defeats T and assume i is minimal such that (i) $\Lambda \subseteq \mathsf{d\text{-}grou}_i(\mathtt{ABF})$ (recall Theorem 15) and (ii) Λ is a defeater of T in $\mathsf{d\text{-}grou}(\mathtt{ABF})$. Since $\Lambda \not< T$, $\min(\Lambda) \geq v(T)$. Assume for a contradiction that $\min(\Lambda) = v(T)$. Let $L \in \min(\Lambda)$. By contraposition there is a $\Psi \subseteq (\Lambda \setminus \{L\}) \cup \{T\}$ for which $\Psi \vdash_{\mathcal{R}} \overline{L}$. Since $v(L) = v(T)$, $(\Lambda \setminus \{L\}) \cup \{T\} \not< L$ and so $\Psi \not< L$ by Fact 14. So Ψ and so also $(\Lambda \setminus \{L\}) \cup \{T\}$ d-defeats $L \in \mathsf{d\text{-}grou}_i(\mathtt{ABF})$. But then there is a $\Omega \subseteq \mathsf{d\text{-}grou}_{i-1}(\mathtt{ABF})$ that d-defeats $(\Lambda \setminus \{L\}) \cup \{T\}$. By the conflict-freeness of $\mathsf{d\text{-}grou}(\mathtt{ABF})$ and since $\Lambda \setminus \{L\} \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$, Ω d-defeats T. But this contradicts the minimality of i. So our assumption was wrong and hence $\min(\Lambda) > v(T)$. □

Definition 24. *Let* $\mathsf{d\text{-}grou}^k(\mathtt{ABF}) = \{A \in \mathsf{d\text{-}grou}(\mathtt{ABF}) \mid v(A) = k\}$ *and* $\mathsf{d\text{-}grou}^{>k}(\mathtt{ABF}) = \{A \in \mathsf{d\text{-}grou}(\mathtt{ABF}) \mid v(A) > k\}$.

Lemma 16. $\mathsf{d\text{-}grou}(\mathtt{ABF})^{-B} \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF}^B)$.

Proof. We show this by a double induction following the inductive characterization of $\mathsf{d\text{-}grou}(\mathtt{ABF})$ in Definition 21. Since Ab is finite, there is a maximal k such that $\mathsf{d\text{-}grou}^k(\mathtt{ABF}) \neq \emptyset$ (see Definition 24). Our outer induction is on $i = k, \ldots, 1$ demonstrating that $\mathsf{d\text{-}grou}^i(\mathtt{ABF})^{-B} \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF}^B)$. Our inner induction is as follows: given a fixed i we show that for every $j \geq 0$, $\mathsf{d\text{-}grou}^i_j(\mathtt{ABF})^{-B} \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$ where $\mathsf{d\text{-}grou}^i_j(\mathtt{ABF}) = \mathsf{d\text{-}grou}^i(\mathtt{ABF}) \cap \mathsf{d\text{-}grou}_j(\mathtt{ABF})$.

Outer Base: $i = k$.

- Inner Base: $j = 0$. Suppose $A \in \mathsf{d\text{-}grou}^k_0(\mathtt{ABF})^{-B}$. Assume for a contradiction that $\Theta \subseteq Ab^{-B}$ d-defeats A in $\mathtt{ABF}^B$. Thus, $\Theta \vdash_{\mathcal{R}\cup\{\to B\}} \overline{A}$, $\Theta \not< A$ and hence $\min(\Theta) \geq k = v(A)$. Since $A \in \mathsf{d\text{-}grou}_0(\mathtt{ABF})$, $\Theta \nvdash_{\mathcal{R}} \overline{A}$. However, by Fact 28, $\Theta \cup \Delta_B \vdash_{\mathcal{R}} \overline{A}$. Since $\Theta \cup \Delta_B$ does not d-defeat A in $\mathtt{ABF}$, $\Theta \cup \Delta_B < A$. By Fact 30 there are $D \in \min(\Delta_B)$ and $\Lambda \subseteq ((\Theta \cup \Delta_B) \setminus \{D\}) \cup \{A\}$ such that $\Lambda \vdash_{\mathcal{R}} \overline{D}$ and $\Lambda \not< D$. Hence, $\mathsf{d\text{-}grou}(\mathtt{ABF})$ defends D from this attack. Thus, there is a $\Delta' \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$ that d-defeats some $T \in \Lambda \subseteq ((\Theta \cup \Delta_B) \setminus \{D\}) \cup \{A\}$. Since $\Delta_B \cup \{A\} \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$, $T \in \Theta$ and thus $v(T) \geq k$. By Lemma 15, there is a $\Omega \subseteq \mathsf{d\text{-}grou}^{>k}(\mathtt{ABF})$ that d-defeats T. This contradicts the maximality of k. Thus, no $\Theta \subseteq Ab^{-B}$ d-defeats A in $\mathtt{ABF}^B$ and so $A \in \mathsf{d\text{-}grou}(\mathtt{ABF}^B)$.

- Inner Step: $j \mapsto j+1$. Suppose $A \in \mathsf{d\text{-}grou}^k_{j+1}(\mathtt{ABF})^{-B}$. Assume for a contradiction that $\Theta \subseteq Ab^{-B}$ d-defeats A in $\mathtt{ABF}^B$. Thus, $\Theta \vdash_{\mathcal{R}\cup\{\to B\}} \overline{A}$ and $\Theta \not< A$.

  - Suppose first that $\Theta \vdash_{\mathcal{R}} \overline{A}$. Then, since $\min(\Theta) \geq k$, there is a $\Delta \subseteq \mathsf{d\text{-}grou}^k_j(\mathtt{ABF})$ that d-defeats Θ in $\mathtt{ABF}$.
  - Suppose now that $\Theta \nvdash_{\mathcal{R}} \overline{A}$. Then, by Fact 24, $\Theta \cup \Delta_B \vdash_{\mathcal{R}} \overline{A}$. If $\Theta \cup \Delta_B \not< A$, $\mathsf{d\text{-}grou}(\mathtt{ABF})$ d-defeats Θ in $\mathtt{ABF}$ (since $\Delta_B \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$). Else, if $\Theta \cup \Delta_B < A$, by Fact 30 there are $D \in (\min(\Delta_B)$ and $\Lambda \subseteq ((\Theta \cup \Delta_B) \setminus \{D\}) \cup \{A\}$ for which $\Lambda \not< D$ and $\Lambda \vdash_{\mathcal{R}} \overline{D}$. Thus, $\mathsf{d\text{-}grou}(\mathtt{ABF})$ d-defeats Λ in some T (in $\mathtt{ABF}$) and since $\Delta_B \cup \{A\} \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$, $T \in \Theta$.

  Hence, in any case, by Lemma 15, there is a $\Delta' \subseteq \mathsf{d\text{-}grou}^{>k}(\mathtt{ABF})$ that d-defeats Θ in $\mathtt{ABF}$ which contradicts the maximality of k. Thus, there is no $\Theta \subseteq Ab^{-B}$ that d-defeats A and so $A \in \mathsf{d\text{-}grou}(\mathtt{ABF}^B)$.

Outer Step: $i \mapsto i-1$.

- Inner Base: $j = 0$. Suppose $A \in \mathsf{d\text{-}grou}^{i-1}_0(\mathtt{ABF})^{-B}$. Suppose $\Theta \subseteq Ab^{-B}$ d-defeats A in $\mathtt{ABF}^B$. Thus, $\Theta \vdash_{\mathcal{R}\cup\{\to B\}} \overline{A}$ and $\Theta \not< A$. Note that $\Theta \nvdash_{\mathcal{R}} \overline{A}$

since $A \in \mathsf{d\text{-}grou}_0(\mathtt{ABF})$ and has thus no defeaters. by Fact 24, $\Theta \cup \Delta_B \vdash_{\mathcal{R}} \overline{A}$. Again, since $A \in \mathsf{d\text{-}grou}_0(\mathtt{ABF})$, $\Theta \cup \Delta_B < A$. Thus, by Fact 30 there are $D \in \min(\Delta_B)$ and $\Lambda \subseteq ((\Theta \cup \Delta_B) \setminus \{D\}) \cup \{A\}$ for which $\Lambda \not< D$ and $\Lambda \vdash_{\mathcal{R}} \overline{D}$. Thus, $\mathsf{d\text{-}grou}(\mathtt{ABF})$ d-defeats Λ in some T in $\mathtt{ABF}$. Since $\Delta_B \cup \{A\} \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$, $T \in \Theta$. By Lemma 15 there is a $\Delta' \subseteq \mathsf{d\text{-}grou}^{\geq i}(\mathtt{ABF})$ that d-defeats Θ since $\min(\Theta) \geq i - 1$. By the inductive hypothesis, $\Delta'^{-B} \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF}^B)$ and by Fact 27, Δ'^{-B} d-defeats Θ in $\mathtt{ABF}^B$.

- Inner Step: $j \mapsto j + 1$. Suppose $A \in \mathsf{d\text{-}grou}^{i-1}_{j+1}(\mathtt{ABF})^{-B}$. Suppose $\Theta \subseteq Ab^{-B}$ d-defeats A in $\mathtt{ABF}^B$. Thus, $\Theta \vdash_{\mathcal{R} \cup \{\to B\}} \overline{A}$, $\Theta \not< A$, and thus $\min(\Theta) \geq v(A) = i - 1$.
  - Suppose first that $\Theta \vdash_{\mathcal{R}} \overline{A}$. Then, there is a $\Delta' \subseteq \mathsf{d\text{-}grou}^{\geq \min(\Theta)}(\mathtt{ABF}) \subseteq \mathsf{d\text{-}grou}^{\geq i-1}(\mathtt{ABF})$ such that Δ' d-defeats Θ in $\mathtt{ABF}$. By Lemma 15, there is a $\Delta \subseteq \mathsf{d\text{-}grou}^{\geq i}(\mathtt{ABF})$ that d-defeats Θ in $\mathtt{ABF}$. By the inductive hypothesis, $\Delta^{-B} \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF}^B)$. By Fact 27, Δ^{-B} also d-defeats Θ in $\mathtt{ABF}^B$.
  - Suppose now that $\Theta \nvdash_{\mathcal{R}} \overline{A}$. Then, by Fact 24, $\Theta \cup \Delta_B \vdash_{\mathcal{R}} \overline{A}$. If $\Theta \cup \Delta_B \not< A$, $\mathsf{d\text{-}grou}(\mathtt{ABF})$ d-defeats Θ (since $\Delta_B \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$). Else, if $\Theta \cup \Delta_B < A$, by Fact 30, there are $D \in \min(\Delta_B)$ and $\Lambda \subseteq ((\Theta \cup \Delta_B) \setminus \{D\}) \cup \{A\}$ for which $\Lambda \not< D$ and $\Lambda \vdash_{\mathcal{R}} \overline{D}$. Thus, $\mathsf{d\text{-}grou}(\mathtt{ABF})$ d-defeats some $T \in \Lambda \subseteq (\Theta \cup \Delta_B \setminus \{D\}) \cup \{A\}$ in $\mathtt{ABF}$. Since $\Delta_B \cup \{A\} \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$, $T \in \Theta$.

    So, in both cases $\mathsf{d\text{-}grou}(\mathtt{ABF})$ d-defeats some $T \in \Theta$. Since $v(T) \geq i - 1$ by Lemma 15, there is a $\Delta' \subseteq \mathsf{d\text{-}grou}^{\geq i}(\mathtt{ABF})$ that d-defeats Θ. By the inductive hypothesis, $\Delta'^{-B} \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF}^B)$ and by Fact 27, Δ'^{-B} d-defeats Θ in $\mathtt{ABF}^B$.

We have shown that $\mathsf{d\text{-}grou}(\mathtt{ABF}^B)$ d-defends A and thus that $A \in \mathsf{d\text{-}grou}(\mathtt{ABF}^B)$ since $\mathsf{d\text{-}grou}(\mathtt{ABF}^B)$ is complete.

□

Theorem 8. Any $\mathtt{ABF}$ closed under contraposition is cumulative for $\mid\!\sim^{\mathsf{grou}}_{\mathsf{d}}$.

Proof. We have to show that: $\mathsf{d\text{-}grou}(\mathtt{ABF}) \vdash_{\mathcal{R}} C$ iff $\mathsf{d\text{-}grou}(\mathtt{ABF}^B) \vdash_{\mathcal{R} \cup \{\to B\}} C$ for every $C \in \mathcal{L}$ and that $\mathsf{d\text{-}grou}(\mathtt{ABF})^{-B} = \mathsf{d\text{-}grou}(\mathtt{ABF}^B)$.

We first note $\mathsf{d\text{-}grou}(\mathtt{ABF}) = \bigcap \mathsf{d\text{-}comp}(\mathtt{ABF})$ and $\mathsf{d\text{-}grou}(\mathtt{ABF}^B) = \bigcap \mathsf{d\text{-}comp}(\mathtt{ABF}^B)$ with Theorem 15.

Suppose $A \notin \mathsf{d\text{-}grou}(\mathtt{ABF})^{-B}$. Since $\mathsf{d\text{-}grou}(\mathtt{ABF}) = \bigcap \mathsf{d\text{-}comp}(\mathtt{ABF})$, there is a $\Delta \in \mathsf{d\text{-}comp}(\mathtt{ABF})$ for which $A \notin \Delta^{-B}$. By Lemma 14, $\Delta^{-B} \in \mathsf{d\text{-}comp}(\mathtt{ABF}^B)$. Thus, $A \notin \mathsf{d\text{-}grou}(\mathtt{ABF}^B) = \bigcap \mathsf{d\text{-}comp}(\mathtt{ABF}^B)$. Altogether, $\mathsf{d\text{-}grou}(\mathtt{ABF}^B) \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})^{-B}$.

With Lemma 16, $\mathsf{d\text{-}grou}(\mathtt{ABF})^{-B} = \mathsf{d\text{-}grou}(\mathtt{ABF}^B)$. Since $\Delta_B \subseteq \mathsf{d\text{-}grou}(\mathtt{ABF})$ and by Fact 28, $\mathsf{d\text{-}grou}(\mathtt{ABF}) \vdash_{\mathcal{R}} C$ iff $\mathsf{d\text{-}grou}(\mathtt{ABF}^B) \vdash_{\mathcal{R}\cup\{\to B\}} C$ for all $C \in \mathcal{L}$. □

G.4 Proof of Theorem 9

Fact 31. *Where* $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$ *and* $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}, \mathsf{grou}\}$, $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} B$, *and* $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$, $\Delta = \Delta^{+B}$.

Proof. We distinguish two cases: $B \in Ab$ and $B \notin Ab$. In the latter case, trivially $\Delta = \Delta^{+B}$. In the former case, $B \in \mathrm{Cn}_{\mathcal{R}}(\Theta)$ for every $\Theta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$. Thus, $B \in \mathrm{Cn}_{\mathcal{R}}(\Delta)$. Since $\mathtt{ABF}$ is flat, $B \in \Delta$. □

Fact 32. *Where* $B \in Ab$, $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$ *and* $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}, \mathsf{grou}\}$, *if* $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} B$ *then* $\emptyset \nvdash_{\mathcal{R}} \overline{B}$.

Proof. Suppose $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} B$ and let $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$. By Fact 31, $B \in \Delta$. Were $\emptyset \vdash_{\mathcal{R}} \overline{B}$ then $\emptyset$ would $\mathtt{x}$-defeat Δ in B which is impossible since Δ cannot defend itself from this attack. □

In the following we suppose that $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}\}$, $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, \upsilon)$ is closed under contraposition and that $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} B$. That means that in every $\Delta \in \mathsf{sem}(\mathtt{ABF})$ there is a $\Delta_B \subseteq \Delta$ for which $\Delta_B \vdash_{\mathcal{R}} B$.

Lemma 17. *Where* $\Delta \in \mathtt{d}\text{-}\mathsf{sem}(\mathtt{ABF})$ *and* Δ^{-B} $\mathtt{d}$*-defeats* $\Theta \subseteq Ab^{-B}$ *in* $\mathtt{ABF}^B$ *then also* Δ $\mathtt{d}$*-defeats* Θ *in* $\mathtt{ABF}$.

Proof. Suppose Δ^{-B} $\mathtt{d}$-defeats Θ in $\mathtt{ABF}^B$. Thus, there is a $T \in \Theta$ and there is a $\Delta' \subseteq \Delta^{-B}$ for which $\Delta' \vdash_{\mathcal{R}\cup\{\to B\}} \overline{T}$ and $\Delta' \not< T$. By Fact 28, there is a $\Delta'' \in \{\Delta', \Delta' \cup \Delta_B\}$ such that $\Delta'' \vdash_{\mathcal{R}} \overline{T}$.

If $\Delta'' = \Delta'$ then also Δ $\mathtt{d}$-defeats Θ in $\mathtt{ABF}$.

Suppose $\Delta'' = \Delta' \cup \Delta_B$. Note that $\Delta'' \subseteq \Delta$ since $\Delta_B \subseteq \Delta$ and $\Delta' \subseteq \Delta^{-B} \subseteq \Delta$. We have two cases: $\Delta'' < T$ or $\Delta'' \not< T$. In the latter case Δ $\mathtt{d}$-defeats T in $\mathtt{ABF}$. In the former case $\Delta_B < T$ and by Fact 30 there is a $D \in \min(\Delta_B)$ and a $\Lambda \subseteq ((\Delta' \cup \Delta_B) \setminus \{D\}) \cup \{T\}$ for which $\Lambda \vdash_{\mathcal{R}} \overline{D}$ and $\Lambda \not< D$. Hence, Δ $\mathtt{d}$-defeats $\Lambda \subseteq \Delta' \cup (\Delta_B \setminus \{D\}) \cup \{T\}$ in $\mathtt{ABF}$. Since $\Delta' \cup \Delta_B \subseteq \Delta$ and Δ is d-conflict-free, Δ $\mathtt{d}$-defeats T in $\mathtt{ABF}$. □

Lemma 18 (CC-sem, $\mathsf{d\text{-}stab/pref}$)**.** $\Delta \in \mathsf{d\text{-}sem}(\mathtt{ABF})$ *implies* $\Delta^{-B} \in \mathsf{d\text{-}sem}(\mathtt{ABF}^B)$.

Proof. Suppose $\Delta \in \mathsf{sem}(\mathtt{ABF})$. We now show that $\Delta^{-B} \in \mathsf{sem}(\mathtt{ABF}^B)$.

$\mathtt{d}$-*conflict-free.* Suppose Δ^{-B} $\mathtt{d}$-defeats some $A \in \Delta^{-B}$ in $\mathtt{ABF}^B$. Then Δ $\mathtt{d}$-defeats A in $\mathtt{ABF}$ by Lemma 17. Since $A \in \Delta$ this is a contradiction to the conflict-freeness of Δ in $\mathtt{ABF}$.

d-*admissible.* Suppose some $\Theta \subseteq Ab^{-B}$ d-defeats some $A \in \Delta^{-B}$ in $\mathtt{ABF}^B$. Thus, $\Theta \vdash_{\mathcal{R} \cup \{\to B\}} \overline{A}$ and $\Theta \not< A$.

- If $\Theta \vdash_{\mathcal{R}} \overline{A}$, Δ d-defeats Θ in $\mathtt{ABF}$.
- Else, by Fact 28, $\Theta \cup \Delta_B \vdash_{\mathcal{R}} \overline{A}$.

  If $\Theta \cup \Delta_B \not< A$ then Δ d-defeats $\Theta \cup \Delta_B$ in some E (in $\mathtt{ABF}$). Since $\Delta_B \subseteq \Delta$ and by the d-conflict-freeness of Δ, $E \in \Theta$.

  Else, by Fact 30, there are $C \in \min(\Delta_B)$ and $\Lambda \subseteq ((\Theta \cup \Delta_B) \setminus \{C\}) \cup \{A\}$ such that $\Lambda \vdash_{\mathcal{R}} \overline{C}$ and $\Lambda \not< C$. Again, Δ d-defeats $\Lambda \subseteq (\Theta \cup (\Delta_B \setminus \{C\})) \cup \{A\}$ in some E (in $\mathtt{ABF}$). Since $\Delta_B \cup \{A\} \subseteq \Delta$ and by the d-conflict-freeness of Δ, $E \in \Theta$.

We have shown that in every case, Δ d-defeats Θ in $\mathtt{ABF}$. By Fact 27, Δ^{-B} d-defeats Θ in $\mathtt{ABF}^B$.

$\mathsf{sem} = \mathsf{stab}$. Suppose $A \in Ab^{-B} \setminus \Delta^{-B}$. Thus, Δ d-defeats A. By Fact 27, Δ^{-B} d-defeats A.

$\mathsf{sem} = \mathsf{pref}$. Assume for a contradiction that $\Delta^{-B} \notin \mathsf{d\text{-}pref}(\mathtt{ABF}^B)$. Since, as shown above, $\Delta^{-B} \in \mathsf{d\text{-}adm}(\mathtt{ABF}^B)$, there is a $\Theta \in Ab^{-B}$ such that $\Theta \supset \Delta^{-B}$ and $\Theta \in \mathsf{d\text{-}adm}(\mathtt{ABF}^B)$. We distinguish two cases: (a) $B \in Ab$ and (b) $B \notin Ab$.

Ad (a). By Fact 31, $B \in \Delta$. Also, $\Theta^{+B} \supset \Delta$. Thus Θ^{+B} is not d-admissible in $\mathtt{ABF}$ since Δ is d-preferred in $\mathtt{ABF}$. We will show that from the assumption that $\Delta^{-B} \notin \mathsf{d\text{-}pref}(\mathtt{ABF}^B)$, it follows that Θ^{+B} is d-admissible in $\mathtt{ABF}$, leading to a contradiction.

We first show that Θ^{+B} is d-conflict-free in $\mathtt{ABF}$. Assume otherwise, then there are $\Theta' \subseteq \Theta^{+B}$ and $A \in \Theta^{+B}$ such that $\Theta' \vdash_{\mathcal{R}} \overline{A}$ and $\Theta' \not< A$. Thus, by Fact 27, Θ'^{-B} d-defeats A in $\mathtt{ABF}^B$ or $A = B$. In the former case we have a contradiction to the d-conflict-freeness of Θ in $\mathtt{ABF}^B$. In the second case, $\Theta' \vdash_{\mathcal{R}} \overline{B}$. With Fact 32, $\Theta' \neq \emptyset$. Let $T \in \min(\Theta')$, then by contraposition $\Lambda \vdash_{\mathcal{R}} \overline{T}$ for some $\Lambda \subseteq (\Theta' \setminus \{T\}) \cup \{B\}$. Hence $\Lambda \setminus \{B\} \vdash_{\mathcal{R} \cup \{\to B\}} \overline{T}$ which again contradicts the d-conflict-freeness of Θ in $\mathtt{ABF}^B$ since $\Lambda \setminus \{B\} \not< T$ as $T \in \min(\Theta')$ and $\Lambda \setminus \{B\} \subseteq \Theta' \setminus \{T\}$. Thus, Θ^{+B} is d-conflict-free in $\mathtt{ABF}$.

For showing d-admissibility of Θ^{+B} in $\mathtt{ABF}$ suppose now some $\Lambda \subseteq Ab$ d-defeats some $D \in \Theta^{+B}$.

- If $D \in \Delta$, Δ d-defeats Λ and hence so does Θ^{+B}.

- Else suppose $D \in \Theta^{+B} \setminus \Delta$ and hence $D \neq B$. Since by Fact 27 Λ^{-B} $\mathtt{d}$-defeats Θ(in D) in $\mathtt{ABF}^B$, Θ $\mathtt{d}$-defeats Λ^{-B} in $\mathtt{ABF}^B$ in view of $\Theta \in \mathsf{d\text{-}adm}(\mathtt{ABF}^B)$. Thus, there is a $\Theta' \subseteq \Theta$ for which $\Theta' \vdash_{\mathcal{R}\cup\{\to B\}} \overline{L}$ for some $L \in \Lambda^{-B}$ such that $\Theta' \not< L$.
  - If $\Theta' \vdash_{\mathcal{R}} \overline{L}$ also Θ^{+B} $\mathtt{d}$-defeats Λ in $\mathtt{ABF}$.
  - Else, $\Theta'^{+B} \vdash_{\mathcal{R}} \overline{L}$.

    If $B \not< L$, Θ^{+B} $\mathtt{d}$-defeats Λ in $\mathtt{ABF}$.

    Else, $\{B\} = \min(\Theta'^{+B})$. By Fact 30 there is a $\Theta^\star \subseteq \Theta' \cup \{L\}$ for which $\Theta^\star \vdash_{\mathcal{R}} \overline{B}$ and $\Theta^\star \not< B$. Thus, $\Theta' \cup \{L\}$ $\mathtt{d}$-defeats Δ in $\mathtt{ABF}$ and hence Δ $\mathtt{d}$-defeats $\Theta' \cup \{L\}$ in $\mathtt{ABF}$. Since Θ^{+B} is $\mathtt{d}$-conflict-free and $\Theta^{+B} \supseteq \Delta$, the attack is in L. Thus, Θ^{+B} $\mathtt{d}$-defeats Λ in $\mathtt{ABF}$.

Altogether we have shown that Θ^{+B} is $\mathtt{d}$-admissible in $\mathtt{ABF}$ which is a contradiction to the $\subset$-maximality of Δ.

Ad (b). Again, Θ is not $\mathtt{d}$-admissible in $\mathtt{ABF}$ since $\Theta \supset \Delta = \Delta^{-B}$ and Δ is $\mathtt{d}$-preferred in $\mathtt{ABF}$. We now show that Θ is $\mathtt{d}$-admissible, leading to a contradiction.

We first show that Θ is $\mathtt{d}$-conflict-free in $\mathtt{ABF}$. Assume the opposite. Then there are $\Theta' \subseteq \Theta$ and $A \in \Theta$ for which Θ' $\mathtt{d}$-defeats A in $\mathtt{ABF}$. Thus, also Θ' $\mathtt{d}$-defeats A in $\mathtt{ABF}^B$ which contradicts the conflict-freeness of Θ in $\mathtt{ABF}^B$.

We now show that Θ is $\mathtt{d}$-admissible in $\mathtt{ABF}$. Suppose some $\Lambda \subseteq Ab$ $\mathtt{d}$-defeats Θ. Thus, Λ also $\mathtt{d}$-defeats Θ in $\mathtt{ABF}^B$. Thus, Θ $\mathtt{d}$-defeats Λ in $\mathtt{ABF}^B$. Hence, there is a $\Theta' \subseteq \Theta$ for which $\Theta' \vdash_{\mathcal{R}\cup\{\to B\}} \overline{L}$ for some $L \in \Lambda$ such that $\Theta' \not< L$. Note that $\Theta' \cup \Delta_B \subseteq \Theta$ since $\Theta' \subseteq \Theta \subseteq \Delta = \Delta^{-B}$ and $\Delta_B \subseteq \Delta$. By Fact 28, $\Theta' \cup \Delta_B \vdash_{\mathcal{R}} \overline{L}$ or $\Theta' \vdash_{\mathcal{R}} \overline{L}$.

- If $\Theta' \vdash_{\mathcal{R}} \overline{L}$ also Θ $\mathtt{d}$-defeats Λ in $\mathtt{ABF}$.
- Else either $\Theta' \cup \Delta_B \not< L$ or $\Theta' \cup \Delta_B < L$.
  - If $\Theta' \cup \Delta_B \not< L$, Θ also $\mathtt{d}$-defeats Λ in $\mathtt{ABF}$.
  - Else, by Fact 30, there are $D \in \min(\Delta_B)$ and $\Lambda \subseteq ((\Theta' \cup \Delta_B) \setminus \{D\}) \cup \{L\}$ for which $\Lambda \vdash_{\mathcal{R}} \overline{D}$ and $\Lambda \not< D$. Since $\Delta_B \subseteq \Delta$ and $\Delta \in \mathsf{d\text{-}adm}(\mathtt{ABF})$, Δ $\mathtt{d}$-defeats $\Lambda \subseteq ((\Theta' \cup \Delta_B) \setminus \{D\}) \cup \{L\}$ in $\mathtt{ABF}$. By the $\mathtt{d}$-conflict-freeness of Θ in $\mathtt{ABF}$ and since $\Theta \supset \Delta$, Δ $\mathtt{d}$-defeats L. Hence, also Θ $\mathtt{d}$-defeats Λ in $\mathtt{ABF}$.

Altogether, we have shown that Θ is $\mathtt{d}$-admissible in $\mathtt{ABF}$ which contradicts the $\subset$-maximality of Δ.

$\square$

Theorem 9. Where $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}\}$, any $\mathtt{ABF}$ closed under contraposition satisfies Cautious Cut for $\mid\!\sim^{\mathsf{sem}}_{\mathtt{d}}$.

Proof. In view of Lemma 18 we have to show that if $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} B$ and $\mathtt{ABF}^B \mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} A$ then $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} A$. We show this via contraposition. Suppose (1) $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} B$ and (2) $\mathtt{ABF} \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} A$. We show that $\mathtt{ABF}^B \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} A$.

By (2) there is a $\Delta \in \mathsf{d}\text{-}\mathsf{sem}(\mathtt{ABF})$ for which (3) there is no $\Delta' \subseteq \Delta$ such that $\Delta' \vdash_{\mathcal{R}} A$. By (1), there is a $\Delta_B \subseteq \Delta$ for which $\Delta_B \vdash_{\mathcal{R}} B$. By Lemma 18, $\Delta^{-B} \in \mathsf{d}\text{-}\mathsf{sem}(\mathtt{ABF}^B)$. Assume for a contradiction that there is a $\Theta \subseteq \Delta^{-B}$ such that $\Theta \vdash_{\mathcal{R}\cup\{\to B\}} A$. By Fact 28, $\Theta \vdash_{\mathcal{R}} A$ or $\Theta \cup \Delta_B \vdash_{\mathcal{R}} A$. But this is a contradiction to (3) since $\Theta \cup \Delta_B \subseteq \Delta$. Since $\Delta^{-B} \in \mathsf{d}\text{-}\mathsf{sem}(\mathtt{ABF}^B)$ and there is no $\Delta' \subseteq \Delta^{-B}$ for which $\Delta' \vdash_{\mathcal{R}\cup\{\to B\}} A$, $\mathtt{ABF}^B \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{d}} A$. $\square$

G.5 Proof of Theorem 10

In the following we suppose that $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ is well-behaved (see Definition 12). Recall for this section also Definition 17.

Fact 33. *Where $\Delta \subseteq Ab$ and $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$, if $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathtt{x}} B$ and $\Delta \in \mathsf{IS}(\mathtt{ABF})$ then $\Delta^{-B} \notin \mathsf{CS}(\mathtt{ABF}^B)$.*

Proof. Suppose $\Delta \in \mathsf{IS}(\mathtt{ABF})$. Thus, there is a $C \in \Delta$ such that $\Delta' \vdash_{\mathcal{R}} \overline{C}$ where $\Delta' = \Delta \setminus \{C\}$.

- Suppose $B \notin \Delta'$.
  - If $B \neq C$, $\Delta = \Delta^{-B}$ and, since $\Delta' \vdash_{\mathcal{R}\cup\{\to B\}} \overline{C}$, $\Delta' \cup \{C\} \in \mathsf{IS}(\mathtt{ABF}^B)$ and so $\Delta^{-B} \notin \mathsf{CS}(\mathtt{ABF}^B)$.
  - Else $B = C$. We first note that $\Delta' \neq \emptyset$ since otherwise $\emptyset \vdash_{\mathcal{R}} \overline{B}$ which with Fact 32 contradicts $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathtt{x}} B$. Let $D \in \min(\Delta')$. By contraposition there is a $\Theta \subseteq (\Delta' \setminus \{D\}) \cup \{B\}$ for which $\Theta \vdash_{\mathcal{R}} \overline{D}$ and thus $\Theta \setminus \{B\} \vdash_{\mathcal{R}\cup\{\to B\}} \overline{D}$. Thus, $\Delta^{-B} \notin \mathsf{CS}(\mathtt{ABF}^B)$ since $(\Theta \setminus \{B\}) \cup \{D\} \in \mathsf{IS}(\mathtt{ABF}^B)$ and $(\Theta \setminus \{B\}) \cup \{D\} \subseteq \Delta^{-B}$.
- Suppose now that $B \in \Delta'$. Then $\Delta'^{-B} \vdash_{\mathcal{R}\cup\{\to B\}} \overline{C}$. Again, $\Delta'^{-B} \cup \{C\} \in \mathsf{IS}(\mathtt{ABF}^B)$ and so $\Delta^{-B} \notin \mathsf{CS}(\mathtt{ABF}^B)$.

$\square$

Lemma 19. *Where $\Delta \subseteq Ab$ and $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$, if $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathtt{x}} B$ and $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$ then $\Delta^{-B} \in \mathsf{MCS}(\mathtt{ABF}^B)$.*

Proof. Let $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$. By Lemma 6, $\Delta \in \mathtt{x}\text{-}\mathsf{stab}(\mathtt{ABF})$. Since $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathtt{x}} B$ there is a $\Delta_B \subseteq \Delta$ for which $\Delta_B \vdash_{\mathcal{R}} B$.

Assume first that $\Delta^{-B} \notin \mathsf{CS}(\mathtt{ABF}^B)$. Thus, there is a $\Delta' \in \mathsf{IS}(\mathtt{ABF}^B)$ such that $\Delta' \subseteq \Delta^{-B}$. Hence, there is a $C \in \Delta'$ for which $\Delta' \setminus \{C\} \vdash_{\mathcal{R}\cup\{\to B\}} \overline{C}$. By Fact 28, $\Delta' \setminus \{C\} \vdash_{\mathcal{R}} \overline{C}$ or $(\Delta' \setminus \{C\}) \cup \Delta_B \vdash_{\mathcal{R}} \overline{C}$ and by sanity there is a $\Delta'' \subseteq (\Delta' \cup \Delta_B) \setminus \{C\}$ for which $\Delta'' \vdash_{\mathcal{R}} \overline{C}$. So, $\Delta' \in \mathsf{IS}(\mathtt{ABF})$ or $\Delta'' \cup \{C\} \in \mathsf{IS}(\mathtt{ABF})$. Since $\Delta'' \cup \{C\} \subseteq \Delta' \cup \Delta_B \subseteq \Delta$ and $\Delta \in \mathsf{CS}(\mathtt{ABF})$ this is impossible. Hence, $\Delta^{-B} \in \mathsf{CS}(\mathtt{ABF}^B)$.

Assume now that there is a $\Theta \supset \Delta^{-B}$ for which $\Theta \subseteq Ab^{-B}$ and $\Theta \in \mathsf{CS}(\mathtt{ABF}^B)$. Since $\Theta^{+B} \supset \Delta$ and $\Delta \in \mathsf{MCS}(\mathtt{ABF})$, $\Theta^{+B} \notin \mathsf{CS}(\mathtt{ABF})$ and thus there is a $\Lambda \in \mathsf{IS}(\mathtt{ABF})$ for which $\Lambda \subseteq \Theta^{+B}$. However, since by Fact 33, $\Lambda^{-B} \notin \mathsf{CS}(\mathtt{ABF}^B)$ this is a contradiction since $\Lambda^{-B} \subseteq \Theta$ and $\Theta \in \mathsf{CS}(\mathtt{ABF}^B)$. Thus, $\Delta^{-B} \in \mathsf{MCS}(\mathtt{ABF}^B)$. $\square$

Fact 34. *Where $\Delta, \Theta \subseteq Ab$ and $B \in Ab$,*

1. *if $B \in \Theta$ and $\Theta \prec \Delta$ then $\Theta^{-B} \prec \Delta^{-B}$; and*
2. *if $\Delta, \Theta \subseteq Ab^{-B}$ and $\Delta \prec \Theta$ then $\Delta^{+B} \prec \Theta^{+B}$.*

Proof. Ad 1. Since $\Theta \prec \Delta$ there is an $i \geq 1$ for which $\pi_i(\Theta) \subset \pi_i(\Delta)$ and for all $j > i$, $\pi_j(\Theta) = \pi_j(\Delta)$. Suppose first $B \notin \Delta$. Then $v(B) < i$ and still $\pi_i(\Theta^{-B}) \subset \pi_i(\Delta^{-B})$ while for all $j > i$, $\pi_j(\Theta^{-B}) = \pi_j(\Delta^{-B})$. Suppose now that $B \in \Delta$. Again, $\pi_i(\Theta^{-B}) \subset \pi_i(\Delta^{-B})$ while for all $j > i$, $\pi_j(\Theta^{-B}) = \pi_j(\Delta^{-B})$.

Ad 2. Trivial. $\square$

Lemma 20. *Where $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$ and $\Delta \subseteq Ab^{-B}$, if $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathtt{x}} B$ and $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF}^B)$ then $\Delta^{+B} \in \mathsf{MCS}(\mathtt{ABF})$.*

Proof. Let $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF}^B)$. Assume first that $\Delta^{+B} \notin \mathsf{CS}(\mathtt{ABF})$. Thus, there is a $\Theta \in \mathsf{IS}(\mathtt{ABF})$ for which $\Theta \subseteq \Delta^{+B}$. Thus, by Fact 33, $\Theta^{-B} \notin \mathsf{CS}(\mathtt{ABF}^B)$ which contradicts that $\Delta \in \mathsf{CS}(\mathtt{ABF}^B)$ since $\Theta^{-B} \subseteq \Delta$. Thus, $\Delta^{+B} \in \mathsf{CS}(\mathtt{ABF})$.

Suppose now that there is a $\Theta \in \mathsf{MCS}(\mathtt{ABF})$ for which $\Theta \supset \Delta^{+B}$. Thus, $\Delta^{+B} \prec \Theta$ by Fact 20. By Fact 34 (Item 1) $\Delta \prec \Theta^{-B}$ and since $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF}^B)$, $\Theta^{-B} \notin \mathsf{CS}(\mathtt{ABF}^B)$ and so $\Theta^{-B} \notin \mathsf{MCS}(\mathtt{ABF}^B)$. By Lemma 19, $\Theta \notin \mathsf{MCS}_{\prec}(\mathtt{ABF})$. By Fact 21, there is a $\Lambda \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$ for which $\Theta \prec \Lambda$. Again by Lemma 19, $\Lambda^{-B} \in \mathsf{MCS}(\mathtt{ABF}^B)$. Thus, by Fact 34 (Item 1), $\Theta^{-B} \prec \Lambda^{-B}$ and thus by the transitivity of $\prec$ (Fact 19), $\Delta \prec \Lambda^{-B}$. This is again a contradiction to $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF}^B)$. Thus, there is no $\Theta \in \mathsf{MCS}(\mathtt{ABF})$ for which $\Theta \supset \Delta^{+B}$ and thus $\Delta^{+B} \in \mathsf{MCS}(\mathtt{ABF})$. $\square$

Lemma 21. *Where $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$ and $\mathtt{ABF} \mid\!\sim^{\mathsf{stab}}_{\mathtt{x}} B$, $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$ iff ($\Delta^{-B} \in \mathsf{MCS}_{\prec}(\mathtt{ABF}^B)$ and $\Delta = \Delta^{+B}$).*

Proof. ($\Rightarrow$) Suppose $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$. By Theorem 13, $\Delta \in \mathtt{x}\text{-}\mathsf{stab}(\mathtt{ABF})$ and by Fact 31, $\Delta = \Delta^{+B}$. By Lemma 19, $\Delta^{-B} \in \mathsf{MCS}(\mathtt{ABF}^B)$. Assume there is a $\Theta \in \mathsf{MCS}_{\prec}(\mathtt{ABF}^B)$ for which $\Delta^{-B} \prec \Theta$. By Lemma 20, $\Theta^{+B} \in \mathsf{MCS}(\mathtt{ABF})$. This contradicts $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$ since $\Delta \prec \Theta^{+B}$ by Fact 34.2.

($\Leftarrow$) Suppose $\Delta^{-B} \in \mathsf{MCS}_{\prec}(\mathtt{ABF}^B)$ and $\Delta = \Delta^{+B}$. $\Delta \in \mathsf{MCS}(\mathtt{ABF})$ by Lemma 20,. Assume for a contradiction that $\Delta \notin \mathsf{MCS}_{\prec}(\mathtt{ABF})$. By Fact 21, there is a $\Theta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$ for which $\Delta \prec \Theta$. By Fact 34 (Item 1), $\Delta^{-B} \prec \Theta^{-B}$. By Lemma 19, $\Theta^{-B} \in \mathsf{MCS}(\mathtt{ABF}^B)$. This contradicts $\Delta^{-B} \in \mathsf{MCS}_{\prec}(\mathtt{ABF}^B)$. $\square$

Corollary 5. *Where* $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}\}$, $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$, *and* $\mathtt{ABF} \mathrel{|\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$,

1. *where* $\Delta \subseteq Ab$, $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$ *iff* $\Delta^{+B} \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$ *iff* $\Delta^{-B} \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}^B)$*;*

2. *where* $\Delta \subseteq Ab^{-B}$, $\Delta^{+B} \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$ *iff* $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}^B)$*.*

Proof. Suppose $\mathtt{ABF} \mathrel{|\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$. Note that with Fact 31, for all $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$, $\Delta = \Delta^{+B}$.

Ad 1. Let $\Delta \subseteq Ab$. $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$ iff [by Theorem 13] $\Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$ iff [by Lemma 21] $\Delta^{-B} \in \mathsf{MCS}_{\prec}(\mathtt{ABF}^B)$ iff [by Theorem 13 and Fact 13] $\Delta^{-B} \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}^B)$.

Ad 2. Let $\Delta \subseteq Ab^{-B}$. $\Delta^{+B} \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$ iff [by Theorem 13 and Fact 13] $\Delta^{+B} \in \mathsf{MCS}_{\prec}(\mathtt{ABF})$ iff [by Lemma 21] $\Delta^{+B^{-B}} = \Delta \in \mathsf{MCS}_{\prec}(\mathtt{ABF}^B)$ iff [by Theorem 13] $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}^B)$. $\square$

Theorem 10. Where $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}\}$ and $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$, any well-behaved $\mathtt{ABF}$ is cumulative for $\mathrel{|\!\sim}^{\mathsf{sem}}_{\mathtt{x}}$.

Proof. Suppose $\mathtt{ABF} \mathrel{|\!\sim}^{\mathsf{sem}}_{\mathtt{x}} B$. Thus, there is a $\Delta_B \subseteq \bigcap \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$ for which $\Delta_B \vdash_{\mathcal{R}} B$. By Fact 31, for all $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$, $\Delta = \Delta^{+B}$.

Cautious Monotony: Suppose $\mathtt{ABF}^B \mathrel{|\!\not\sim}^{\mathsf{sem}}_{\mathtt{x}} A$. Thus, there is a $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}^B)$ such that for all $\Delta' \subseteq \Delta$, $\Delta' \nvdash_{\mathcal{R}\cup\{\to B\}} A$. By Corollary 5 (Item 2), $\Delta^{+B} \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$. Assume for a contradiction that there is a $\Delta' \subseteq \Delta^{+B}$ for which $\Delta' \vdash_{\mathcal{R}} A$. But then $\Delta'^{-B} \vdash_{\mathcal{R}\cup\{\to B\}} A$ which is a contradiction since $\Delta'^{-B} \subseteq \Delta$. Thus, $\mathtt{ABF} \mathrel{|\!\not\sim}^{\mathsf{sem}}_{\mathtt{x}} A$.

Cautious Cut: Suppose $\mathtt{ABF} \mathrel{|\!\not\sim}^{\mathsf{sem}}_{\mathtt{x}} A$. Thus, there is a $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$ such that there is no $\Delta' \subseteq \Delta$ for which $\Delta' \vdash_{\mathcal{R}} A$. By Corollary 5 (Item 1), $\Delta^{-B} \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}^B)$. Note again that if there were a $\Delta' \subseteq \Delta^{-B}$ for which $\Delta' \vdash_{\mathcal{R}\cup\{\to B\}} A$ then, by Fact 28, $\Delta' \cup \Delta_B \vdash_{\mathcal{R}} A$ or $\Delta' \vdash_{\mathcal{R}} A$ which was excluded in view of $\Delta' \cup \Delta_B \subseteq \Delta$. Thus, also $\mathtt{ABF}^B \mathrel{|\!\not\sim}^{\mathsf{sem}}_{\mathtt{x}} A$. $\square$

G.6 Proof of Theorems 11 and 12

Let in the following $\mathtt{ABF} = (\mathcal{L}, \mathcal{R}, Ab, \overline{}, \mathbb{V}, \leq, v)$ be well-behaved where $\leq\, = \mathbb{V}^2$. Also suppose in the following that $B \in Ab$ and $\overline{B} \notin \mathrm{Cn}_{\mathcal{R}}(\emptyset)$.

Fact 35. *Every* $\mathtt{r}$*-defeat is also a* $\mathtt{d}$*-defeat.*

Proof. Assume for a contradiction that A $\mathtt{r}$-defeats Δ but does not $\mathtt{d}$-defeat Δ. Thus, $\Delta' \vdash_{\mathcal{R}} \overline{A}$ for some Δ' for which $\Delta' < A$. However, the latter is impossible since $\leq\, = \mathbb{V}^2$. $\square$

Lemma 22. *Where* $\mathsf{sem} \in \{\mathsf{stab}, \mathsf{pref}\}$ *and* $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$*, if* $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}^B)$ *then* $\Delta^{+B} \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$.

Proof. By Theorem 13 and Fact 13 we need only consider the case $\mathsf{sem} = \mathsf{stab}$ and $\mathtt{x} = \mathtt{d}$. Suppose $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}^B)$.

Conflict-freeness. Suppose Δ^{+B} $\mathtt{d}$-defeats some $A \in \Delta^{+B}$ in $\mathtt{ABF}$. Thus, there is a $\Delta' \subseteq \Delta^{+B}$ and a $A \in \Delta^{+B}$ for which $\Delta' \vdash_{\mathcal{R}} \overline{A}$. Thus, $\Delta' \setminus \{B\} \vdash_{\mathcal{R} \cup \{\to B\}} \overline{A}$. Since $\Delta' \setminus \{B\} \subseteq \Delta$ and Δ is conflict-free, $A \notin \Delta$. So, $A = B$ and $\Delta' \vdash_{\mathcal{R}} \overline{B}$. By the sanity of $\mathtt{ABF}$, there is a $\Delta'' \subseteq \Delta' \setminus \{B\}$ for which $\Delta'' \vdash_{\mathcal{R}} \overline{B}$. Note that by the assumption that $\overline{B} \notin \mathrm{Cn}_{\mathcal{R}}(\emptyset)$, $\Delta'' \neq \emptyset$. Let $C \in \Delta''$. By contraposition, there is a $\Theta \subseteq (\Delta'' \setminus \{C\}) \cup \{B\}$ such that $\Theta \vdash_{\mathcal{R}} \overline{C}$ and thus $\Theta \setminus \{B\} \vdash_{\mathcal{R} \cup \{\to B\}} \overline{C}$. Since $(\Theta \setminus \{B\}) \cup \{C\} \subseteq \Delta$ this is in contradiction to the conflict-freeness of Δ in $\mathtt{ABF}^B$.

Stability. Let $A \in Ab \setminus \Delta^{+B}$. Thus, $A \in Ab^{-B} \setminus \Delta$. Thus, Δ $\mathtt{d}$-defeats A in $\mathtt{ABF}^B$. Hence, there is a $\Delta' \subseteq \Delta$ for which $\Delta' \vdash_{\mathcal{R} \cup \{\to B\}} \overline{A}$ which implies $\Delta' \vdash_{\mathcal{R}} \overline{A}$ or $\Delta' \cup \{B\} \vdash_{\mathcal{R}} \overline{A}$. Thus, Δ^{+B} $\mathtt{d}$-defeats A in $\mathtt{ABF}$. $\square$

Theorem 11. *Where* $\mathsf{sem} \in \{\mathsf{pref}, \mathsf{stab}\}$, $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$, $\mathtt{ABF}$ *is Ab-monotonic for* $\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}}$.

Proof. In view of Lemma 22 we still have to show that $\mathtt{ABF} \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} A$ then $\mathtt{ABF}^B \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} A$. We show the contraposition.

Suppose $\mathtt{ABF}^B \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} A$. Thus, there is a $\Delta \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF}^B)$ such that $\Delta' \nvdash_{\mathcal{R} \cup \{\to B\}} A$ for all $\Delta' \subseteq \Delta$. By Lemma 22, $\Delta^{+B} \in \mathtt{x}\text{-}\mathsf{sem}(\mathtt{ABF})$. Assume for a contradiction that there is a $\Delta' \subseteq \Delta^{+B}$ for which $\Delta' \vdash_{\mathcal{R}} A$. But then $\Delta' \setminus \{B\} \vdash_{\mathcal{R} \cup \{\to B\}} A$ which is a contradiction since $\Delta' \setminus \{B\} \subseteq \Delta$. So, $\mathtt{ABF} \not\mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} A$. $\square$

Lemma 23. *Where* $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$, $\mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF})^{-B} \subseteq \mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF}^B)$.

Proof. We show via induction on $i \geq 0$ that if $A \in \mathtt{x}\text{-}\mathsf{grou}_i(\mathtt{ABF})^{-B}$ then $A \in \mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF}^B)$. In view of Fact 35 we only consider $\mathtt{x} = \mathtt{d}$.

$i = 0$. Let $A \in \mathtt{x}\text{-}\mathsf{grou}_0(\mathtt{ABF})^{-B}$. Assume for a contradiction that some $\Theta \subseteq Ab^{-B}$ defeats A in $\mathtt{ABF}^B$. Thus, $\Theta' \vdash_{\mathcal{R} \cup \{\to B\}} \overline{A}$ for a $\Theta' \subseteq \Theta$. Moreover, $\Theta' \nvdash_{\mathcal{R}} \overline{A}$ since

A has no defeaters in $\mathtt{ABF}$. Thus, $\Theta'^{+B} \vdash_{\mathcal{R}} \overline{A}$ which contradicts the fact that $A \in \mathsf{grou}_0(\mathtt{ABF})$. Thus, A has no defeaters in $\mathtt{ABF}^B$ and so $A \in \mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF}^B)$.

$i \mapsto i+1$. Let $A \in \mathtt{x}\text{-}\mathsf{grou}_{i+1}(\mathtt{ABF})^B$. Suppose some $\Theta \subseteq Ab^{-B}$ defeats A in $\mathtt{ABF}^B$. Thus, there is a $\Theta' \subseteq \Theta$ for which $\Theta' \vdash_{\mathcal{R}\cup\{\to B\}} \overline{A}$. Thus, $\Theta' \vdash_{\mathcal{R}} \overline{A}$ or $\Theta' \cup \{B\} \vdash_{\mathcal{R}} \overline{A}$. Hence, there is a $\Lambda \subseteq \mathtt{x}\text{-}\mathsf{grou}_i(\mathtt{ABF})$ and a $T \in \Theta' \cup \{B\}$ such that $\Lambda \vdash_{\mathcal{R}} \overline{T}$. So, $\Lambda \setminus \{B\} \vdash_{\mathcal{R}\cup\{\to B\}} \overline{T}$.

- If $T \in \Theta'$, since by the inductive hypothesis $\Lambda \setminus \{B\} \subseteq \mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF}^B)$, Θ is defeated by $\mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF}^B)$.
- Assume for a contradiction that $T \notin \Theta'$ and so $T = B$ and $\Lambda \setminus \{B\} \vdash_{\mathcal{R}\cup\{\to B\}} \overline{B}$. By the sanity of $\mathtt{ABF}$, there is a $\Lambda' \subseteq \Lambda \setminus \{B\}$ for which $\Lambda' \vdash_{\mathcal{R}} \overline{B}$. Since $\overline{B} \notin \mathrm{Cn}_{\mathcal{R}}(\emptyset)$, $\Lambda' \neq \emptyset$. Let $C \in \Lambda'$. By contraposition, there is a $\Omega \subseteq (\Lambda' \setminus \{C\}) \cup \{B\}$ for which $\Omega \vdash_{\mathcal{R}} \overline{C}$. So, $\Omega \setminus \{B\} \vdash_{\mathcal{R}\cup\{\to B\}} \overline{C}$. But this contradicts the conflict-freeness of $\mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF}^B)$ since by the inductive hypothesis $(\Omega \setminus \{B\}) \cup \{C\} \subseteq \Lambda \setminus \{B\} \subseteq \mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF}^B)$.

We have shown that $\mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF}^B)$ x-defends A and hence $A \in \mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF}^B)$. ☐

Theorem 12. *Where* $\mathtt{x} \in \{\mathtt{d}, \mathtt{r}\}$, $\mathtt{ABF}$ *is Ab-monotonic for* $\mid\!\sim^{\mathsf{grou}}_{\mathtt{x}}$.

Proof. In view of Lemma 23, we have to show that $\mathtt{ABF} \mid\!\sim^{\mathsf{grou}}_{\mathtt{x}} A$ implies $\mathtt{ABF}^B \mid\!\sim^{\mathsf{grou}}_{\mathtt{x}} A$. Suppose $\mathtt{ABF} \mid\!\sim^{\mathsf{grou}}_{\mathtt{x}} A$. Thus, $\Theta \vdash_{\mathcal{R}} A$ for some $\Theta \subseteq \mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF})$. So, $\Theta^{-B} \vdash_{\mathcal{R}\cup\{\to B\}} A$. By Lemma 23, $\Theta^{-B} \subseteq \mathtt{x}\text{-}\mathsf{grou}(\mathtt{ABF}^B)$ and thus $\mathtt{ABF}^B \mid\!\sim^{\mathsf{sem}}_{\mathtt{x}} A$. ☐

 Received 19 September 2018

Groups of Worldview Transformations Implied by Isotropy of Space

Judit X. Madarász
Rényi Institute, Hungary
madarasz.judit@renyi.hu

Mike Stannett
University of Sheffield, UK
m.stannett@sheffield.ac.uk

Gergely Székely
Rényi Institute, Hungary
University of Public Service, Hungary
szekely.gergely@renyi.hu

Abstract

Given any Euclidean ordered field, Q, and any 'reasonable' group, G, of (1+3)-dimensional spacetime symmetries, we show how to construct a model $\mathcal{M}_{\mathsf{G}}$ of kinematics for which the set $\mathbb{W}$ of worldview transformations between inertial observers satisfies $\mathbb{W} = \mathsf{G}$. This holds in particular for all relevant subgroups of Gal, $c\mathsf{Poi}$, and $c\mathsf{Eucl}$ (the groups of Galilean, Poincaré and Euclidean transformations, respectively, where $c \in Q$ is a model-specific parameter corresponding to the speed of light in the case of Poincaré transformations).

In doing so, by an elementary geometrical proof, we demonstrate our main contribution: spatial isotropy is enough to entail that the set $\mathbb{W}$ of worldview transformations satisfies either $\mathbb{W} \subseteq \mathsf{Gal}$, $\mathbb{W} \subseteq c\mathsf{Poi}$, or $\mathbb{W} \subseteq c\mathsf{Eucl}$ for some $c > 0$. So assuming spatial isotropy is enough to prove that there are only 3 possible cases: either the world is classical (the worldview transformations between inertial observers are Galilean transformations); the world is relativistic (the worldview transformations are Poincaré transformations); or the world is Euclidean (which gives a nonstandard kinematical interpretation to Euclidean geometry). This result considerably extends previous results in this field, which assume a priori the (strictly stronger) special principle of relativity, while also restricting the choice of Q to the field $\mathbb{R}$ of reals.

Vol. 8 No. 3 2021
Journal of Applied Logics — IfCoLog Journal of Logics and their Applications

As part of this work, we also prove the rather surprising result that, for any G containing translations and rotations fixing the time-axis $\mathbf{t}$, the requirement that G be a subgroup of one of the groups Gal, $c\mathsf{Poi}$ or $c\mathsf{Eucl}$ is logically equivalent to the somewhat simpler requirement that, for all $g \in \mathsf{G}$: $g[\mathbf{t}]$ is a line, and if $g[\mathbf{t}] = \mathbf{t}$ then g is a trivial transformation (i.e. g is a linear transformation that preserves Euclidean length and fixes the time-axis setwise).

1 Introduction

Physical theories conventionally define coordinate systems and transformations using values and functions defined over the field of reals, $\mathbb{R}$. However, this assumption is not well-founded in physical observation because all physical measurements yield only finite-accuracy values — even quantum electrodynamics (QED), one of the most precisely tested physical theories, is only accurate to around 12 decimal digits [26]. Since we have no empirical reason to make this assumption, it is worth investigating what happens to our expectations of physical theories if we generalize by assuming less about the physical quantities used in measurements. In this paper, we assume only that every positive element in the ordered field of quantities has a square root, but it is worth noting that special relativity can also be modelled over the field of rational numbers [21], in which even this assumption fails. It remains an open question whether the new results presented here generalize over arbitrary ordered fields.

Starting in 1910, Ignatovsky's [18, 19, 20] attempt to derive special relativity assuming only Einstein's principle of relativity initiated a new research direction investigating the consequences of assuming the principle of relativity without Einstein's light postulate. However, Frank and Rothe [10] quickly identified (1911) that hidden assumptions were implicitly used by both Einstein and Ignatovsky, and it is still not uncommon over a century later to find hidden assumptions in related works.

One notable investigation was that of Borisov [7] (see also [17, §10, pp. 60-61]). Borisov explicitly introduced all the assumptions used in his framework investigating the consequences of the principle of relativity. Then he showed that there are basically two possible cases: either the world is classical and the worldview transformations between inertial observers are Galilean; or the world is relativistic and the worldview transformations are Poincaré transformations.[1]

In [23], we made Borisov's framework even more explicit using first-order logic, and investigated the role of his assumption that the structure of physical quantities

[1] Metric geometries corresponding to these two structures also appear among Cayley-Klein geometries; see, e.g., [32] and [28, §6].

is the field of real numbers. We showed that over non-Archimedean fields there is a third possibility: the worldview transformations can also be Euclidean isometries.[2]

In this paper, we present a general axiom system for kinematics using a simple language talking only about quantities, inertial observers (coordinate systems), and the worldview transformations between them. Our axiom system is based on just a few natural assumptions, e.g., instead of assuming that the structure of physical quantities is the field of real numbers we assume only that it is an ordered field Q in which all non-negative values have square roots. Using this framework, we investigate what happens if instead of the principle of relativity we make the weaker assumption that space is isotropic. We show that isotropy is already enough to ensure that the worldview transformations are either Euclidean isometries, or Galilean or Poincaré transformations; see Theorem 5.5 (Classification).

There is an abundance of axiom systems for special relativity in the literature using various basic concepts and basic assumptions, see e.g., [5, 8, 13, 24, 27, 33, 29, 34]. It is natural to ask whether they all capture the same thing – and if not, what is the significance of their differences? Recently, Andréka and Németi have initiated a research project answering these questions by connecting two of these axiom systems by interpretations (logical translation functions) as a first step, see [4]. The investigation presented in this paper forms part of the wider Andréka–Németi school's general project of logic-based axiomatic foundations of relativity theories, see e.g., [1, 2, 3]. Friend and Molinini [11, 12] discuss the significance of this project and the underlying methodology from the viewpoints of epistemology and explanation in science. One important feature of using a first-order logic-based axiomatic framework is that it helps avoid hidden assumptions, which is fundamental in foundational analyses of this nature. Another feature is that it opens up the possibility of machine verification of the results, see e.g., [31, 16].

2 Framework

We are concerned in this paper with two sorts of objects, *(inertial)* ***observers*** and ***quantities***, which we represent as elements of non-empty sets IOb and Q, respectively.

Observers are interpreted to be labels for inertial coordinate systems. Quantities are used to specify coordinates, lengths and related quantities, and we assume that Q is equipped with the usual binary operations, $\cdot$ (multiplication) and $+$ (addition); constants, 0 and 1 (additive and multiplicative identities); and a binary relation, $\leq$

[2] That the principle of relativity is consistent with worldview transformations being Euclidean isometries has previously been shown by Gyula Dávid [9].

(ordering).

Although the results presented here can also be generalized to higher-dimensional spaces (though not necessarily lower-dimensional ones — see Sect. 8), we assume for definiteness that observers inhabit 4-dimensional spacetime, Q^4, and locations in spacetime are accordingly represented as 4-tuples over Q. We often write $\vec{p}$, $\vec{q}$ and $\vec{r}$ to denote generic spacetime locations.

For each pair of observers $k, h \in IOb$, we assume the existence of a function $\mathsf{w}_{kh} \colon Q^4 \to Q^4$, called the ***worldview transformation*** *from the worldview of h to the worldview of k*, which we interpret as representing the idea that observers may see (i.e. coordinatize) the same events, but at different spacetime locations: whatever is seen by h at $\vec{p}$ is seen by k at $\mathsf{w}_{kh}(\vec{p})$.[3]

Formally, this framework corresponds to using a two sorted first-order language where the models are of the following form

$$\mathcal{M} = (IOb, Q, +, \cdot, 0, 1, \leq, \mathsf{w}),$$

where: IOb and Q are two sorts; $+$ and $\cdot$ are binary operations on Q; 0 and 1 are constants on Q; $\leq$ is a binary relation on Q; and w is a function from $IOb \times IOb \times Q^4$ to Q^4. In this language, the worldview transformation between fixed observers k and h can be introduced as:

$$\mathsf{w}_{kh}(t, x, y, z) \stackrel{\text{def}}{=} \mathsf{w}(k, h, t, x, y, z).$$

3 Axioms

In this section, we describe the general axiom system, KIN, used to represent kinematics in this paper. Additional axioms representing spatial isotropy and the special principle of relativity will be introduced in Section 4.

3.1 Quantities

We assume that $(Q, +, \cdot, 0, 1, \leq)$ exhibits the most fundamental algebraic properties expected of the real numbers ($\mathbb{R}$), so that calculations can be performed and results compared with one another. We also assume that square-roots are defined for non-negative values (i.e. that Q is a ***Euclidean field*** [25]).

[3] In more general theories, for example in general relativity, this relation need not be a function or even defined on the whole Q^4, because an event seen by k may be invisible to h or may appear at one or more different spacetime locations from h's point of view, but in this paper we assume that all observers completely and unambiguously coordinatize the same universe — they all see the same events, albeit in different locations relative to one another.

AxEField $(Q, +, \cdot, 0, 1, \leq)$ is a Euclidean field, i.e. a linearly ordered field in which every non-negative element has a square root.

Assuming AxEField also means that the derived operations of subtraction $(-)$, division $(/)$, square root $(\sqrt{\ })$, dot product of vectors $(\cdot)$, Euclidean length of vectors, etc., are well-defined on their domains, and allows us to assume the usual vector space structure of Q^4 over Q. We will generally omit the multiplication symbol.

3.2 Worldview transformations

The following axiom states informally that: (i) the worldview transformation from an observer's worldview to itself is just the identity transformation, $\mathsf{Id}\colon Q^4 \to Q^4$; and (ii) switching from k's worldview to h's and then to m's has the same effect as switching directly from k's worldview to m's.

AxWvt For all $k, h, m \in IOb$:

(i) $\mathsf{w}_{kk} = \mathsf{Id}$;

(ii) $\mathsf{w}_{mh} \circ \mathsf{w}_{hk} = \mathsf{w}_{mk}$.

3.3 Lines, worldlines and motion

By assumption, all of the locations under discussion in this paper are points in Q^4. We often write (t, x, y, z) to indicate the coordinates of a generic point in Q^4. Given any $n > 0$ and $\vec{p} = (p_1, p_2, \ldots, p_n) \in Q^n$, its ***squared length***, $|\vec{p}\,|^2$, is defined by

$$|\vec{p}\,|^2 \stackrel{\text{def}}{=} p_1^2 + \ldots + p_n^2.$$

(This is just the standard Euclidean squared length of $\vec{p}$.)

To simplify our notation, we write $\vec{\mathsf{o}} \stackrel{\text{def}}{=} (0, 0, 0, 0)$ for the zero-vector (origin) in Q^4. More generally, we sometimes write $\vec{0}$ for any tuple of zeroes (the length will always be clear from context). We define the ***time-axis***, $\mathbf{t}$, and the ***present simultaneity***, $\mathbf{S}$, to be the set

$$\mathbf{t} \stackrel{\text{def}}{=} \{(t, 0, 0, 0) : t \in Q\}$$

and the spatial hyperplane

$$\mathbf{S} \stackrel{\text{def}}{=} \{(0, x, y, z) : x, y, z \in Q\},$$

respectively. We write $\vec{\mathsf{t}}$ for the unit time vector $(1,0,0,0)$, and likewise $\vec{\mathsf{x}} \stackrel{\text{def}}{=} (0,1,0,0)$, $\vec{\mathsf{y}} \stackrel{\text{def}}{=} (0,0,1,0)$ and $\vec{\mathsf{z}} \stackrel{\text{def}}{=} (0,0,0,1)$. If $\vec{p} = (t,x,y,z) \in Q^4$, we call $\vec{p}_t \stackrel{\text{def}}{=} t$ the ***time component***, and $\vec{p}_s \stackrel{\text{def}}{=} (x,y,z)$ the ***space component***, of $\vec{p}$. Finally, if $t \in Q$ and $\vec{s} \in Q^3$, we write $(t, \vec{s})$ for the point with time component t and space component $\vec{s}$.

The ***worldline*** *of observer h according to observer k* is defined as

$$\mathsf{wl}_k(h) \stackrel{\text{def}}{=} \mathsf{w}_{kh}[\mathbf{t}].$$

In particular, if we assume AxWvt and take $k = h$, we have $\mathsf{wl}_h(h) = \mathsf{w}_{hh}[\mathbf{t}] = \mathbf{t}$. This corresponds to the convention that observers consider themselves to be at the spatial origin relative to which measurements are made: from their own viewpoint their worldline is the time-axis; and $\mathsf{wl}_k(h) = \mathsf{w}_{kh}[\mathbf{t}] = \mathsf{w}_{kh}[\mathsf{wl}_h(h)]$ describes the same worldline but from k's point of view.

When we say that one observer moves *inertially* with respect to another, we mean that neither of them accelerates relative to the other, so that linear motions seen by one remain linear when seen by the other. Since each observer considers its own worldline to be the line $\mathbf{t}$, we would expect all inertial observers to agree that each others' world lines are lines.

Formally, a subset $\ell \subseteq Q^4$ is a ***line*** iff there are $\vec{p}, \vec{v} \in Q^4$, where $\vec{v} \neq \vec{\mathsf{o}}$ and $\ell = \{\vec{p} + \lambda\vec{v} \, : \, \lambda \in Q\}$. The next axiom states that worldlines of observers according to observers are lines.

AxLine For every $k, h \in IOb, \mathsf{wl}_k(h)$ is a line.

According to AxLine, the worldlines of observers are lines, and by AxWvt each observer considers its own worldline to be the time-axis; we can therefore express the idea that *observer k is **moving according to** observer m* by saying that $\mathsf{wl}_m(k)$ is not parallel to $\mathbf{t}$,[4] or more simply, that w_{mk} takes the time-unit vector $\vec{\mathsf{t}}$ and the zero-vector $\vec{\mathsf{o}}$ to coordinate points having different spatial components, i.e. $\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_s \neq \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_s$. In the same spirit, we say that k ***is at rest according to*** m iff $\mathsf{w}_{mk}(\vec{\mathsf{t}})_s = \mathsf{w}_{mk}(\vec{\mathsf{o}})_s$.

We will sometimes need to assume explicitly the existence of observers moving relative to one another, which we express using the following formula:

$\exists$MovingIOb There are observers $m, k \in IOb$ such that $\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_s \neq \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_s$.

[4]As one would expect, being in motion relative to another observer — and likewise being at rest — are symmetric relations; see Lemma 6.6.2 (Rest).

3.4 Trivial transformations

We say that a linear transformation $T : Q^4 \to Q^4$ is a ***linear trivial transformation*** provided it fixes (setwise) both the time-axis and the present simultaneity, and preserves squared lengths in both, i.e.

- if $\vec{p} \in \mathbf{t}$, then $T(\vec{p}\,) \in \mathbf{t}$ and $T(\vec{p}\,)_t^2 = \vec{p}_t^{\,2}$; and
- if $\vec{p} \in \mathbf{S}$, then $T(\vec{p}\,) \in \mathbf{S}$ and $|T(\vec{p}\,)_s|^2 = |\vec{p}_s|^2$.

Remark 3.1. Assuming $\mathsf{AxEField}$, the statement that T is a linear trivial transformation is equivalent to the statement that T is a linear transformation that preserves Euclidean length and fixes the time-axis setwise.[5] □

A map $f : Q^4 \to Q^4$ is a ***translation*** iff there is $\vec{q} \in Q^4$ such that $f(\vec{p}\,) = \vec{p} + \vec{q}$ for every $\vec{p} \in Q^4$. We write Trans for the *set of all translations*.

A transformation is called a ***trivial transformation*** if it is a linear trivial transformation composed with a translation. We write Triv for the *set of all trivial transformations*.

We say that two observers k and k' are ***co-located*** if they consider themselves to share the same worldline: $\mathsf{wl}_k(k) = \mathsf{wl}_k(k')$ (assuming AxWvt, this relationship is symmetric; see Lemma 6.3.5 (Equal Worldlines)). The following axiom says that, if observers k and k' are co-located, then their worldviews are related to one another by a trivial transformation. In other words, even though inertial observers following the same worldline may use different coordinate systems, these coordinate systems can only differ by using a different orthonormal basis for coordinatizing space and/or a different direction and origin of time.[6]

$\underline{\mathsf{AxColocate}}$ For all $k, k' \in IOb$, if $\mathsf{wl}_k(k) = \mathsf{wl}_k(k')$, then $\mathsf{w}_{kk'} \in \mathsf{Triv}$.

[5] This claim follows by Lemma 6.3.2 ($\mathsf{Triv} = \bigcap_\kappa \mathsf{Iso}$), but can also be proven directly. Suppose T is linear, preserves Euclidean length and fixes $\mathbf{t}$ setwise. It follows immediately that $T(\vec{\mathsf{t}}\,) = \pm\vec{\mathsf{t}}$. Now choose any $(0, \vec{s}\,) \in \mathbf{S}$, and suppose $T(0, \vec{s}\,) = (t', \vec{s}\,')$. Then $|T(\pm 1, \vec{s}\,)|^2 = |T(0, \vec{s}\,) \pm T(\vec{\mathsf{t}}\,)|^2 = (t' \pm 1)^2 + |\vec{s}\,'|^2$. Since $|(1, \vec{s}\,)|^2 = |(-1, \vec{s}\,)|^2$ and T preserves Euclidean length, we therefore require $(t' + 1)^2 + |\vec{s}\,'|^2 = (t' - 1)^2 + |\vec{s}\,'|^2$, whence $t' = 0$. Thus, T also fixes $\mathbf{S}$, so it is a linear trivial transformation. The converse is trivial.

[6] By AxWvt, if k and k' are co-located, i.e. $\mathsf{wl}_k(k) = \mathsf{wl}_k(k')$, then $\mathsf{w}_{kk'}[\mathbf{t}] = \mathbf{t}$. This is why we do not need to assume explicitly in the statement of $\mathsf{AxColocate}$ that co-located observers share the same time-axis.

3.5 Spatial rotations.

A linear trivial transformation $R : Q^4 \to Q^4$ is called a ***spatial rotation*** iff it preserves the direction of time and the orientation of space, i.e. $R(\vec{\mathsf{t}}) = \vec{\mathsf{t}}$ and the determinant of 3×3 matrix $[R(\vec{\mathsf{x}})_s, R(\vec{\mathsf{y}})_s, R(\vec{\mathsf{z}})_s]$ is positive.[7] We denote the *set of all spatial rotations* by SRot.

The following axiom says that translated and spatially rotated versions of any inertial coordinate system are also inertial coordinate systems.[8]

$\underline{\mathsf{AxRelocate}}$ For all $k \in IOb$ and for all $T \in \mathsf{Trans} \cup \mathsf{SRot}$, there is $h \in IOb$ such that $\mathsf{w}_{kh} = T$.

The underlying axiom system with which we are concerned in this paper is

$$\mathsf{KIN} \stackrel{\text{def}}{=} \{\mathsf{AxEField}, \mathsf{AxWvt}, \mathsf{AxLine}, \mathsf{AxRelocate}, \mathsf{AxColocate}\},$$

which defines our basic theory of the ***kinematics*** of inertial observers.

4 The special principle of relativity, isotropy and set of worldview transformations

There are many different formal interpretations of the principle of relativity [14, 15, 22]. In this paper, we interpret the ***special principle of relativity*** (*SPR*) to mean that all inertial observers agree as to how they are related to other observers, so that no observer can be distinguished from any other in terms of the things they can and cannot (potentially) observe. We express this via the following axiom:

$\underline{\mathsf{AxSPR}}$ For every $k, k^*, h \in IOb$, there exists $h^* \in IOb$ such that $\mathsf{w}_{kh} = \mathsf{w}_{k^*h^*}$,

that is, given observers k, k^*, h, there must (potentially) be some h^* which is related to k^* in exactly the same way that h is related to k, i.e. the geometrical structure of spacetime cannot forbid such an observer.

[7] This can be expressed in our formal language without any assumption about the structure of quantities as: $R(\vec{\mathsf{x}})_2 R(\vec{\mathsf{y}})_3 R(\vec{\mathsf{z}})_4 + R(\vec{\mathsf{x}})_4 R(\vec{\mathsf{y}})_2 R(\vec{\mathsf{z}})_3 + R(\vec{\mathsf{x}})_3 R(\vec{\mathsf{y}})_4 R(\vec{\mathsf{z}})_2 > R(\vec{\mathsf{x}})_4 R(\vec{\mathsf{y}})_3 R(\vec{\mathsf{z}})_2 + R(\vec{\mathsf{x}})_2 R(\vec{\mathsf{y}})_4 R(\vec{\mathsf{z}})_3 + R(\vec{\mathsf{x}})_3 R(\vec{\mathsf{y}})_2 R(\vec{\mathsf{z}})_4$, here $R(\vec{p})_2$, $R(\vec{p})_3$, and $R(\vec{p})_4$ denote the second, third and fourth component of $R(\vec{p}) \in Q^4$, i.e. if $R(\vec{p}) = (t, x, y, z)$, then $R(\vec{p})_2 = x$, $R(\vec{p})_3 = y$, and $R(\vec{p})_4 = z$.

[8] The quantification over T in $\mathsf{AxRelocate}$ appears at first sight to be second-order. However, because translations are determined by the image of the origin, while spatial rotations are determined by the images of the three spatial unit vectors, this axiom can be formalized in our first-order logic language by quantifying over the 4 parameters representing the image of the origin and the 12 parameters representing the images of the three spatial unit vectors.

In contrast, ***isotropy*** refers to the weaker constraint that there is no distinguished direction in space, i.e. no matter which direction we face, we should be able to perform the same experiments and observe the same outcomes. Isotropy can be expressed in much the same way as SPR, except that we only require equivalence as to what can be observed (h) when the relevant observers (k and k^*) are related via a spatial rotation (see Figure 1):

AxIsotropy For every $k, k^*, h \in IOb$, if $\mathsf{w}_{kk^*} \in \mathsf{SRot}$, there exists $h^* \in IOb$ such that $\mathsf{w}_{kh} = \mathsf{w}_{k^*h^*}$.

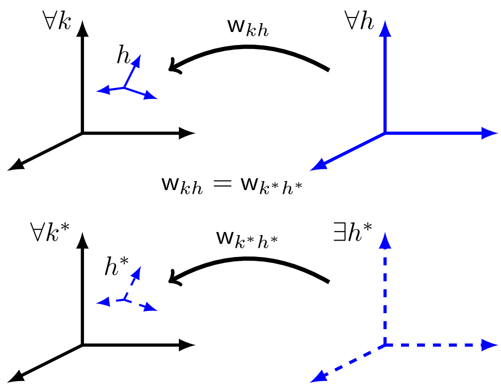


Figure 1: Isotropy and the special principle of relativity. The special principle, AxSPR, says that given any k, h and k^*, there exists an h^* that is related to k^* the same way that h is related to k (i.e. there are no distinguished inertial coordinate systems). Spatial isotropy, AxIsotropy, is similar, except that we only require h^* to exist when w_{kk^*} is a spatial rotation (i.e. rotating ones spatial coordinate system has no effect on what can and cannot potentially be seen).

In order to investigate these ideas, we will need to consider various sets of worldview transformations, and attempt to establish both their algebraic properties and the relationships between them. The set $\mathbb{W}_k$ of worldview transformations associated with a specific observer $k \in IOb$ will be defined by

$$\mathbb{W}_k \stackrel{\text{def}}{=} \{\mathsf{w}_{kh} : h \in IOb\}$$

and the set of *all* worldview transformations is then given by

$$\mathbb{W} \stackrel{\text{def}}{=} \{\mathsf{w}_{kh} : k, h \in IOb\} = \bigcup\{\mathbb{W}_k : k \in IOb\}.$$

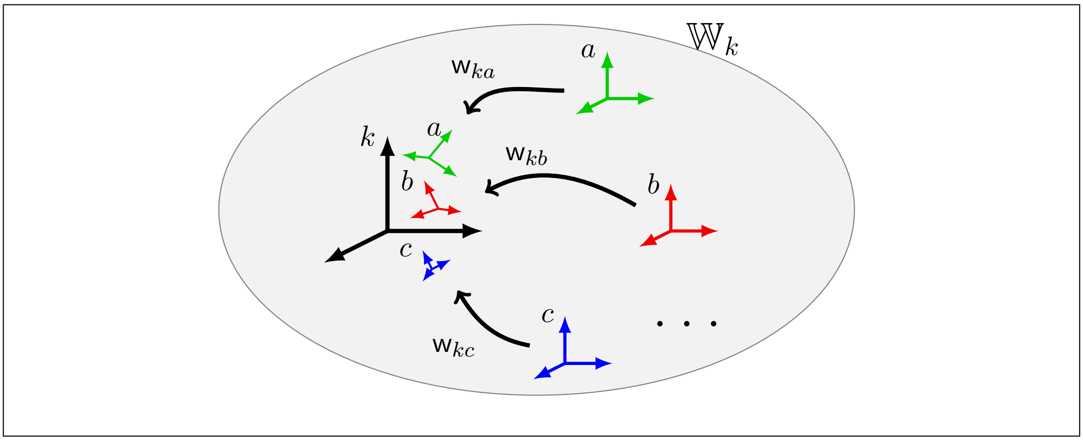


Figure 2: The set $\mathbb{W}_k$ of all worldview transformations into k's coordinate system. For each observer $a, b, c, \ldots$, the set $\mathbb{W}_k$ contains the associated transformation $\mathsf{w}_{ka}, \mathsf{w}_{kb}, \mathsf{w}_{kc}, \ldots$.

Using these notations AxSPR can be reformulated as saying that all inertial observers have essentially the same worldview, i.e. $\mathbb{W}_k = \mathbb{W}_{k^*}$ for all $k, k^* \in IOb$. Although it is not immediately obvious that any $\mathbb{W}_k$ can form a group, if we assume AxWvt it can be proven that AxSPR is equivalent to saying that there is at least one k for which $\mathbb{W}_k$ forms a group under composition, which is itself equivalent to saying that $\mathbb{W}_k = \mathbb{W}$. For the proof of this and other equivalent formulations of AxSPR, see [23, Prop. 2.1]. Similarly, AxIsotropy is equivalent to saying, for all $k, k^* \in IOb$, if $\mathsf{w}_{kk^*} \in \mathsf{SRot}$, then $\mathbb{W}_k = \mathbb{W}_{k^*}$.

Remark 4.1. We have already noted that AxSPR entails AxIsotropy, so that the special principle of relativity is at least as strong assumption as spatial isotropy. In fact, it is strictly stronger, because $\mathbb{W}$ is a group in all models of KIN + AxIsotropy, but $\mathbb{W}_k$ need not be. In particular, therefore, KIN + AxIsotropy does not imply AxSPR. This remains true even if we add the restriction that $(Q, +, \cdot, 0, 1, \leq)$ is the ordered field of real numbers. However, if we add the assumption that co-located observers agree on the direction of time, then it can be shown that KIN + AxIsotropy implies AxSPR.

For easy reference, Table 1 summarizes the axioms used in this paper and discussed above.

KIN	Axiom	Description
✓	AxEField	the set Q of quantities is an ordered field in which all non-negative values have square roots
✓	AxWvt	w_{kk} transforms k's worldview to itself identically; and going from k's worldview to h's and then to m's is same as going directly from k's worldview to m's
✓	AxLine	inertial observers see each other's worldlines as lines
✓	AxColocate	if two observers are co-located, their worldviews are trivially related to one another
✓	AxRelocate	translated and spatially rotated versions of inertial coordinate systems are also inertial
	AxSPR	the special principle of relativity
	AxIsotropy	isotropy of space

Table 1: Our axioms and their intuitive meanings.

5 Main theorems

First let us introduce the transformations that will be used in this paper to characterize the worldviews of observers. In this section, we assume that $(Q, +, \cdot, 0, 1)$ is a field. Table 2 summarizes the various transformation groups referred to in the theorems.

Trans	translations		
SRot	spatial rotations		
Triv	trivial transformations		
${}_{\kappa}\mathsf{Iso}$	κ-isometries		
$c\mathsf{Poi}$	c-Poincaré transformations	$=$	${}_{1/c^2}\mathsf{Iso}$
$c\mathsf{Eucl}$	c-Euclidean transformations	$=$	${}_{-1/c^2}\mathsf{Iso}$
Gal	Galilean transformations	$=$	${}_{0}\mathsf{Iso}$

Table 2: Transformation groups considered in this paper.

5.1 κ-isometries

Given $\vec{p} = (t, x, y, z)$, the ***(squared) κ-length*** of $\vec{p}$ is defined by

$$\|(t, x, y, z)\|_\kappa^2 \stackrel{\text{def}}{=} t^2 - \kappa(x^2 + y^2 + z^2),$$

or in other words,

$$\|\vec{p}\,\|_\kappa^2 \stackrel{\text{def}}{=} \vec{p}_t^{\,2} - \kappa|\vec{p}_s|^2.$$

Taking $\kappa = 1$ gives the squared *Minkowski length* $\|\vec{p}\,\|_1^2 = t^2 - (x^2 + y^2 + z^2)$ of $\vec{p}$, while $\kappa = -1$ gives its squared *Euclidean length*, $\|\vec{p}\,\|_{-1}^2 = |\vec{p}\,|^2 = t^2 + (x^2 + y^2 + z^2)$.

Definition 5.1.1 (κ-isometry, $\kappa \neq 0$)**.** If $\kappa \neq 0$, we call a linear transformation $f : Q^4 \to Q^4$ a ***linear κ-isometry*** provided it preserves κ-length, i.e. for every $\vec{p} \in Q^4$,

$$\|f(\vec{p}\,)\|_\kappa^2 = \|\vec{p}\,\|_\kappa^2 .$$

In the case of $\kappa = 0$, we require more than simply preserving 0-length, for while 0-length takes account of temporal extent, it ignores spatial structure. We therefore need to add an extra condition to the definition of 0-isometry to ensure that spatial structure is also respected when considering points with equal time coordinates.[9]

Definition 5.1.2 (κ-isometry, $\kappa = 0$)**.** Let $f : Q^4 \to Q^4$ be a linear transformation. We call f a ***linear 0-isometry*** provided, for every $\vec{p} \in Q^4$,

$$f(\vec{p}\,)_t^2 = \vec{p}_t^{\,2} \text{ and } \left(\vec{p}_t = 0 \;\Rightarrow\; |f(\vec{p}\,)_s|^2 = |\vec{p}_s|^2\right). \tag{5.1}$$

We call the composition of a linear κ-isometry and a translation a ***κ-isometry***, and write ${}_\kappa\mathsf{Iso}$ for the *set of all κ-isometries.*

Definition 5.1.3 ($c\mathsf{Poi}$, $c\mathsf{Eucl}$ and Gal)**.** For $c > 0$, $1/c^2$-isometries will be called ***c-Poincaré transformations*** and $-1/c^2$-isometries will be called ***c-Euclidean isometries***. Parameter c in c-Poincaré transformations corresponds to the "speed of light". A 0-isometry is also called a ***Galilean symmetry***. We denote these sets of transformations by $c\mathsf{Poi}$, $c\mathsf{Eucl}$ and Gal, respectively.

It is easily verified that each of these sets forms a group under function composition. In general, when we speak about a set G of transformations as a group, we mean G under function composition, i.e. $(\mathsf{G}, \circ)$. As usual, we write $\mathsf{H} \leq \mathsf{G}$ to mean that H is a subgroup of G, and $\mathsf{H} < \mathsf{G}$ to mean that the inclusion is proper.

[9]Although every 0-isometry preserves 0-length, the converse is not true.

We note that 1-Poincaré transformations form the usual group Poi of Poincaré transformations and 1-Euclidean isometries form the usual group Eucl of Euclidean isometries. Notice also that trivial transformations, translations and spatial rotations are κ-isometries for all values of κ. Moreover, by Lemma 6.3.2 ($\mathsf{Triv} = \bigcap {}_{\kappa}\mathsf{Iso}$),

$$\mathsf{Trans} \cup \mathsf{SRot} \subset \mathsf{Triv} = \bigcap_{\kappa \in Q} {}_{\kappa}\mathsf{Iso} = {}_{x}\mathsf{Iso} \cap {}_{y}\mathsf{Iso} \tag{5.2}$$

for any two distinct $x, y \in Q$. It follows immediately that $\mathsf{Trans} \cup \mathsf{SRot} \subset c\mathsf{Poi} \cap c\mathsf{Eucl} \cap \mathsf{Gal}$.

5.2 The theorems

Our first result, Theorem 5.1 (Characterisation), tells us that if space is isotropic then all worldview transformations are κ-isometries for some κ, and shows how to calculate the value of κ in the case that two observers can be found which move relative to one another.

Theorem 5.1 (Characterisation)**.** Assume $\mathsf{KIN} + \mathsf{AxIsotropy}$. Then there is a $\kappa \in Q$ such that the set of worldview transformations is a set of κ-isometries, i.e.

$$\mathbb{W} \subseteq {}_{\kappa}\mathsf{Iso}.$$

In other terms,

$$\text{either } \mathbb{W} \subseteq c\mathsf{Poi},\ \mathbb{W} \subseteq \mathsf{Gal}, \text{ or } \mathbb{W} \subseteq c\mathsf{Eucl} \text{ for some } c > 0.$$

Moreover,

- if $\neg\exists\mathsf{MovingIOb}$ is assumed, then $\mathbb{W} \subseteq \mathsf{Triv}$;
- if $\exists\mathsf{MovingIOb}$ is assumed, this κ is uniquely determined by the w_{mk}-images of $\vec{\mathsf{o}}$ and $\vec{\mathsf{t}}$ where m and k are observers moving relative to one another, and can be calculated as
$$\kappa = \frac{\left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_t - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_t\right|^2 - 1}{\left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_s - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_s\right|^2}.$$

□

For all positive $c \in Q$, the group $c\mathsf{Poi}$ is isomorphic to group Poi (via natural inner automorphisms of the affine group, representing the effects of changing the spatial or temporal units of measurements) and similarly group $c\mathsf{Eucl}$ is isomorphic to the Euclidean transformation group Eucl (via the same inner automorphisms);

see [23, Prop. 6.9]. So essentially there are only three nontrivial cases: either all the worldview transformations are relativistic; all of them are classical; or all of them are Euclidean isometries. Subject to this constraint, however, Theorem 5.3 (Model Construction) says that all 'reasonable' transformation groups (groups containing the translations and spatial rotations, which we know must be present) can occur as the group of worldview transformations in a model of $\mathsf{KIN} + \mathsf{AxSPR}$.

To present a general model construction, let us write $\mathsf{Sym}(Q^4)$ for the set of all permutations of Q^4. Given any transformation group $\mathsf{G} \leq \mathsf{Sym}(Q^4)$, we define a model $\mathcal{M}_\mathsf{G}$ of our language by taking $IOb := \mathsf{G}$ and $\mathsf{w}_{mk} := m \circ k^{-1}$ for $k, m \in \mathsf{G}$. Theorem 5.2 (Satisfaction) connects the axioms of KIN to properties of G.

Theorem 5.2 (Satisfaction)**.** Let $\mathsf{G} \leq \mathsf{Sym}(Q^4)$. Then

(a) $\mathcal{M}_\mathsf{G}$ satisfies AxWvt, AxSPR and $\mathbb{W} = \mathsf{G}$.

(b) $\mathcal{M}_\mathsf{G}$ satisfies $\mathsf{AxRelocate}$ iff $\mathsf{SRot} \cup \mathsf{Trans} \subseteq \mathsf{G}$.

(c) $\mathcal{M}_\mathsf{G}$ satisfies AxLine iff $g[\mathbf{t}]$ is a line for all $g \in \mathsf{G}$.

(d) $\mathcal{M}_\mathsf{G}$ satisfies $\mathsf{AxColocate}$ iff $g \in \mathsf{Triv}$ whenever $g \in \mathsf{G}$ and $g[\mathbf{t}] = \mathbf{t}$.

Theorem 5.3 (Model Construction)**.** Assume $\mathsf{AxEField}$. Let G be a group such that

- $\mathsf{SRot} \cup \mathsf{Trans} \subseteq \mathsf{G} \leq c\mathsf{Poi}$ for some $c \in Q$; or
- $\mathsf{SRot} \cup \mathsf{Trans} \subseteq \mathsf{G} \leq c\mathsf{Eucl}$ for some $c \in Q$; or
- $\mathsf{SRot} \cup \mathsf{Trans} \subseteq \mathsf{G} \leq \mathsf{Gal}$.

Then $\mathcal{M}_\mathsf{G}$ is a model of $\mathsf{KIN} + \mathsf{AxSPR}$ for which $\mathbb{W} = \mathsf{G}$.

By Theorem 5.1 (Characterisation), Theorem 5.3 (Model Construction) and Theorem 5.2 (Satisfaction), in order to determine whether a group of symmetries has to be a subgroup of one of the groups $c\mathsf{Poi}$, $c\mathsf{Eucl}$ and Gal, it is sufficient to consider its members' actions on $\mathbf{t}$:

Theorem 5.4 (Determination)**.** Let $(Q, +, \cdot, 0, 1, \leq)$ be a Euclidean field, and let G be a group satisfying $\mathsf{SRot} \cup \mathsf{Trans} \subseteq \mathsf{G} \leq \mathsf{Sym}(Q^4)$. Then

(i) For all $g \in \mathsf{G}$, $g[\mathbf{t}]$ is a line, and if $g[\mathbf{t}] = \mathbf{t}$, then $g \in \mathsf{Triv}$. $\iff$ (ii) $\mathsf{G} \leq c\mathsf{Poi}$, $\mathsf{G} \leq c\mathsf{Eucl}$ or $\mathsf{G} \leq \mathsf{Gal}$ for some positive $c \in Q$.

Our next result, Theorem 5.5 (Classification), tells us that we can classify all possible models by looking at how observers' clocks run relative to one another. Based on the difference between the time components of the w_{mk}-image of $\vec{\mathsf{t}}$ and $\vec{\mathsf{o}}$, we can decide whether observer k's clock is fast, slow or accurate relative to observer m's clock; see Figure 3. Using these notions, we can capture the following situations:

$\exists\mathsf{SlowClock}$ There are observers $m, k \in IOb$ such that

$$\left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_t - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_t\right| > 1.$$

$\exists\mathsf{FastClock}$ There are observers $m, k \in IOb$ such that

$$\left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_t - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_t\right| < 1.$$

$\exists\mathsf{MovingAccurateClock}$ There are observers $m, k \in IOb$ such that

$$\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_s \neq \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_s \text{ and } \left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_t - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_t\right| = 1.$$

$\forall\mathsf{MovingClockSlow}$ For all observers $m, k \in IOb$,

$$\text{if } \mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_s \neq \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_s, \text{ then } \left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_t - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_t\right| > 1.$$

$\forall\mathsf{MovingClockFast}$ For all observers $m, k \in IOb$,

$$\text{if } \mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_s \neq \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_s, \text{ then } \left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_t - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_t\right| < 1.$$

$\forall\mathsf{ClockAccurate}$ For all observers $m, k \in IOb$, $\left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_t - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_t\right| = 1$.

Theorem 5.5 (Classification)**.** Assume $\mathsf{KIN} + \mathsf{AxIsotropy}$. Then precisely one of the following four cases holds:

1. There exists a slow clock ($\exists\mathsf{SlowClock}$). In this case, there exists a moving observer ($\exists\mathsf{MovingIOb}$), all moving clocks are slow ($\forall\mathsf{MovingClockSlow}$), and

$$\mathbb{W} \subseteq c\mathsf{Poi} \text{ for some positive } c \in Q.$$

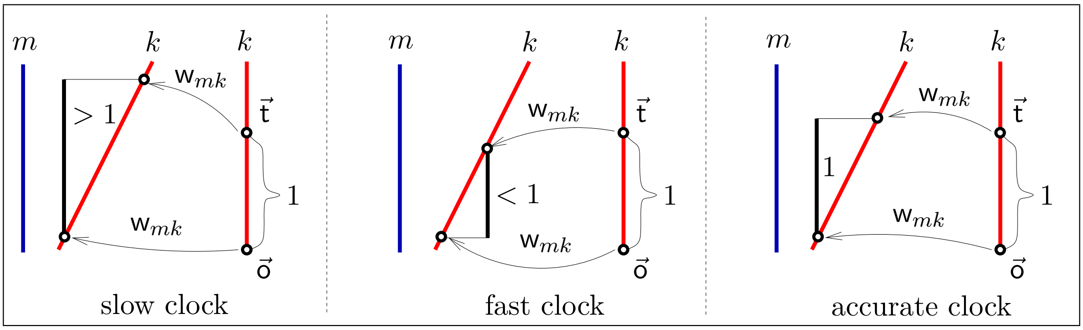


Figure 3: k's clock can be fast, slow or accurate according to m

2. There exists a fast clock ($\exists\mathsf{FastClock}$). In this case, there exists a moving observer ($\exists\mathsf{MovingIOb}$), all moving clocks are fast ($\forall\mathsf{MovingClockFast}$), and

$$\mathbb{W} \subseteq c\mathsf{Eucl} \text{ for some positive } c \in Q.$$

3. There exists a moving accurate clock ($\exists\mathsf{MovingAccurateClock}$). In this case, all clocks are accurate ($\forall\mathsf{ClockAccurate}$) and

$$\mathbb{W} \subseteq \mathsf{Gal}.$$

4. There are no moving observers ($\neg\exists\mathsf{MovingIOb}$). In this case,

$$\mathbb{W} \subseteq \mathsf{Triv}.$$

By Theorem 5.6 (Consistency), all of these situations can indeed arise.

Theorem 5.6 (Consistency)**.** The following axiom systems are all consistent (they all have models):

1. $\mathsf{KIN} + \mathsf{AxSPR} + \exists\mathsf{SlowClock}$,

2. $\mathsf{KIN} + \mathsf{AxSPR} + \exists\mathsf{FastClock}$,

3. $\mathsf{KIN} + \mathsf{AxSPR} + \exists\mathsf{MovingAccurateClock}$,

4. $\mathsf{KIN} + \mathsf{AxSPR} + \neg\exists\mathsf{MovingIOb}$.

6 Subsidiary theorems and lemmas

Because we use only a small number of basic axioms, we have a large number of intermediate lemmas to prove before we can prove our main theorems. This section is accordingly split into six subsections, each focussing on a key stage in the overall proof of our main findings. Each stage builds on its predecessor(s) and together they establish the following subsidiary theorems. Informally stated, they assert (subject to various conditions) that:

Theorem 6.1 (Observer Lines Lemma)
If ℓ is a possible worldline, then all lines of the same slope as ℓ are also possible worldlines.

Theorem 6.2 (Line-to-Line Lemma)
Each worldview transformation is a bijection taking lines to lines, planes to planes and hyperplanes to hyperplanes.

Theorem 6.3 (*tx*-Plane Lemma)
If w_{km} maps the tx-plane to itself, then it also maps the yz-plane to itself; moreover, if w_{km} is linear, there is some positive λ such that $|\mathsf{w}_{km}(\vec{p}\,)| = \lambda|\vec{p}\,|$ for all $\vec{p}$ in the yz-plane.

Theorem 6.4 (Same-Speed Lemma)
Suppose at least one observer considers h and k to be travelling with the same speed. Then w_{hk} is a κ-isometry for some κ.

Theorem 6.5 (Fundamental Lemma)
Suppose no observers move with infinite speed, and that $\mathsf{speed}_k(m) = u > 0$. Then there exists $\varepsilon > 0$ for which, given any positive $v \leq u + \varepsilon$, there is some h with $\mathsf{speed}_k(h) = v$ and $\mathsf{speed}_m(h) = \mathsf{speed}_m(k)$.

Theorem 6.6 (Main Lemma)
There exists at least one observer k and one κ for which all worldview transformations w_{mk} involving observers m who agree with k about the origin are κ-isometries.

The order of implications in the proofs that follow is:

$$\begin{tikzcd}
& & & \text{Same-Speed} \arrow[dr] & \\
\begin{matrix}\text{Observer}\\ \text{Lines}\end{matrix} \arrow[r] & \text{Line-to-Line} \arrow[r] & tx\text{-Plane} \arrow[ur] \arrow[dr] & & \text{Main} \\
& & & \text{Fundamental} \arrow[ur] &
\end{tikzcd}$$

6.1 Observer Lines Lemma

We say that a subset $\ell \subseteq Q^4$ is an ***observer line*** *for* k if there is some observer h for which $\ell = \mathsf{wl}_k(h)$, and write $\mathsf{ObLines}(k)$ for the set of k-observer lines. We say that ℓ is an *observer line* if there is some k for which it is an observer line. By AxLine, all observer lines are lines (because they are worldlines). In this section, we prove that if k can see an observer travelling along a worldline, then every other line with the same slope is also a worldline as far as k is concerned; there are none of these lines from which observers are banned.

Now suppose $\mathsf{AxEField}$ holds. If ℓ is a line and $\vec{p}$, $\vec{q}$ are distinct points in ℓ, we define its ***slope*** by

$$\mathsf{slope}(\ell) \stackrel{\text{def}}{=} \begin{cases} |\vec{p}_s - \vec{q}_s| \big/ |\vec{p}_t - \vec{q}_t| & \text{if } \vec{p}_t \neq \vec{q}_t, \\ \infty & \text{otherwise .} \end{cases}$$

Theorem 6.1 (Observer Lines Lemma)**.** Assume $\mathsf{AxEField}$, AxWvt, $\mathsf{AxRelocate}$, AxLine and $\mathsf{AxIsotropy}$. Suppose either

(a) $\mathsf{slope}(\ell) = \mathsf{slope}(\ell') \neq \infty$; or else

(b) $\mathsf{slope}(\ell) = \mathsf{slope}(\ell') = \infty$ and there exist $\vec{p} \in \ell$ and $\vec{q} \in \ell'$ whose time coordinates are equal.

Then for any observer k, we have $\ell \in \mathsf{ObLines}(k)$ iff $\ell' \in \mathsf{ObLines}(k)$. $\square$

In order to prove this result, we require various supporting lemmas (the more elementary ones are re-used in subsequent proofs). These lemmas refer to a concept we call *F-transformation* that relates the worldviews of any two observers via that of a third (see Figure 4). To illustrate the concept, suppose that I am observing two planets, k and k^*, in the night sky. From my point of view, people living on those planets would see the world quite differently, but they nonetheless see the same world I do, so I ought to be able to find some function (F) that transforms "what I think k sees" into "what I think k^* sees". From my point of view, I can say that "k^* is an 'F-transformed' version of k."

Definition 6.1.1 (F-transforms)**.** Given any bijection $F : Q^4 \to Q^4$, we say that *k^* is an* F-**transformed** *version of k according to h*, and write $k \overset{F}{\rightsquigarrow}_h k^*$ if

$$\mathsf{w}_{hk^*} = F \circ \mathsf{w}_{hk}. \tag{6.1}$$

□

Remark 6.1. Assuming AxWvt, $k \overset{\mathsf{Id}}{\rightsquigarrow}_h k^*$ is equivalent to $\mathsf{w}_{k^*k} = \mathsf{Id}$, in particular $k \overset{\mathsf{Id}}{\rightsquigarrow}_h k$; relations $k \overset{F}{\rightsquigarrow}_h k^*$ and $k^* \overset{G}{\rightsquigarrow}_h k'$ imply $k \overset{G\circ F}{\rightsquigarrow}_h k'$; and $k \overset{F}{\rightsquigarrow}_h k^*$ implies $k^* \overset{F^{-1}}{\rightsquigarrow}_h k$.

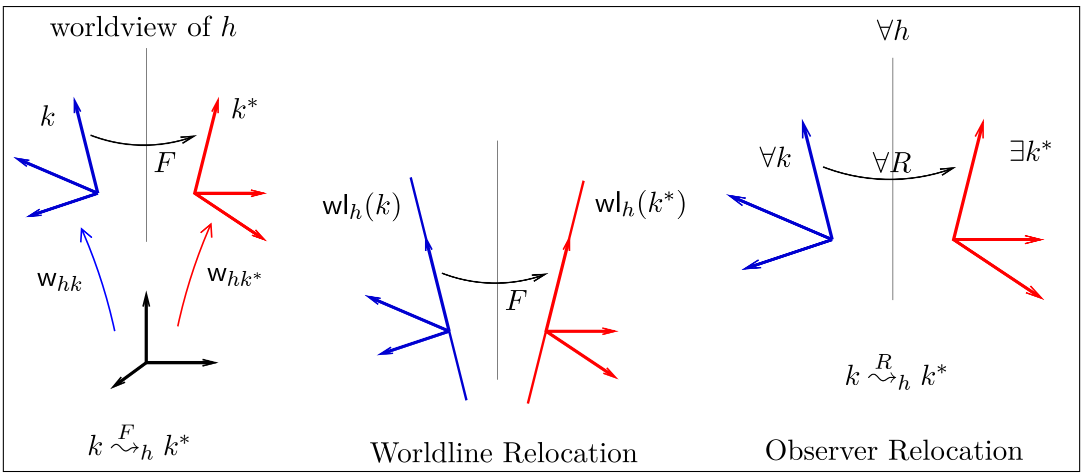


Figure 4: F-transforms (left) describe how h can transform what it considers to be k's worldview — and worldline (middle) — into k^*'s (Definition 6.1.1, Lemma 6.1.3 (Worldline Relocation)). Lemma 6.1.4 (Observer Rotation) tells us that all spatial rotations can be interpreted as F-transforms (right).

6.1.1 Supporting lemmas

Some of these initial lemmas are quite elementary, but they form the bedrock of what follows, and we need to prove them formally to ensure they definitely follow from our somewhat restricted first-order axiom set. The supporting lemmas can be informally described as follows:

Lemma 6.1.2 (WVT)
This describes various elementary properties concerning worldview transformations. We often use these results without further mention.

Lemma 6.1.3 (Worldline Relocation)
If h can F-transform k into k^*, then that transformation maps k's worldline into k^*'s.

Lemma 6.1.4 (Observer Rotation)
Every spatial rotation can be interpreted as an F-transform.

Lemma 6.1.5 (Transformed Observer Lines)
If ℓ is an observer line for k, then $\mathsf{w}_{hk}[\ell]$ is an observer line for h.

Lemma 6.1.6 (Rotated Observer Lines)
If ℓ is an observer line for k, so is any spatially rotated copy of ℓ.

Lemma 6.1.7 (Horizontal Rotation)
This is a technical lemma telling us when one pair of mutually orthogonal horizontal vectors can be spatially rotated into another (where "horizontal" means "orthogonal to the time-axis").

Lemma 6.1.8 (Same-Slope Rotation)
If two lines have the same slope and both pass through the origin, it is possible to spatially rotate one into the other.

Lemma 6.1.9 (Observer Line Intersections)
Suppose two intersecting lines have the same slope. If one of them is an observer line for k, then so is the other.

Lemma 6.1.10 (Triangulation)
Suppose $\mathbf{t}'$ is a line parallel to the time-axis, $\mathbf{t}$, and that $\vec{p}$ is not on $\mathbf{t}'$. Given any positive λ we can find lines ℓ_1 and ℓ_2 which intersect at $\vec{p}$, meet $\mathbf{t}'$ at different points, and have the same slope, λ. In other words, we can find an isosceles triangle whose base is along $\mathbf{t}'$ and vertex at $\vec{p}$, and whose equal non-base sides both have slope λ.

6.1.2 Proofs of the supporting lemmas

Lemma 6.1.2 (WVT)**.** Assume AxWvt. Then, for every $k, h, m \in IOb$,

(i) $\mathsf{wl}_k(k) = \mathbf{t}$;

(ii) $\mathsf{w}_{hk}\,[\mathsf{wl}_k(m)] = \mathsf{wl}_h(m)$;

(iii) $\mathsf{w}_{hk} : Q^4 \to Q^4$ is a bijection from Q^4 onto itself;

(iv) $\mathsf{w}_{hk}^{-1} = \mathsf{w}_{kh}$.

Proof. (i) $\mathsf{wl}_k(k) = \mathsf{w}_{kk}[\mathbf{t}] = \mathsf{Id}[\mathbf{t}] = \mathbf{t}$.

(ii) Since $\mathsf{wl}_k(m) = \mathsf{w}_{km}\,[\mathbf{t}]$, we have $\mathsf{w}_{hk}\,[\mathsf{wl}_k(m)] = \mathsf{w}_{hk}\,[\mathsf{w}_{km}\,[\mathbf{t}]] = \mathsf{w}_{hm}\,[\mathbf{t}] = \mathsf{wl}_h(m)$, as required.

(iii), (iv): It follows from $\mathsf{w}_{kh} \circ \mathsf{w}_{hk} = \mathsf{w}_{kk} = \mathsf{Id}$ and $\mathsf{w}_{hk} \circ \mathsf{w}_{kh} = \mathsf{w}_{hh} = \mathsf{Id}$ that w_{kh} and w_{hk} are mutual inverses, and hence that they are both bijections. □

Lemma 6.1.3 (Worldline Relocation)**.** Assume AxWvt, and suppose $k \overset{F}{\rightsquigarrow}_h k^*$ for some bijection $F : Q^4 \to Q^4$. Then F maps $\mathsf{wl}_h(k)$ onto $\mathsf{wl}_h(k^*)$; see Fig. 4 (middle).

Proof. Recall that $k \overset{F}{\rightsquigarrow}_h k^*$ means $\mathsf{w}_{hk^*} = F \circ \mathsf{w}_{hk}$. So

$$\mathsf{wl}_h(k^*) = \mathsf{w}_{hk^*}\,[\mathbf{t}] = (F \circ \mathsf{w}_{hk})\,[\mathbf{t}] = F\,[\mathsf{wl}_h(k)]\,.$$

□

Lemma 6.1.4 (Observer Rotation)**.** Assume $\mathsf{AxEField}$, AxWvt, $\mathsf{AxRelocate}$ and $\mathsf{AxIsotropy}$. Then given any spatial rotation $R \in \mathsf{SRot}$ and $k, h \in IOb$, there exists an observer k^* such that $k \overset{R}{\rightsquigarrow}_h k^*$; see Fig. 4 (right).

Proof. By $\mathsf{AxRelocate}$, there exists an observer h^* for which $\mathsf{w}_{hh^*} = R$. Because h and h^* are related via a spatial rotation, $\mathsf{AxIsotropy}$ tells us there exists $k^* \in IOb$ which is related to h^* the same way k is related to h, i.e. $\mathsf{w}_{h^*k^*} = \mathsf{w}_{hk}$. It follows immediately that $\mathsf{w}_{hk^*} = \mathsf{w}_{hh^*} \circ \mathsf{w}_{h^*k^*} = R \circ \mathsf{w}_{hk}$, i.e. $k \overset{R}{\rightsquigarrow}_h k^*$, as claimed. □

Lemma 6.1.5 (Transformed Observer Lines)**.** Assume AxWvt. Then $\ell \in \mathsf{ObLines}(k)$ iff $\mathsf{w}_{hk}[\ell] \in \mathsf{ObLines}(h)$.

Proof. This follows immediately from Lemma 6.1.2 (WVT), since all k-observer lines are worldlines. □

Lemma 6.1.6 (Rotated Observer Lines)**.** Assume $\mathsf{AxEField}$, AxWvt, $\mathsf{AxRelocate}$ and $\mathsf{AxIsotropy}$. If $\ell \in \mathsf{ObLines}(k)$ and $R \in \mathsf{SRot}$ is any spatial rotation, then $R[\ell] \in \mathsf{ObLines}(k)$.

Proof. Choose $h \in IOb$ such that $\ell = \mathsf{wl}_k(h)$. By Lemma 6.1.4 (Observer Rotation), there is some $h^* \in IOb$ for which $h \overset{R}{\rightsquigarrow}_k h^*$, i.e. $\mathsf{w}_{kh^*} = R \circ \mathsf{w}_{kh}$. By Lemma 6.1.3 (Worldline Relocation), we have that $\mathsf{wl}_k(h^*) = R[\mathsf{wl}_k(h)] = R[\ell]$, and this worldline is in $\mathsf{ObLines}(k)$, as required. □

Lemma 6.1.7 (Horizontal Rotation)**.** Let $(Q, +, \cdot, 0, 1, \leq)$ be an ordered field and suppose $\vec{p}_1, \vec{q}_1, \vec{p}_2, \vec{q}_2 \in Q^4$ satisfy:

(a) $\vec{p}_1$ and $\vec{p}_2$ have the same length, as do $\vec{q}_1$ and $\vec{q}_2$: $|\vec{p}_1|^2 = |\vec{p}_2|^2$ and $|\vec{q}_1|^2 = |\vec{q}_2|^2$;

(b) $\vec{p}_1$ and $\vec{q}_1$ are horizontal and mutually orthogonal: $\vec{p}_1 \cdot \vec{\mathsf{t}} = \vec{q}_1 \cdot \vec{\mathsf{t}} = \vec{p}_1 \cdot \vec{q}_1 = 0$; and

(c) $\vec{p}_2$ and $\vec{q}_2$ are horizontal and mutually orthogonal: $\vec{p}_2 \cdot \vec{\mathsf{t}} = \vec{q}_2 \cdot \vec{\mathsf{t}} = \vec{p}_2 \cdot \vec{q}_2 = 0$.

Then there exists a spatial rotation $R \in \mathsf{SRot}$ such that $R(\vec{p}_1) = \vec{p}_2$ and $R(\vec{q}_1) = \vec{q}_2$; see the left-hand side of Figure 5.

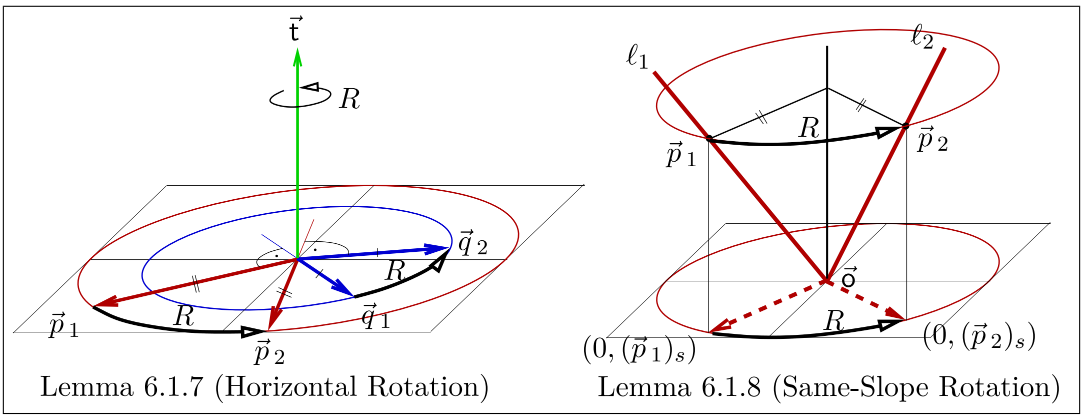


Figure 5: Illustrations for Lemma 6.1.7 (Horizontal Rotation) and Lemma 6.1.8 (Same-Slope Rotation).

Proof. Consider the linear map that takes $\alpha\vec{\mathsf{t}} + \beta\vec{p}_1 + \gamma\vec{q}_1$ to $\alpha\vec{\mathsf{t}} + \beta\vec{p}_2 + \gamma\vec{q}_2$. It is easy to see that this map is a linear Euclidean isometry between two subspaces of Q^4 which are each at most three-dimensional. Hence, by the refinement of Witt's theorem [30, Thm 234.1, p.234] there is an extension $R : Q^4 \to Q^4$ which is a linear Euclidean isometry with determinant 1. This R must be a spatial rotation, because $R(\vec{\mathsf{t}}) = \vec{\mathsf{t}}$. $\square$

Lemma 6.1.8 (Same-Slope Rotation)**.** Let $(Q, +, \cdot, 0, 1, \leq)$ be a Euclidean field. Assume ℓ_1 and ℓ_2 are lines such that $\mathsf{slope}(\ell_1) = \mathsf{slope}(\ell_2)$ and $\vec{\mathsf{o}} \in \ell_1 \cap \ell_2$. Then there exists $R \in \mathsf{SRot}$ such that $R[\ell_1] = \ell_2$.

Proof. Let $\vec{p}_1 \in \ell_1$ and $\vec{p}_2 \in \ell_2$ be such that $\vec{p}_1 \neq \vec{o} \neq \vec{p}_2$ and $(\vec{p}_1)_t = (\vec{p}_2)_t$, see the right-hand side of Figure 5. Then $|(0,(\vec{p}_1)_s)|^2 = |(0,(\vec{p}_2)_s)|^2$. Taking $\vec{q}_1 = \vec{q}_2 = \vec{o}$, Lemma 6.1.7 (Horizontal Rotation) now tells us there exists a spatial rotation R that takes $(0,(\vec{p}_1)_s)$ to $(0,(\vec{p}_2)_s)$ and leaves $\vec{o}$ fixed. Since spatial rotations leave time coordinates unchanged, this R takes $\vec{p}_1$ to $\vec{p}_2$, and since it also fixes the origin it must take ℓ_1 to ℓ_2. □

Lemma 6.1.9 (Observer Line Intersections)**.** Assume $\mathsf{AxEField}$, AxWvt, AxLine, $\mathsf{AxRelocate}$, and $\mathsf{AxIsotropy}$. If two lines ℓ_1, ℓ_2 intersect one another and have equal slope, then for any $k \in IOb$ we have $\ell_1 \in \mathsf{ObLines}(k)$ iff $\ell_2 \in \mathsf{ObLines}(k)$.

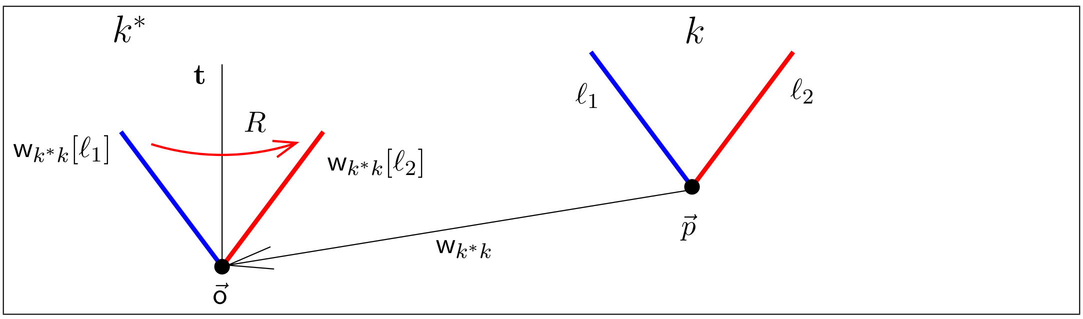


Figure 6: Illustration for the proof of Lemma 6.1.9 (Observer Line Intersections).

Proof. Let $\vec{p}$ be the point of intersection of ℓ_1 and ℓ_2, and let T be the translation taking $\vec{p}$ to the origin, $\vec{o}$. By $\mathsf{AxRelocate}$, there exists some $k^* \in IOb$ such that $\mathsf{w}_{k^*k} = T$; see Figure 6.

Note first that the images of ℓ_1 and ℓ_2 under w_{k^*k} are lines of equal slope because $\mathsf{w}_{k^*k} = T$ is a translation, and translations map lines to lines and leave slopes unchanged. Moreover, both of these lines pass through $T(\vec{p}) = \vec{o}$, so Lemma 6.1.8 (Same-Slope Rotation) tells us there exists a spatial rotation R taking $\mathsf{w}_{k^*k}[\ell_1]$ to $\mathsf{w}_{k^*k}[\ell_2]$.

The claim now follows. For suppose ℓ_1 is a k-observer line; we have to show that ℓ_2 is also a k-observer line. Since $\mathsf{w}_{k^*k}[\ell_1] \in \mathsf{ObLines}(k^*)$ by Lemma 6.1.5 (Transformed Observer Lines), it follows that $\mathsf{w}_{k^*k}[\ell_2] \in \mathsf{ObLines}(k^*)$ as well, by Lemma 6.1.6 (Rotated Observer Lines). Applying Lemma 6.1.5 (Transformed Observer Lines) in the opposite direction now tells us that $\ell_2 \in \mathsf{ObLines}(k)$, as required.

The converse follows by symmetry. □

Lemma 6.1.10 (Triangulation)**.** Assume $\mathsf{AxEField}$. Let $\mathbf{t}'$ be a line parallel to the time-axis and let $\vec{p}$ be any point not on $\mathbf{t}'$. Given any positive $\lambda \in Q$, there exist lines ℓ_1, ℓ_2 with

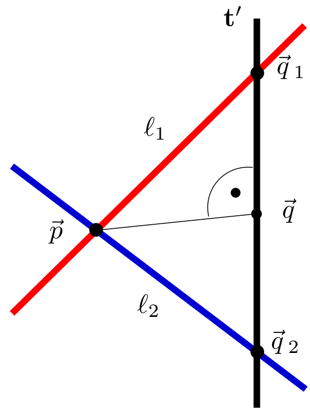
(i) $\mathsf{slope}(\ell_1) = \mathsf{slope}(\ell_2) = \lambda$,

(ii) $\vec{p} \in \ell_1 \cap \ell_2$,

(iii) $\ell_1 \cap \mathbf{t}' \neq \varnothing$,

(iv) $\ell_2 \cap \mathbf{t}' \neq \varnothing$,

(v) $\ell_1 \cap \ell_2 \cap \mathbf{t}' = \varnothing$.



Proof. Let $\vec{q} \in \mathbf{t}'$ be the point on $\mathbf{t}'$ with $\vec{q}_t = \vec{p}_t$. We know that $\vec{p}_s \neq \vec{q}_s$ because $\vec{p} \notin \mathbf{t}'$. Consider the points

$$\vec{q}_1 := \vec{q} + (|\vec{p}_s - \vec{q}_s| / \lambda, 0, 0, 0) \quad \text{and} \quad \vec{q}_2 := \vec{q} - (|\vec{p}_s - \vec{q}_s| / \lambda, 0, 0, 0)$$

and let ℓ_1 be the line passing through $\vec{p}$ and $\vec{q}_1$, and ℓ_2 the line passing through $\vec{p}$ and $\vec{q}_2$. Then direct calculation shows that ℓ_1 and ℓ_2 have the required properties. $\square$

6.1.3 Main proof

We now complete the proof of Theorem 6.1 (Observer Lines Lemma).

We use the word *plane* in the usual Euclidean sense to mean a 2-dimensional slice of Q^4, and refer to 3-dimensional 'slices' as *hyperplanes*. Formally, a subset $P \subseteq Q^4$ is a ***plane*** iff there are linearly independent vectors $\vec{v}, \vec{w} \neq \vec{\mathsf{o}} \in Q^4$ and a point $\vec{p} \in Q^4$, such that $P = \{\vec{p} + \lambda\vec{v} + \mu\vec{w} \ : \ \lambda, \mu \in Q\}$ (hyperplanes are defined analogously). By $\mathsf{AxEField}$, the usual properties of Euclidean planes hold. In particular, a plane P can be specified by giving a line $\ell \subseteq P$ and a point $\vec{p} \in P \setminus \ell$, or three distinct non-collinear points $\vec{p}, \vec{q}, \vec{r} \in P$, or two distinct but intersecting lines in P. Moreover, given a line $\ell \subseteq P$ and a point $\vec{p} \in P \setminus \ell$, there is exactly one line ℓ_p through $\vec{p}$ that is parallel to ℓ (indeed, if we assume $\mathsf{AxEField}$, the way in which we have defined *line* and *plane* allows us to uniquely determine ℓ_p in the usual way once $\vec{p}$ and ℓ are specified).

Proof of Theorem 6.1 (Observer Lines Lemma). Let ℓ, ℓ' be lines of equal slope, i.e. $\mathsf{slope}(\ell) = \mathsf{slope}(\ell')$. If $\ell = \ell'$, there is nothing to prove, so assume that $\ell \neq \ell'$. Also, if $\mathsf{slope}(\ell) = \mathsf{slope}(\ell') = 0$, then ℓ and ℓ' are both parallel to the time-axis, and it follows easily from $\mathsf{AxRelocate}$ that $\ell, \ell' \in \mathsf{ObLines}(k)$.

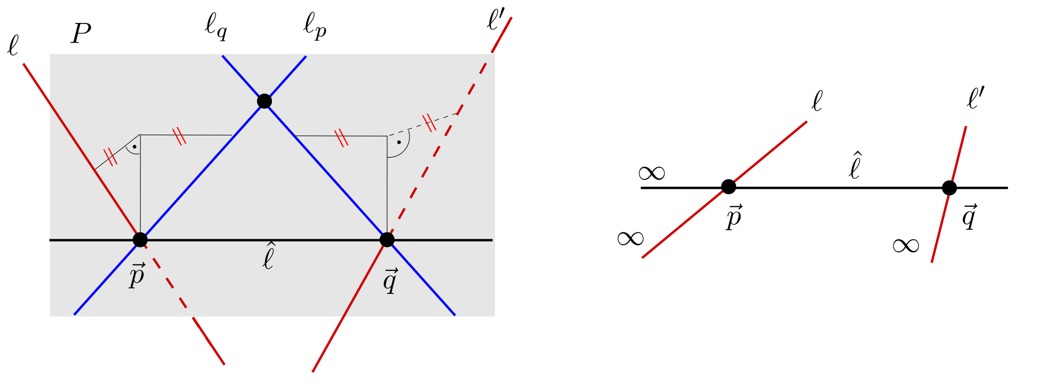
Suppose, therefore, that $\mathsf{slope}(\ell) = \mathsf{slope}(\ell') \neq 0$.

Note first that there exist $\vec{p}, \vec{q} \in Q^4$ such that

$$\vec{p} \in \ell, \quad \vec{q} \in \ell', \quad \vec{p} \neq \vec{q}, \quad \text{and} \quad \vec{p}_t = \vec{q}_t.$$

This is true by assumption for case (b), where $\mathsf{slope}(\ell) = \mathsf{slope}(\ell') = \infty$, and it is easy to see that such $\vec{p}, \vec{q}$ also exist in case (a) where $\mathsf{slope}(\ell) = \mathsf{slope}(\ell')$ is finite.[10]

Let $\hat{\ell}$ be the line containing $\vec{p}$ and $\vec{q}$. Because $\vec{p}, \vec{q}$ have the same time coordinate, $\mathsf{slope}(\hat{\ell}) = \infty$; see Figure 7.



(a) finite slopes (b) infinite slopes

Figure 7: Illustration for the proof of Theorem 6.1 (Observer Lines Lemma)

We now consider cases (a) and (b) in turn.

Case (a): finite slopes. By assumption, $0 < \mathsf{slope}(\ell) = \mathsf{slope}(\ell') \neq \infty$ and $\mathsf{slope}(\hat{\ell}) = \infty$. Let P be the plane containing $\hat{\ell}$ and parallel to $\mathbf{t}$.[11]

Let $\mathbf{t}_p$ be the line parallel to $\mathbf{t}$ which passes through $\vec{p}$, and notice that this line lies in P. Choose any point $\vec{p}\,' \in P \setminus \mathbf{t}_p$ and let $\lambda = \mathsf{slope}(\ell) = \mathsf{slope}(\ell')$. Then Lemma 6.1.10 (Triangulation) tells us that we can find two distinct lines which pass through $\vec{p}\,'$, lie in P (because they meet both $\vec{p}\,'$ and $\mathbf{t}_p$), and have slope λ. Applying the translation taking $\vec{p}\,'$ to $\vec{p}$, the images of those two lines will still lie in P and still have slope λ, but will intersect at $\vec{p}$. Similarly, we can find two distinct lines

[10] Pick any point $\vec{p}$ on ℓ that isn't on ℓ' and consider the 'horizontal time slice' containing it; because ℓ' has finite slope, it must also pass through this time slice. Take $\vec{q}$ to be the corresponding point of intersection on ℓ'.

[11] P is parallel to $\mathbf{t}$ iff P contains a line parallel to $\mathbf{t}$.

of slope λ which lie in P and pass through $\vec{q}$. Pick one of the lines passing through $\vec{q}$, and call it ℓ_q. Since the two lines through $\vec{p}$ are distinct, they cannot both be parallel to ℓ_q — let ℓ_p be one that isn't. Since ℓ_p and ℓ_q are non-parallel lines lying in the same plane, they must intersect.

The claim now follows. For suppose $\ell \in \mathsf{ObLines}(k)$. Then ℓ and ℓ_p are lines of equal slope which intersect at $\vec{p}$, so Lemma 6.1.9 (Observer Line Intersections) tells us that ℓ_p is also in $\mathsf{ObLines}(k)$, whence (applying the same argument twice more) so are ℓ_q (because it meets ℓ_p) and ℓ' (since it meets ℓ_q).

Case (b): infinite slopes. If $\mathsf{slope}(\ell) = \infty$, then ℓ and $\hat{\ell}$ are two lines of infinite slope which intersect at $\vec{p}$. Likewise, ℓ' and $\hat{\ell}$ are lines of infinite slope that intersect at $\vec{q}$. As before it now follows by Lemma 6.1.9 (Observer Line Intersections) that

$$\ell \in \mathsf{ObLines}(k) \iff \hat{\ell} \in \mathsf{ObLines}(k) \iff \ell' \in \mathsf{ObLines}(k).$$

In both cases, therefore, we have $\ell \in \mathsf{ObLines}(k) \iff \ell' \in \mathsf{ObLines}(k)$, as required. □

6.2 Line-to-Line Lemma

Theorem 6.2 (Line-to-Line Lemma)**.** Assume $\mathsf{AxEField}$, AxWvt, AxLine, $\mathsf{AxIsotropy}$, $\mathsf{AxRelocate}$ and $\exists\mathsf{MovingIOb}$. Then given any $k, h \in IOb$, the worldview transformation w_{hk} is a bijection that takes lines to lines, planes to planes, and hyperplanes to hyperplanes.

6.2.1 Supporting lemmas

A number of the supporting lemmas refer to the concept of an *observer line triad*:

Definition 6.2.1 (Observer Line Triads)**.** If $\ell_1, \ell_2, \ell_3 \in \mathsf{ObLines}(k)$ are three (necessarily coplanar) lines, each pair of which intersect in a point, and whose pairwise intersections are not collinear, we shall call the set $\{\ell_1, \ell_2, \ell_3\}$ an *observer line triad for k*, or simply a *k-triad.* □

The lemmas can be described informally as follows:

Lemma 6.2.4 (Speed)
: Speeds are well-defined, and the terms *at rest* and *in motion* have their expected meanings.

Lemma 6.2.5 (Triads)
: If one observer considers that three worldlines form a triad, all other observers agree.

Lemma 6.2.6 (Plane-to-Plane)
Suppose plane P contains a k-triad whose slopes are either all finite or else all infinite. Then $\mathsf{w}_{hk}[P]$ is contained in a plane.

Lemma 6.2.7 (Infinite Speeds $\Rightarrow$ Lines are Observer Lines)
If infinite speeds occur, then all lines are observer lines.

6.2.2 Proofs of the supporting lemmas

Definition 6.2.2. Suppose $\mathsf{AxEField}$ and AxLine holds. If $\ell = \mathsf{wl}_k(h)$, we call the slope, $\mathsf{slope}(\ell)$, of line ℓ the ***speed*** *of h according to k*, i.e.

$$\mathsf{speed}_k(h) \stackrel{\text{def}}{=} \mathsf{slope}(\mathsf{wl}_k(h)).$$

Definition 6.2.3. Recall that observer $k \in IOb$ is moving according to observer $m \in IOb$ iff $\mathsf{w}_{mk}(\vec{\mathsf{t}})_s \neq \mathsf{w}_{mk}(\vec{\mathsf{o}})_s$ and at rest according to m otherwise. We say that *observer $k \in IOb$* ***is moving instantaneously according to*** *observer m* iff $\mathsf{w}_{mk}(\vec{\mathsf{t}})_t = \mathsf{w}_{mk}(\vec{\mathsf{o}})_t$.

Lemma 6.2.4 (Speed)**.** Assume AxWvt, $\mathsf{AxEField}$ and AxLine. Then for every $m, k \in IOb$, $\mathsf{speed}_m(k)$ is well-defined, and

- k is at rest according to m iff $\mathsf{speed}_m(k) = 0$,
- k is moving according to m iff $\mathsf{speed}_m(k) \neq 0$, and
- k is moving instantaneously according to m iff $\mathsf{speed}_m(k) = \infty$.

Proof. By $\mathsf{AxEField}$ and AxLine, it follows that $\mathsf{speed}_m(k)$ is unambiguously defined for all k and m. The proof is straightforward after noticing that $\mathsf{w}_{mk}(\vec{\mathsf{t}}) \neq \mathsf{w}_{mk}(\vec{\mathsf{o}})$ which holds because w_{mk} is a bijection by Lemma 6.1.2 (WVT). $\square$

Lemma 6.2.5 (Triads)**.** Suppose $\mathsf{AxEField}$, AxWvt, AxLine. Let $k, h \in IOb$. If $T = \{\ell_1, \ell_2, \ell_3\}$ is a k-triad, then $\mathsf{w}_{hk}[T] := \{\mathsf{w}_{hk}[\ell_1], \mathsf{w}_{hk}[\ell_2], \mathsf{w}_{hk}[\ell_3]\}$ is an h-triad.

Proof. Each ℓ_i is a k-observer line, so by Lemma 6.1.5 (Transformed Observer Lines), each $\ell_i' = \mathsf{w}_{hk}[\ell_i]$ is an h-observer line (and hence a line). Because w_{hk} is a bijection, we know that any two of the lines in $\mathsf{w}_{hk}[T]$ has non-empty intersection, and that they have three distinct pairwise intersections in total. It follows that the three lines are coplanar and that their three pairwise intersection points are not collinear. That is, $\mathsf{w}_{hk}[T]$ is an h-triad as claimed. $\square$

Lemma 6.2.6 (Plane-to-Plane)**.** Assume AxEField, AxWvt, AxLine, AxRelocate and AxIsotropy. Choose $k, h \in IOb$, let P be a plane which contains a k-triad $\{\ell_1, \ell_2, \ell_3\}$, and suppose that the slopes of these lines are either all finite, or else all infinite. Then $\mathsf{w}_{hk}[P]$ is contained in a plane.

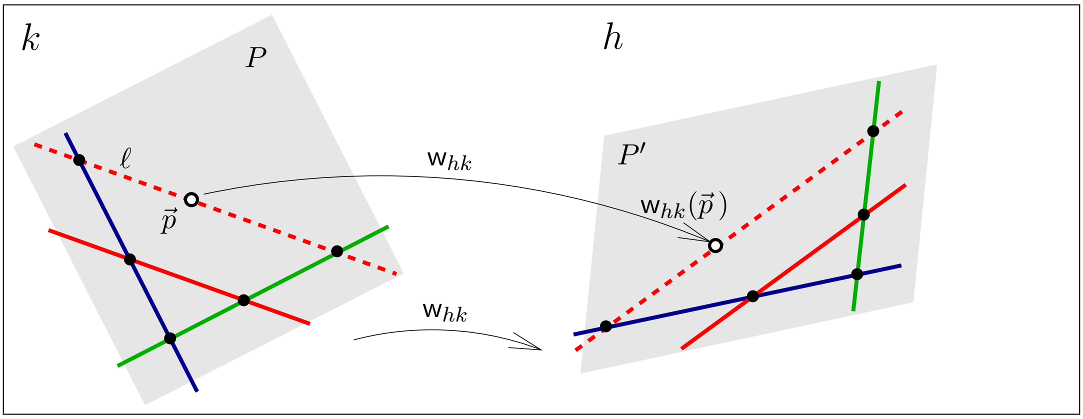


Figure 8: Illustration for the proof of Lemma 6.2.6 (Plane-to-Plane)

Proof. According to Lemma 6.2.5 (Triads), the lines $\mathsf{w}_{hk}[\ell_i]$ $(i = 1, 2, 3)$ form an h-triad. We can therefore define P', the plane spanned by this triad. We will prove that $\mathsf{w}_{hk}[P] \subseteq P'$.

Choose any $\vec{p} \in P$. If $\vec{p}$ lies on any of the lines ℓ_i, then the conclusion $\mathsf{w}_{hk}(\vec{p}\,) \in P'$ is trivial. Suppose, then, that $\vec{p}$ does not lie on any of these lines. Because the lines form a triad we can draw a line ℓ through $\vec{p}$ which is parallel to one of the lines (wlog, ℓ_1) and which intersects the other two lines (ℓ_2 and ℓ_3) in distinct points.

We claim that $\ell \in \mathsf{ObLines}(k)$. If all three lines have finite slope, this follows from Theorem 6.1 (Observer Lines Lemma) because ℓ and ℓ_1 have equal (hence finite) slopes and ℓ_1 is a k-observer line. On the other hand, if all three lines (and hence also ℓ) have infinite slope, this means there exist t_1, t_2 and t_3 such that all points on ℓ_i $(i = 1, 2, 3)$ have time component t_i. But we know that the lines intersect one another, so we must have $t_1 = t_2 = t_3$. Since ℓ lies in the plane spanned by these lines it follows that points on ℓ share the same time component as points on ℓ_1, and we can again apply Theorem 6.1 (Observer Lines Lemma) to ℓ and ℓ_1 to deduce that $\ell \in \mathsf{ObLines}(k)$.

As claimed, therefore, ℓ is a k-observer line. Therefore, ℓ, ℓ_2 and ℓ_3 form a k-triad and Lemma 6.2.5 (Triads) tells us that $\mathsf{w}_{hk}[\ell]$, $\mathsf{w}_{hk}[\ell_2]$ and $\mathsf{w}_{hk}[\ell_3]$ form an h-triad. It follows that $\mathsf{w}_{hk}[\ell]$ lies in the same plane as $\mathsf{w}_{hk}[\ell_2]$ and $\mathsf{w}_{hk}[\ell_3]$, i.e. P', and hence $\mathsf{w}_{hk}(\vec{p}\,) \in \mathsf{w}_{hk}[\ell] \subseteq P'$, as required. $\square$

The following formula says that instantaneously moving observers exists.

$\exists\infty\mathsf{Speed}$ There are observers $m, k \in IOb$ such that $\mathsf{w}_{mk}(\vec{\mathsf{o}})_t = \mathsf{w}_{mk}(\vec{\mathsf{t}})_t$.

Lemma 6.2.7 (Infinite Speeds $\Rightarrow$ Lines are Observer Lines)**.** Assume AxEField, AxWvt, AxLine, AxRelocate, AxIsotropy and $\exists\infty$Speed. Then for any observer, every line is an observer line.

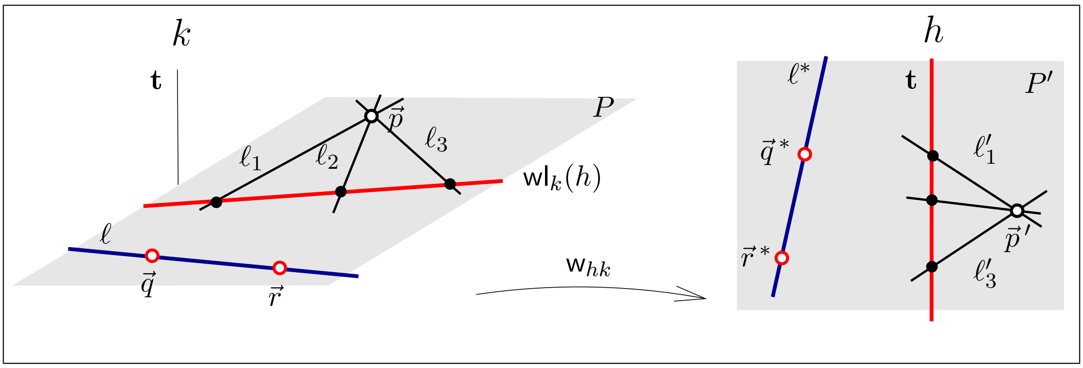


Figure 9: Illustration for the proof of Lemma 6.2.7 (Infinite Speeds $\Rightarrow$ Lines are Observer Lines)

Proof. Choose $k, h \in IOb$ such that $\mathsf{speed}_k(h) = \infty$, and recall that this means that $\mathsf{slope}(\mathsf{wl}_k(h)) = \infty$. Thus, there exists some $t \in Q$ such that every point on $\mathsf{wl}_k(h)$ has time component t. Let P be any 'horizontal' plane containing $\mathsf{wl}_k(h)$, i.e. all points in P have this same time component t. Then every line in P is in $\mathsf{ObLines}(k)$ by Theorem 6.1 (Observer Lines Lemma) because every line in P is of slope ∞.

Choose $\vec{p} \in P \setminus \mathsf{wl}_k(h)$, and notice that the plane P is determined by $\vec{p}$ and $\mathsf{wl}_k(h)$. It follows from Lemma 6.2.6 (Plane-to-Plane) that $\mathsf{w}_{hk}[P]$ is contained in a plane containing both $\mathsf{w}_{hk}(\vec{p})$ and $\mathsf{w}_{hk}[\mathsf{wl}_k(h)]$. In other words, if we define $\vec{p}^{\,\prime} = \mathsf{w}_{hk}(\vec{p})$, observe that $\mathsf{w}_{hk}[\mathsf{wl}_k(h)] = \mathbf{t}$, and define P' to be the plane generated by $\vec{p}^{\,\prime}$ and $\mathbf{t}$, then $\mathsf{w}_{hk}[P] \subseteq P'$.

We will show first that the reverse inclusion also holds, so that $\mathsf{w}_{hk}[P]$ is the whole of P'. To this end, choose three lines ℓ_i ($i = 1, 2, 3$) in P which pass through $\vec{p}$ and whose intersections with $\mathsf{wl}_k(h)$ are three distinct points; as observed above, these are all k-observer lines. Thus, if we define, for each $i = 1, 2, 3$, $\ell'_i := \mathsf{w}_{hk}[\ell_i]$ then $\ell'_1, \ell'_2, \ell'_3$ and $\mathbf{t}$ ($= \mathsf{w}_{hk}[\mathsf{wl}_k(h)]$) are all h-observer lines in P'. Since w_{hk} is a bijection by Lemma 6.1.2 (WVT), all four of these lines are distinct and moreover, each ℓ'_i passes through $\vec{p}\,'$, and they meet $\mathbf{t}$ in three distinct points.

Since at most one of the lines ℓ'_i can have infinite slope (and $\mathsf{slope}(\mathbf{t}) = 0$), we have therefore shown that there exists in P' a k-triad of observer lines, all with finite slope. By Lemma 6.2.6 (Plane-to-Plane), it follows that $\mathsf{w}_{kh}[P'] \subseteq P$, and hence $P' \subseteq \mathsf{w}_{hk}[P]$. Thus, $\mathsf{w}_{hk}[P] = P'$, as claimed.

Now we will prove that every line in P' is in $\mathsf{ObLines}(h)$. Let $\ell^* \subseteq P'$ be a line and let $\vec{q}\,^*, \vec{r}\,^*$ be two distinct points on ℓ^*. Then $\vec{q} := \mathsf{w}_{kh}(\vec{q}\,^*)$, $\vec{r} := \mathsf{w}_{kh}(\vec{r}\,^*)$ are two distinct points in P because $\mathsf{w}_{kh}[P'] \subseteq P$ and w_{kh} is a bijection. Let ℓ be the line connecting $\vec{q}$ and $\vec{r}$. Then ℓ lies in P, and must therefore be in $\mathsf{ObLines}(k)$. Since $\mathsf{w}_{hk}[\ell] = \ell^*$, it follows by Lemma 6.1.5 (Transformed Observer Lines) that $\ell^* \in \mathsf{ObLines}(h)$ as claimed.

Now we use the fact that $\mathbf{t} \subseteq P'$ to prove that *every* line is in $\mathsf{ObLines}(h)$. Let ℓ be an arbitrary line. Then there is some $\ell^* \subseteq P'$ which has the same slope as ℓ because $\mathbf{t} \subseteq P'$ and therefore lines of every positive slope occur in P' by Lemma 6.1.10 (Triangulation), while if $\mathsf{slope}(\ell) = 0$ we can take $\ell^* = \mathbf{t}$, and if $\mathsf{slope}(\ell) = \infty$ we can take ℓ^* to be the line joining $\vec{p}\,'$ to $((\vec{p}\,')_t, \vec{0})$. Moreover, by using translations 'up or down' the time-axis as necessary, ℓ^* can be chosen such that there are $\vec{p} \in \ell$, $\vec{q} \in \ell^*$ such that $\vec{p}_t = \vec{q}_t$. We know that $\ell^* \in \mathsf{ObLines}(h)$ because every line in P' is in $\mathsf{ObLines}(h)$. But now $\ell \in \mathsf{ObLines}(h)$ by Theorem 6.1 (Observer Lines Lemma). So $\mathsf{ObLines}(h)$ is the set of all lines, as claimed.

Finally, it is easy to see that because $\mathsf{ObLines}(h)$ is the set of all lines for one observer h, the same holds for every other observer m. For suppose ℓ' is a line, and choose distinct points $\vec{p}\,', \vec{q}\,' \in \ell'$. By Lemma 6.1.2 (WVT), the points $\vec{p} := \mathsf{w}_{hm}(\vec{p}\,')$ and $\vec{q} := \mathsf{w}_{hm}(\vec{q}\,')$ are again distinct, so they define a line ℓ. As we've just seen, ℓ must be an h-observer line. It follows from Lemma 6.1.5 (Transformed Observer Lines) that $\mathsf{w}_{mh}[\ell]$ is an m-observer line, and hence a line. This means that ℓ' and $\mathsf{w}_{mh}[\ell]$ are both lines passing through the two points $\vec{p}\,' \neq \vec{q}\,'$, so they must be the same line. In other words, $\ell' = \mathsf{w}_{mh}[\ell] \in \mathsf{ObLines}(m)$, as claimed. $\square$

6.2.3 Main proof

We now complete the proof of Theorem 6.2 (Line-to-Line Lemma).

Definition 6.2.8 (Observer Planes)**.** Whenever a plane P contains at least one k-observer line, we shall say that P is an *observer plane for k*, or a k-observer plane. We write $\mathsf{ObPlanes}(k)$ for the set of all k-observer planes. $\square$

Proof of Theorem 6.2 (Line-to-Line Lemma). We have already observed that every worldview transformation w_{kh} is a bijection; we will show first that they also take lines to lines.

Suppose m, m' are observers in motion relative to one another, i.e. $\mathsf{speed}_m(m') > 0$ — such observers exist by $\exists\mathsf{MovingIOb}$ and Lemma 6.2.4 (Speed). There are two cases to consider, depending on whether $\mathsf{speed}_m(m')$ can or cannot be infinite.

(Case 1: $\exists\infty\mathsf{Speed}$): If m, m' can be chosen such that $\mathsf{speed}_m(m') = \infty$, then Lemma 6.2.7 (Infinite Speeds $\Rightarrow$ Lines are Observer Lines) tells us that all lines belong to $\mathsf{ObLines}(h)$ and we know that w_{kh} takes observer lines to observer lines (which are again lines). So in this case, the result is immediate.

(Case 2: $\neg\exists\infty\mathsf{Speed}$): Assume, therefore, that all observers move with finite speed relative to one another (so that, given any observer o and $\ell \in \mathsf{ObLines}(o)$, we have $\mathsf{slope}(\ell) \neq \infty$); in particular, $0 < \mathsf{speed}_m(m') \neq \infty$. Our proof will be given in four stages; we will show that

(1) if a plane P contains a k-triad, then $\mathsf{w}_{hk}[P]$ is again a plane;

(2) that for every observer o there is some $\ell \in \mathsf{ObLines}(o)$ for which $\mathsf{slope}(\ell) \neq 0$;

(3) if $P \in \mathsf{ObPlanes}(k)$ there exists a k-triad lying entirely within P. Items (1) and (3) imply that w_{hk} maps k-observer planes to h-observer planes.

(4) Finally, we use this information to show that every line can be obtained as the intersection of two k-observer planes — since the images of these planes intersect in a line, the result then follows.

(1) We prove that if a plane P contains a k-triad, then $\mathsf{w}_{hk}[P]$ is a plane. Let $\{\ell_1, \ell_2, \ell_3\}$ be a k-triad contained in P, and for each $i = 1, 2, 3$ define $\ell'_i := \mathsf{w}_{hk}[\ell_i]$. Because all observer lines are assumed to have finite slopes, Lemma 6.2.6 (Plane-to-Plane) tells us that $\mathsf{w}_{hk}[P] \subseteq P'$, where P' is the plane generated by $\{\ell'_1, \ell'_2, \ell'_3\}$. Since, by Lemma 6.2.5 (Triads), $\{\ell'_1, \ell'_2, \ell'_3\}$ is likewise an h-triad contained in P' and comprising finite-slope lines, we can again apply Lemma 6.2.6 (Plane-to-Plane) to deduce that $\mathsf{w}_{kh}[P'] \subseteq P$. Consequently, $\mathsf{w}_{hk}[P] = P'$, and $\mathsf{w}_{hk}[P]$ is a plane as claimed.

(2) Next we show that for every observer o there is some $\ell \in \mathsf{ObLines}(o)$ for which $\mathsf{slope}(\ell) \neq 0$. To this end, let ℓ' be the line parallel to $\mathsf{wl}_m(m')$ which passes through the origin $\vec{\mathsf{o}}$, and note that this line cannot be the time-axis (which has slope 0). Since $\mathsf{wl}_m(m')$ is an m-observer line, so is ℓ' (by Theorem 6.1 (Observer Lines Lemma)). It follows that ℓ' and $\mathbf{t} = \mathsf{wl}_m(m)$ are non-identical intersecting m-observer lines, whence $\mathsf{w}_{om}[\ell']$ and $\mathsf{w}_{om}[\mathbf{t}]$ are non-identical intersecting o-observer

lines. If these both had zero slope, they would be the same line. So at least one of them has non-zero slope and hence can be taken to be ℓ.

(3) Now we prove that for every k, if $P \in \mathsf{ObPlanes}(k)$ there exists a k-triad lying entirely in P. Suppose $P \in \mathsf{ObPlanes}(k)$, and choose some k-observer line $\ell = \mathsf{wl}_k(h) \subseteq P$ and some $\vec{p} \in P \setminus \ell$, see Figure 10. Transforming to h's worldview we have $\mathsf{w}_{hk}[\ell] = \mathsf{w}_{hk}[\mathsf{wl}_k(h)] = \mathsf{wl}_h(h) = \mathbf{t}$ and $\vec{p}\,' := \mathsf{w}_{hk}(\vec{p}) \notin \mathbf{t}$. By (2), we know there is some $\ell' \in \mathsf{ObLines}(h)$ for which $\mathsf{slope}(\ell') \neq 0$, and by assumption $\mathsf{slope}(\ell') \neq \infty$. Thus, by Lemma 6.1.10 (Triangulation) there exist lines ℓ'_1, ℓ'_2 passing through $\vec{p}\,'$ which have the same slope as ℓ', such that $\{\mathbf{t}, \ell'_1, \ell'_2\}$ is a k-triad (see Figure 10), and we know that $\ell'_1, \ell'_2 \in \mathsf{ObLines}(h)$ by Theorem 6.1 (Observer Lines Lemma). Taking $\ell_1 := \mathsf{w}_{kh}[\ell'_1]$ and $\ell_2 := \mathsf{w}_{kh}[\ell'_2]$, and recalling that $\mathsf{w}_{kh}[\mathbf{t}] = \ell$, it follows that all three lines are k-observer lines, and together they form a k-triad lying entirely within P because their pairwise intersections comprise the point $\vec{p} \notin \ell$ together with two distinct points on ℓ.

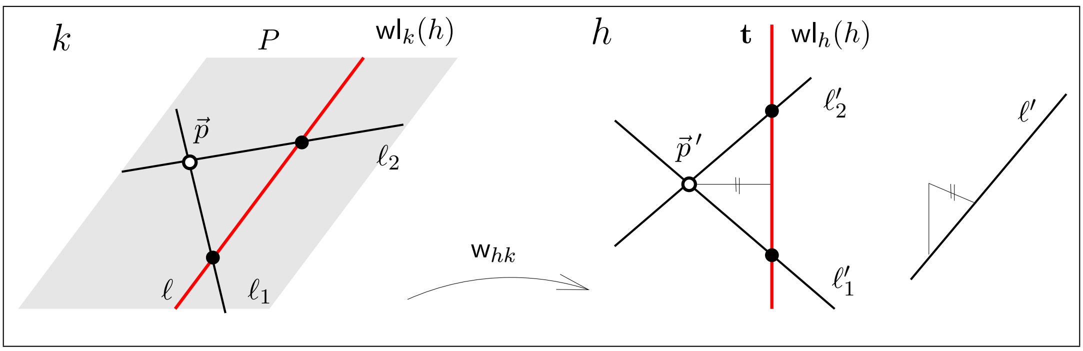


Figure 10: Illustration for item (3) of the proof of Theorem 6.2 (Line-to-Line Lemma).

Taken together, these results imply that whenever $P \in \mathsf{ObPlanes}(k)$, then $\mathsf{w}_{hk}[P]$ is a plane.

(4) Now let $k \in IOb$. We want to prove that any line can be obtained as the intersection of two planes in $\mathsf{ObPlanes}(k)$. To see this, let ℓ be any line, and choose any $\vec{p} \in \ell$, see Figure 11. As we have just seen, we can also choose $\ell' \in \mathsf{ObLines}(k)$ such that $\mathsf{slope}(\ell') \neq 0$ and (by assumption) $\mathsf{slope}(\ell') \neq \infty$. Let ℓ_1, ℓ_2 be lines passing through $\vec{p}$, having the same slope as ℓ', such that ℓ, ℓ_1 and ℓ_2 are not co-planar (such lines can be obtained from ℓ' by a combination of translation and spatial rotation). It follows from Theorem 6.1 (Observer Lines Lemma) that $\ell_1, \ell_2 \in \mathsf{ObLines}(k)$. For each $i = 1, 2$, let P_i be the plane containing ℓ_i and ℓ. Then P_1, P_2 are k-observer planes and their intersection is ℓ, as required.

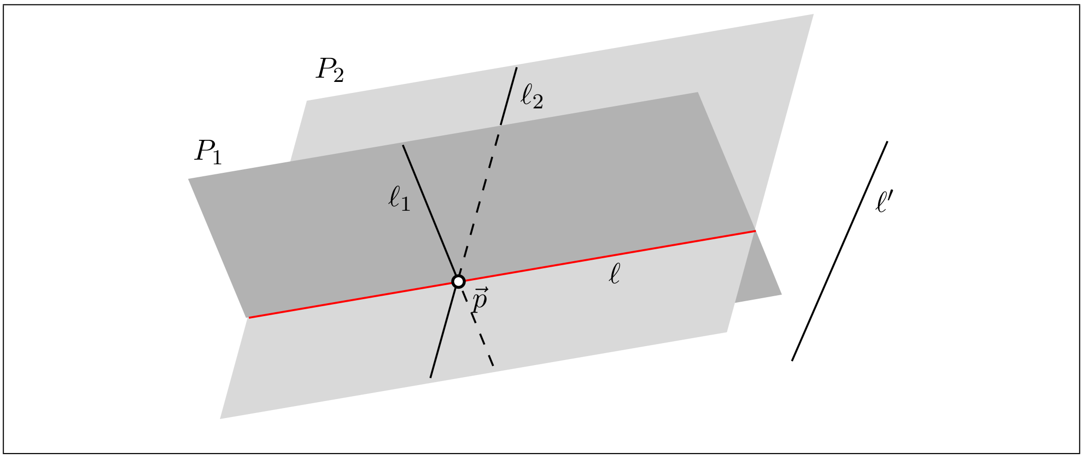


Figure 11: Illustration for item (4) of the proof of Theorem 6.2 (Line-to-Line Lemma).

It now follows, once again, that given any $k, h \in IOb$, the worldview transformation w_{hk} is a bijection that takes lines to lines. For if ℓ is any line, choose k-observer planes P_1, P_2 such that $\ell = P_1 \cap P_2$. Since w_{hk} is one-to-one, $\mathsf{w}_{hk}[\ell] = \mathsf{w}_{hk}[P_1] \cap \mathsf{w}_{hk}[P_2]$ and $\mathsf{w}_{hk}[P_1] \neq \mathsf{w}_{hk}[P_2]$ (as $P_1 \neq P_2$). Since $\mathsf{w}_{hk}[P_1]$ and $\mathsf{w}_{hk}[P_2]$ are distinct intersecting planes, their intersection $\mathsf{w}_{hk}[\ell]$ is a line.

This completes the proof that lines are mapped to lines. The claim for planes and hyperplanes now follows easily. Given a plane, choose three non-collinear points. These determine three distinct intersecting lines and their images determine the image plane. Likewise, we can choose four non-coplanar points in a hyperplane whose images determine the image hyperplane. □

6.3 The tx-Plane Lemma

Definition 6.3.1 (Principal Observer)**.** We now fix one observer o for the rest of the paper (the *principal observer*) and define

$$IOb_o \stackrel{\text{def}}{=} \{k \in IOb : \mathsf{w}_{ko}(\vec{\mathsf{o}}) = \vec{\mathsf{o}}\}$$

to be the set of observers who agree with o (and hence each other) as to the location of the origin. □

Analogously to the definition of the time-axis $\mathbf{t}$, the three spatial axes ($\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$) are defined in the usual way as:

$$\mathbf{x} \stackrel{\text{def}}{=} \{(0, x, 0, 0) : x \in Q\}, \quad \mathbf{y} \stackrel{\text{def}}{=} \{(0, 0, y, 0) : y \in Q\}, \quad \mathbf{z} \stackrel{\text{def}}{=} \{(0, 0, 0, z) : z \in Q\}.$$

We write $\mathsf{plane}(\mathbf{t},\mathbf{x})$ for the tx-plane and $\mathsf{plane}(\mathbf{y},\mathbf{z})$ for the yz-plane. More generally, if $\ell \neq \ell'$ are intersecting lines, then $\mathsf{plane}(\ell,\ell')$ denotes the plane containing ℓ and ℓ'.

Theorem 6.3 (tx-Plane Lemma)**.** Assume $\mathsf{KIN} + \mathsf{AxIsotropy}$. Let $m, k \in IOb_o$ such that $\mathsf{w}_{km}[\mathsf{plane}(\mathbf{t},\mathbf{x})] = \mathsf{plane}(\mathbf{t},\mathbf{x})$. Then

$$\mathsf{w}_{km}[\mathsf{plane}(\mathbf{y},\mathbf{z})] = \mathsf{plane}(\mathbf{y},\mathbf{z}) \tag{6.2}$$

and

$$\text{if } \vec{q}\,,\vec{p}\, \in \mathsf{plane}(\mathbf{y},\mathbf{z}) \text{ and } |\vec{p}\,| = |\vec{q}\,|, \text{ then } |\mathsf{w}_{km}(\vec{p}\,)| = |\mathsf{w}_{km}(\vec{q}\,)|\,. \tag{6.3}$$

Moreover, if w_{km} is also linear, then there is a positive $\lambda \in Q$ such that

$$|\mathsf{w}_{km}(\vec{p}\,)| = \lambda|\vec{p}\,| \tag{6.4}$$

for all $\vec{p}\, \in \mathsf{plane}(\mathbf{y},\mathbf{z})$.

6.3.1 Supporting lemmas

The supporting lemmas can be informally described as:

Lemma 6.3.2 ($\mathsf{Triv} = \bigcap_\kappa \mathsf{Iso}$)
A transformation is trivial if and only if it is a κ-isometry for at least two different choices of κ.

Lemma 6.3.3 (IOb$_o$)
Elementary results concerning worldview transformations involving members of IOb_o.

Lemma 6.3.4 (Affine)
Suppose f is a bijection on Q^4 taking lines to lines. Then there is an automorphism φ of Q and an affine transformation A such that $f = A \circ \widetilde{\varphi}$ (where $\widetilde{\varphi}$ is the coordinatewise extension of φ to Q^4).

Lemma 6.3.5 (Equal Worldlines)
If any one observer considers $m, m^* \in IOb$ to have the same worldline, then all other observers do so as well.

Lemma 6.3.7 (Colocate)
If two observers share the same worldline, the worldview transformation between them is trivial.

6.3.2 Proofs of the supporting lemmas

Lemma 6.3.2 ($\mathsf{Triv} = \bigcap_\kappa \mathsf{Iso}$). Assume that $(Q, +, \cdot, 0, 1)$ is a field and choose $x, y \in Q$ such that $x \neq y$. Then

$$\mathsf{Triv} = {}_x\mathsf{Iso} \cap {}_y\mathsf{Iso}.$$

In particular, every trivial transformation is a Euclidean isometry.

Proof. ($\subseteq$) Choose any $x, y \in Q$, $T \in \mathsf{Triv}$ and $\vec{p} = (t, \vec{s}\,) \in Q^4$. We will show that $T \in {}_x\mathsf{Iso}$. Without loss of generality we can assume that T is linear (since it is the composition of a linear map with a translation, and all translations are x-isometries). It follows that $T(\vec{p}\,) = T(t, \vec{0}\,) + T(0, \vec{s}\,)$. However, because T is trivial, we know that it fixes and preserves squared lengths in both $\mathbf{t}$ and $\mathbf{S}$, so there exist t', $\vec{s}\,'$ such that $T(t, \vec{0}\,) = (t', \vec{0}\,)$ and $T(0, \vec{s}\,) = (0, \vec{s}\,')$, where $|t|^2 = |t'|^2$ and $|\vec{s}\,|^2 = |\vec{s}\,'|^2$. It follows immediately that $\|\vec{p}\,\|_x = |t|^2 - x|\vec{s}\,|^2 = |t'|^2 - x|\vec{s}\,'|^2 = \|T(\vec{p}\,)\|_x$, i.e. T preserves squared κ-lengths. It now follows that $T \in {}_x\mathsf{Iso}$ when $x \neq 0$, and because $|\vec{s}\,|^2 = |\vec{s}\,'|^2$ no matter what the value of t, we also have $T \in {}_x\mathsf{Iso}$ when $x = 0$. Finally, because x can be any value in Q we also have $T \in {}_y\mathsf{Iso}$, and hence $\mathsf{Triv} \subseteq {}_x\mathsf{Iso} \cap {}_y\mathsf{Iso}$, as claimed.

($\supseteq$) To show the converse, choose any $x \neq y \in Q$ and any $T \in {}_x\mathsf{Iso} \cap {}_y\mathsf{Iso}$. We will show that $T \in \mathsf{Triv}$.

Assume first that T is linear. Choose any $\vec{p} = (t, \vec{s}\,) \in Q^4$ and suppose $T(\vec{p}\,) = (t', \vec{s}\,')$. Because T is in both ${}_x\mathsf{Iso}$ and ${}_y\mathsf{Iso}$, we have both $\|T(\vec{p}\,)\|_x = \|\vec{p}\,\|_x$ and $\|T(\vec{p}\,)\|_y = \|\vec{p}\,\|_y$, i.e.

$$|t'|^2 - x|\vec{s}\,'|^2 = |t|^2 - x|\vec{s}\,|^2 \text{ and} \tag{6.5}$$
$$|t'|^2 - y|\vec{s}\,'|^2 = |t|^2 - y|\vec{s}\,|^2. \tag{6.6}$$

Subtracting (6.6) from (6.5) gives

$$(x - y)|\vec{s}\,'|^2 = (x - y)|\vec{s}\,|^2$$

whence division by $(x - y) \neq 0$ gives both

$$|\vec{s}\,'|^2 = |\vec{s}\,|^2 \tag{6.7}$$

and hence (by either (6.5) or (6.6))

$$|t'|^2 = |t|^2. \tag{6.8}$$

Therefore,

$$\text{if } t = 0, \text{ then } t' = 0, \text{ and}$$
$$\text{if } \vec{s} = \vec{0}, \text{ then } \vec{s}\,' = \vec{0},$$

which together with (6.7) and (6.8) show that $T \in \mathsf{Triv}$.

If T is not itself linear, notice that we can write $T = L \circ \tau$ where τ is a translation and L is a linear x-isometry. Since $T \in {}_y\mathsf{Iso}$ and $L = T \circ \tau^{-1}$ differs from T only by a translation (and all translations are in ${}_y\mathsf{Iso}$), we see that L is in ${}_y\mathsf{Iso}$ too. Thus, L is a linear map in ${}_x\mathsf{Iso} \cap {}_y\mathsf{Iso}$ (in other words, the "linear" and "translation" parts of T are the same in ${}_x\mathsf{Iso}$ as in ${}_y\mathsf{Iso}$) whence it follows from what we have just shown that L is trivial. Because τ is trivial, we now conclude that $T = L \circ \tau$ is itself trivial, as claimed.

In particular, we have $\mathsf{Triv} = ({}_0\mathsf{Iso} \cap {}_{-1}\mathsf{Iso}) \subseteq {}_{-1}\mathsf{Iso}$, i.e. all trivial transformations are Euclidean isometries. $\square$

Lemma 6.3.3 (IOb_o)**.** Assume AxWvt. Let $k, h \in IOb_o$ and $m \in IOb$. Then (a)–(c) below hold.

(a) $\mathsf{w}_{kh}(\vec{\mathsf{o}}) = \vec{\mathsf{o}}$ and $\vec{\mathsf{o}} \in \mathsf{wl}_k(h)$.

(b) If $\mathsf{w}_{km}(\vec{\mathsf{o}}) = \vec{\mathsf{o}}$, then $m \in IOb_o$.

(c) If $R : Q^4 \to Q^4$, $R(\vec{\mathsf{o}}) = \vec{\mathsf{o}}$ and $k \overset{R}{\rightsquigarrow}_h m$, then $m \in IOb_o$.

Proof. The proof involves only straightforward applications of Lemma 6.1.2 (WVT), and we omit the details. $\square$

Lemma 6.3.4 (Affine)**.** Assume $\mathcal{Q} = (Q, +, \cdot, 0, 1, \leq)$ is a Euclidean field, and suppose $f \colon Q^4 \to Q^4$ is a bijection taking lines to lines. Then there is an ordered-field automorphism φ of $\mathcal{Q}$ and an affine transformation A on Q^4 such that $f = A \circ \widetilde{\varphi}$, where $\widetilde{\varphi} \colon Q^4 \to Q^4$ is the map $\widetilde{\varphi} \colon (t, x, y, z) \mapsto (\varphi(t), \varphi(x), \varphi(y), \varphi(z))$.

Proof. By the Fundamental Theorem of Affine Geometry [6, Thm. 2.6.3, p. 52], there is an automorphism φ of field $(Q, +, \cdot, 0, 1)$ and an affine transformation A such that $f = A \circ \widetilde{\varphi}$. To complete the proof of the lemma, we only have to show that φ is order preserving, i.e. $\varphi(a) \leq \varphi(b)$ iff $a \leq b$. Since $x \leq y$ iff $0 \leq y - x$, it is enough to show that $0 \leq \varphi(z)$ iff $0 \leq z$ — and this follows directly from the Euclidean property, i.e. $0 \leq d$ iff $d = c^2$ for some $c \in Q$. $\square$

Lemma 6.3.5 (Equal Worldlines)**.** Assume AxWvt. Suppose $m, m^* \in IOb$, and suppose $\mathsf{wl}_k(m) = \mathsf{wl}_k(m^*)$ for some $k \in IOb$. Then $\mathsf{wl}_j(m) = \mathsf{wl}_j(m^*)$ for all $j \in IOb$.

Proof. By Lemma 6.1.2 (WVT), $\mathsf{wl}_j(m) = \mathsf{w}_{jk}[\mathsf{wl}_k(m)] = \mathsf{w}_{jk}[\mathsf{wl}_k(m^*)] = \mathsf{wl}_j(m^*)$ for all $j \in IOb$. □

Definition 6.3.6. Let $m, m^* \in IOb$. If $\mathsf{wl}_k(m) = \mathsf{wl}_k(m^*)$ for some $k \in IOb$, we say that m and m^* ***share the same worldline***.

Lemma 6.3.7 (Colocate)**.** Assume AxWvt and let $m, m^* \in IOb$. Suppose m and m^* share the same worldline. If $\mathsf{AxColocate}$ holds, then $\mathsf{w}_{mm^*} \in \mathsf{Triv}$.

Proof. Saying that m and m^* share the same worldline means that $\mathsf{wl}_k(m) = \mathsf{wl}_k(m^*)$ for some $k \in IOb$. By Lemma 6.3.5 (Equal Worldlines), this equation therefore holds for *all* choices of k, and in particular for $k = m$, i.e. $\mathsf{wl}_m(m) = \mathsf{wl}_m(m^*)$. The claim now follows immediately by $\mathsf{AxColocate}$. □

6.3.3 Main proof

We now complete the proof of Theorem 6.3 (tx-Plane Lemma).

Proof of Theorem 6.3 (tx-Plane Lemma). Let observers $m, k \in IOb_o$ be such that $\mathsf{w}_{km}[\mathsf{plane}(\mathbf{t}, \mathbf{x})] = \mathsf{plane}(\mathbf{t}, \mathbf{x})$. By Lemma 6.3.3 (IOb$_o$), $\mathsf{w}_{mk}(\vec{\mathsf{o}}) = \mathsf{w}_{km}(\vec{\mathsf{o}}) = \vec{\mathsf{o}}$.

Let us first prove the following claim

$$\text{If } R \in \mathsf{SRot} \text{ fixes } \mathsf{plane}(\mathbf{t}, \mathbf{x}) \text{ pointwise, then there exists } k^* \in IOb \\ \text{such that (a) } \mathsf{w}_{kk^*} = \mathsf{w}_{km} \circ R \circ \mathsf{w}_{mk} \text{ and (b) } \mathsf{w}_{kk^*} \in \mathsf{Triv}. \tag{6.9}$$

Proof of claim (6.9). (a) By Lemma 6.1.4 (Observer Rotation), there exists some k^* such that $k \overset{R}{\rightsquigarrow}_m k^*$, i.e. $\mathsf{w}_{mk^*} = R \circ \mathsf{w}_{mk}$. Hence, $\mathsf{w}_{kk^*} = \mathsf{w}_{km} \circ \mathsf{w}_{mk^*} = \mathsf{w}_{km} \circ R \circ \mathsf{w}_{mk}$. (b) By Lemma 6.1.3 (Worldline Relocation), we have $\mathsf{wl}_m(k^*) = R[\mathsf{wl}_m(k)]$, and because $\mathsf{wl}_m(k) = \mathsf{w}_{mk}[\mathbf{t}] \subseteq \mathsf{plane}(\mathbf{t}, \mathbf{x})$ and R leaves $\mathsf{plane}(\mathbf{t}, \mathbf{x})$ pointwise-fixed, we have that $R[\mathsf{wl}_m(k)] = \mathsf{wl}_m(k)$. Thus, $\mathsf{wl}_m(k^*) = R[\mathsf{wl}_m(k)] = \mathsf{wl}_m(k)$, i.e. k and k^* share the same worldline. So $\mathsf{w}_{kk^*} \in \mathsf{Triv}$ by Lemma 6.3.7 (Colocate). Thus, (6.9) holds.

Proof of statement (6.2). Choose any $\vec{p} \in \mathsf{plane}(\mathbf{y}, \mathbf{z})$ and write $\vec{p}\,' := \mathsf{w}_{km}(\vec{p})$. We have to prove that $\vec{p}\,' \in \mathsf{plane}(\mathbf{y}, \mathbf{z})$.

We will show that $\vec{p}\,' \cdot \vec{q} = 0$ for every $\vec{q} \in \mathsf{plane}(\mathbf{t}, \mathbf{x})$, whence it follows easily that $\vec{p}\,' \in \mathsf{plane}(\mathbf{y}, \mathbf{z})$.

By Lemma 6.3.4 (Affine), Theorem 6.2 (Line-to-Line Lemma) and the fact that $\mathsf{w}_{km}(\vec{\mathsf{o}}) = \vec{\mathsf{o}}$, we know that w_{km} can be written as a composition $\mathsf{w}_{km} = L \circ \widetilde{\varphi}$ of a linear transformation, L, and a map induced by a field automorphism, φ. Therefore, $\mathsf{w}_{km}(-\vec{p}) = L(\widetilde{\varphi}(-\vec{p})) = L(-\widetilde{\varphi}(\vec{p})) = -L(\widetilde{\varphi}(\vec{p})) = -\mathsf{w}_{km}(\vec{p}) = -\vec{p}\,'$.

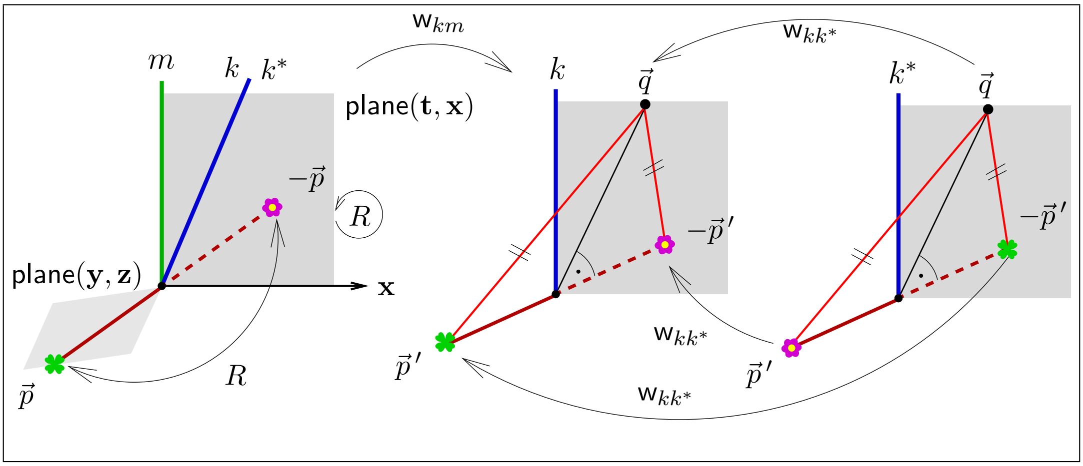


Figure 12: Illustration for the proof of (6.2) of Theorem 6.3 (tx-Plane Lemma).

Let R be the linear transformation that takes $\vec{\mathsf{t}},\vec{\mathsf{x}},\vec{\mathsf{y}},\vec{\mathsf{z}}$ to $\vec{\mathsf{t}},\vec{\mathsf{x}},-\vec{\mathsf{y}},-\vec{\mathsf{z}}$, respectively. Then R is a self-inverse spatial rotation that leaves plane$(\mathbf{t},\mathbf{x})$ pointwise fixed and takes $\vec{p}$ to $-\vec{p}$, see Figure 12. So by (6.9), there is $k^* \in IOb$ such that $\mathsf{w}_{kk^*} \in \mathsf{Triv}$ and $\mathsf{w}_{kk^*} = \mathsf{w}_{km} \circ R \circ \mathsf{w}_{mk}$.

Let $\vec{q} \in$ plane$(\mathbf{t},\mathbf{x})$ be arbitrary. Now note that $\mathsf{w}_{mk}(\vec{q}\,) \in$ plane$(\mathbf{t},\mathbf{x})$, hence $R(\mathsf{w}_{mk}(\vec{q}\,)) = \mathsf{w}_{mk}(\vec{q}\,)$. Note also that $\mathsf{w}_{kk^*}(\vec{p}\,') = -\vec{p}\,'$ and $\mathsf{w}_{kk^*}(\vec{q}\,) = \vec{q}$ because

$$\mathsf{w}_{kk^*}(\vec{p}\,') = \mathsf{w}_{km}(R(\mathsf{w}_{mk}(\vec{p}\,'))) = \mathsf{w}_{km}(R(\vec{p}\,)) = \mathsf{w}_{km}(-\vec{p}\,) = -\vec{p}\,',$$

$$\mathsf{w}_{kk^*}(\vec{q}\,) = \mathsf{w}_{km}(R(\mathsf{w}_{mk}(\vec{q}\,))) = \mathsf{w}_{km}(\mathsf{w}_{mk}(\vec{q}\,)) = \vec{q}\,.$$

Now, because w_{kk^*} is trivial, we know from Lemma 6.3.2 ($\mathsf{Triv} = \bigcap_\kappa \mathsf{Iso}$) that it is a Euclidean isometry. Moreover, because every trivial map is the composition of a linear map and a translation, and since it fixes $\vec{\mathsf{o}}$ (because w_{km}, w_{mk} and R all do so), w_{kk^*} must be linear.

It follows that $|\vec{q} - \vec{p}\,'| = |\mathsf{w}_{kk^*}(\vec{q} - \vec{p}\,')| = |\mathsf{w}_{kk^*}(\vec{q}\,) - \mathsf{w}_{kk^*}(\vec{p}\,')| = |\vec{q} + \vec{p}\,'|$, whence $(\vec{q} - \vec{p}\,')\cdot(\vec{q} - \vec{p}\,') = (\vec{q} + \vec{p}\,')\cdot(\vec{q} + \vec{p}\,')$, and so $\vec{p}\,' \cdot \vec{q} = 0$.

Since this holds for any $\vec{q} \in$ plane$(\mathbf{t},\mathbf{x})$, in particular it holds for both $\vec{\mathsf{t}}$ and $\vec{\mathsf{x}}$. Consequently, $\vec{p}\,' \in$ plane$(\mathbf{y},\mathbf{z})$ as claimed.

Proof of statement (6.3). Let $\vec{p},\vec{q} \in$ plane$(\mathbf{y},\mathbf{z})$ and write $\vec{p}\,' := \mathsf{w}_{km}(\vec{p}\,)$ and $\vec{q}\,' := \mathsf{w}_{km}(\vec{q}\,)$. Assume $|\vec{p}\,| = |\vec{q}\,|$. We want to prove that $|\vec{p}\,'| = |\vec{q}\,'|$. By Lemma 6.1.7 (Horizontal Rotation), there is a spatial rotation that takes $\vec{\mathsf{x}}$ to $\vec{\mathsf{x}}$ and $\vec{p}$ to $\vec{q}$. Let $R' \in \mathsf{SRot}$ be such a spatial rotation. Then R' leaves plane$(\mathbf{t},\mathbf{x})$ pointwise fixed and takes $\vec{p}$ to $\vec{q}$. By (6.9), there is $k^* \in IOb$ such that $\mathsf{w}_{kk^*} \in \mathsf{Triv}$ and

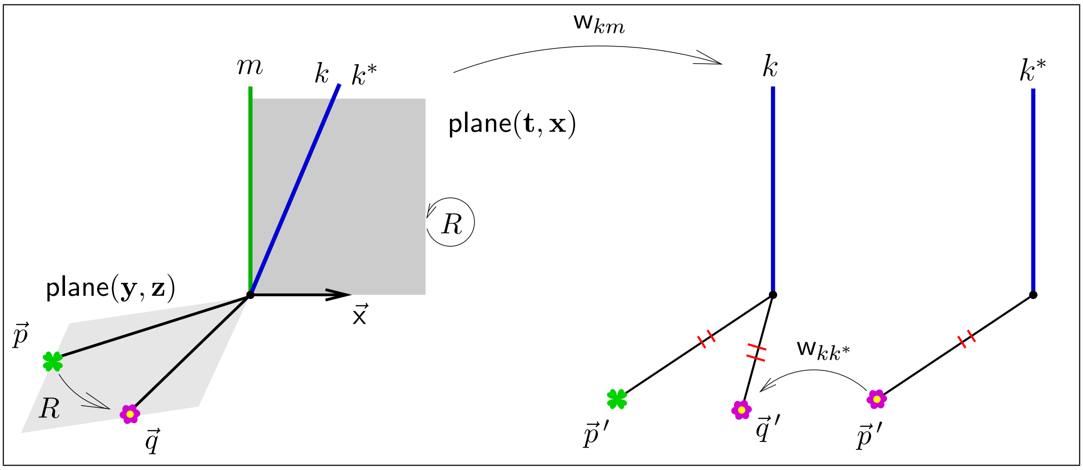


Figure 13: Illustration for the proof of (6.3) of Theorem 6.3 (tx-Plane Lemma).

$\mathsf{w}_{kk^*} = \mathsf{w}_{km} \circ R' \circ \mathsf{w}_{mk}$, see Figure 13. It follows that

$$\mathsf{w}_{kk^*}(\vec{p}\,') = \mathsf{w}_{km}(R'(\mathsf{w}_{mk}(\vec{p}\,'))) = \mathsf{w}_{km}(R'(\vec{p}\,)) = \mathsf{w}_{km}(\vec{q}\,) = \vec{q}\,'.$$

Finally, because w_{kk^*} is trivial, Lemma 6.3.2 ($\mathsf{Triv} = \bigcap_\kappa \mathsf{Iso}$) tells us that it is a Euclidean isometry. It now follows that $|\vec{p}\,'| = |\mathsf{w}_{kk^*}(\vec{p}\,')| = |\vec{q}\,'|$, as claimed. Thus, (6.3) holds.

Proof of statement (6.4). Now assume that w_{km} is linear. Let $\lambda := |\mathsf{w}_{km}(\vec{\mathsf{y}}\,)|$. This λ is positive since $\mathsf{w}_{km}(\vec{\mathsf{y}}\,) \neq \vec{\mathsf{o}}$ as $m, k \in IOb_o$. We will prove that $|\mathsf{w}_{km}(\vec{p}\,)| = \lambda\,|\vec{p}\,|$ for every $\vec{p} \in \mathsf{plane}(\mathbf{y},\mathbf{z})$. Clearly for $\vec{p} = \vec{\mathsf{o}}$ this holds, so assume that $\vec{p} \in \mathsf{plane}(\mathbf{y},\mathbf{z}) \setminus \{\vec{\mathsf{o}}\,\}$, and note that

$$\left|\frac{\vec{p}}{|\vec{p}\,|}\right| = 1 = |\vec{\mathsf{y}}\,|\,.$$

Then, by (6.3),

$$\left|\mathsf{w}_{km}\left(\frac{\vec{p}}{|\vec{p}\,|}\right)\right| = |\mathsf{w}_{km}(\vec{\mathsf{y}}\,)| = \lambda.$$

Therefore, by linearity of w_{km},

$$|\mathsf{w}_{km}(\vec{p}\,)| = \left|\,|\vec{p}\,|\,\mathsf{w}_{km}\left(\frac{\vec{p}}{|\vec{p}\,|}\right)\right| = \lambda\,|\vec{p}\,|\,.$$

□

6.4 The Same-Speed Lemma

Theorem 6.4 (Same-Speed Lemma)**.** Assume KIN, $\mathsf{AxIsotropy}$, and that $k, m, h \in IOb_o$. If $\mathsf{speed}_m(k) = \mathsf{speed}_m(h)$, then

(a) there exists κ such that w_{hk} is a κ-isometry;

(b) $\mathsf{speed}_k(h) = \mathsf{speed}_h(k)$;

(c) $\mathsf{speed}_h(m) = \mathsf{speed}_k(m)$.

6.4.1 Supporting lemmas

The supporting lemmas can be informally described as:

Lemma 6.4.1 (Translation to IOb$_o$)
Every observer can be translated into IOb_o.

Lemma 6.4.2 (Vertical Plane Rotation)
Every vertical plane can be rotated into the tx-plane.

Lemma 6.4.3 (LinTriv ⇒ Same Speed)
If w_{mm^*} is both linear and trivial, then every j agrees that m and m^* are moving at the same speed, and likewise m and m^* agree on the speed of j.

6.4.2 Proofs of the supporting lemmas

Lemma 6.4.1 (Translation to IOb$_o$)**.** Assume AxWvt and $\mathsf{AxRelocate}$. Given any $k \in IOb$ there exists $k^o \in IOb_o$ such that $\mathsf{w}_{k^o k}$ is a translation.

Proof. Let T be the translation taking $\mathsf{w}_{ko}(\vec{\mathsf{o}})$ to the origin and let k^o be an observer such that $\mathsf{w}_{k^o k} = T$ (such an observer exists by $\mathsf{AxRelocate}$). Then $\mathsf{w}_{k^o o}(\vec{\mathsf{o}}) = (\mathsf{w}_{k^o k} \circ \mathsf{w}_{ko})(\vec{\mathsf{o}}) = T(\mathsf{w}_{ko}(\vec{\mathsf{o}})) = \vec{\mathsf{o}}$, so $k^o \in IOb_o$ as required. □

Lemma 6.4.2 (Vertical Plane Rotation)**.** Assume $(Q, +, \cdot, 0, 1, \leq)$ is a Euclidean field, that P is a plane in Q^4 containing the time-axis $\mathbf{t}$, and that $\vec{p} \in P \setminus \mathbf{t}$. Then there exists a spatial rotation R that takes P and $\vec{p}$ to $\mathsf{plane}(\mathbf{t}, \mathbf{x})$ and $(\vec{p}_t, |\vec{p}_s|, 0, 0)$, respectively.

Proof. By Lemma 6.1.7 (Horizontal Rotation), there is $R \in \mathsf{SRot}$ which takes $(0, \vec{p}_s)$ to $(0, |\vec{p}_s|, 0, 0)$ and $\vec{\mathsf{o}}$ to $\vec{\mathsf{o}}$; see Figure 14. It is easy to see that this R has the desired properties. □

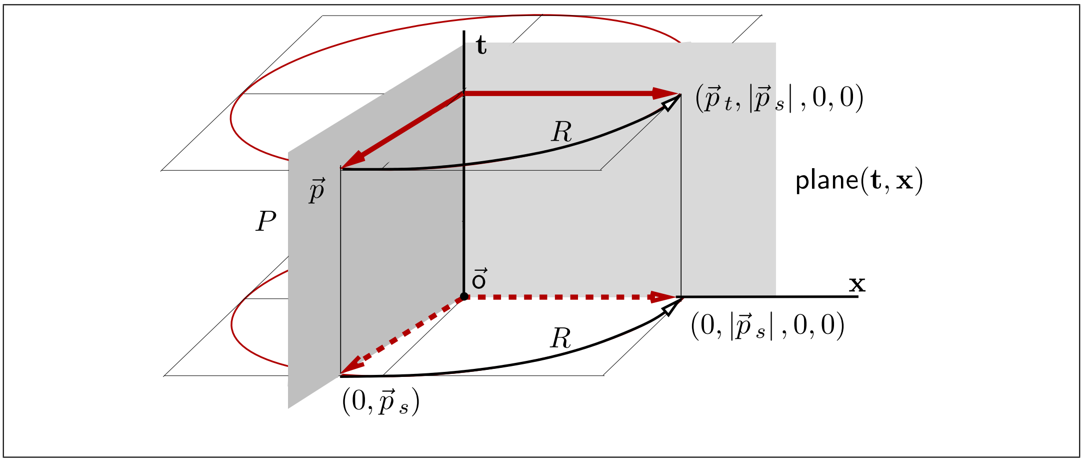


Figure 14: Illustration for Lemma 6.4.2 (Vertical Plane Rotation).

Lemma 6.4.3 (LinTriv ⇒ Same Speed)**.** Assume AxWvt and AxEField and suppose $m, m^* \in IOb$ and w_{m^*m} is a linear trivial transformation. Then $\mathsf{wl}_j(m) = \mathsf{wl}_j(m^*)$ for every observer $j \in IOb$. Furthermore, if AxLine is assumed, then $\mathsf{speed}_j(m) = \mathsf{speed}_j(m^*)$ and $\mathsf{speed}_m(j) = \mathsf{speed}_{m^*}(j)$ for every $j \in IOb$.

Proof. Recall that $\mathsf{wl}_{m^*}(m) = \mathsf{w}_{m^*m}[\mathbf{t}]$. Since w_{m^*m} is a linear trivial transformation, we have $\mathsf{w}_{m^*m}[\mathbf{t}] = \mathbf{t} = \mathsf{wl}_{m^*}(m^*)$. Thus, $\mathsf{wl}_{m^*}(m) = \mathsf{wl}_{m^*}(m^*)$. Hence, for every $j \in IOb$, $\mathsf{wl}_j(m) = \mathsf{wl}_j(m^*)$ by Lemma 6.3.5 (Equal Worldlines).

Now, assume AxLine and let $j \in IOb$. Then $\mathsf{speed}_j(m) = \mathsf{speed}_j(m^*)$ since $\mathsf{wl}_j(m) = \mathsf{wl}_j(m^*)$. It is easy to see that $\mathsf{slope}(\ell) = \mathsf{slope}(f[\ell])$ holds for every trivial transformation f and line ℓ. Therefore, $\mathsf{speed}_m(j) = \mathsf{slope}(\mathsf{wl}_m(j)) = \mathsf{slope}(\mathsf{w}_{m^*m}[\mathsf{wl}_m(j)]) = \mathsf{slope}(\mathsf{wl}_{m^*}(j)) = \mathsf{speed}_{m^*}(j)$. □

6.4.3 Main proof

We now complete the proof of Theorem 6.4 (Same-Speed Lemma).

Proof of Theorem 6.4 (Same-Speed Lemma). Suppose $\mathsf{speed}_m(k) = \mathsf{speed}_m(h)$ for some $m, k, h \in IOb_o$.

(a) If $\mathsf{wl}_m(k) = \mathsf{wl}_m(h)$, then w_{hk} is a trivial transformation by Lemma 6.3.7 (Colocate), hence it is a κ-isometry by Lemma 6.3.2 ($\mathsf{Triv} = \bigcap_\kappa \mathsf{Iso}$).

Assume, therefore, that $\mathsf{wl}_m(k) \neq \mathsf{wl}_m(h)$. Because k and h have the same speed in m's worldview, their worldlines have the same slope according to m. By Lemma 6.3.3 (IOb_o), $\vec{o} \in \mathsf{wl}_m(k) \cap \mathsf{wl}_m(h)$ because $m, k, h \in IOb_o$.

Let $\vec{p}_1 \in \mathsf{wl}_m(k)$ and $\vec{p}_2 \in \mathsf{wl}_m(h)$ be such that $\vec{p}_1 \neq \vec{\mathsf{o}} \neq \vec{p}_2$ and $(\vec{p}_1)_t = (\vec{p}_2)_t$, see Figure 15. Let $t^* := (\vec{p}_1)_t$ be the common time component of $\vec{p}_1$ and $\vec{p}_2$.

Let $\vec{s}_1 := (\vec{p}_1)_s$ and $\vec{s}_2 := (\vec{p}_2)_s$. Then $|\vec{s}_1| = |\vec{s}_2|$ because lines $\mathsf{wl}_m(k)$ and $\mathsf{wl}_m(h)$ are of same slope. Thinking of $\vec{s}_1$ and $\vec{s}_2$ as points in Q^3, let $\vec{s}^*$ be the point mid-way between them, i.e. $\vec{s}^* = (\vec{s}_1 + \vec{s}_2)/2$, and let ℓ be a line in Q^3 passing through $\vec{0}$ and $\vec{s}^*$. If we now define ρ to be the map which rotates Q^3 through $180°$ about axis ℓ, then the map R given by $R(t, \vec{s}) := (t, \rho(\vec{s}))$ is a self-inverse spatial rotation.[12]

We claim that $R(\vec{p}_1) = \vec{p}_2$. To see this, notice that the points $\vec{s}_1$ and $\vec{s}_2$ form the base of an isosceles triangle in Q^3 whose vertex is $\vec{0}$; it follows easily that the line ℓ bisects and is orthogonal to the line joining $\vec{s}_1$ to $\vec{s}_2$, whence the rotation ρ about ℓ maps $\vec{s}_1$ to $\vec{s}_2$ (and vice versa) in Q^3. Thus, $R(\vec{p}_1) = R(t^*, \vec{s}_1) = (t^*, \rho(\vec{s}_1)) = (t^*, \vec{s}_2) = \vec{p}_2$. Since R also fixes $\vec{\mathsf{o}}$, it must take $\mathsf{wl}_m(k)$ to $\mathsf{wl}_m(h)$. Point $\vec{s}^*$ is fixed by ρ because this point is on ρ's axis of rotation. Therefore, $(0, \vec{s}^*)$ is fixed by R.

So we have $R \in \mathsf{SRot}$, $R[\mathsf{wl}_m(k)] = \mathsf{wl}_m(h)$, $R^{-1} = R$ and $R(0, \vec{s}^*) = (0, \vec{s}^*)$.

Choose $h' \in IOb_o$ such that $k \overset{R}{\rightsquigarrow}_m h'$. Such h' exists by Lemma 6.1.4 (Observer Rotation) and Lemma 6.3.3 (IOb_o). By Lemma 6.1.3 (Worldline Relocation), we have $\mathsf{wl}_m(h') = R[\mathsf{wl}_m(k)]$, and since $R[\mathsf{wl}_m(k)] = \mathsf{wl}_m(h)$, we must have

$$\mathsf{wl}_m(h) = \mathsf{wl}_m(h'), \tag{6.10}$$

i.e. h and h' share the same worldline. It follows, by Lemma 6.3.7 (Colocate) and $h, h' \in IOb_o$, that

$$\mathsf{w}_{hh'} \text{ is a linear trivial transformation.} \tag{6.11}$$

Our goal is to prove that $\mathsf{w}_{hk} \in {}_\kappa\mathsf{Iso}$ for some κ. Since $\mathsf{w}_{hk} = \mathsf{w}_{hh'} \circ \mathsf{w}_{h'k}$ and (as we have just seen) $\mathsf{w}_{hh'}$ is trivial, it is enough to prove that $\mathsf{w}_{h'k} \in {}_\kappa\mathsf{Iso}$ for some κ.

By $k \overset{R}{\rightsquigarrow}_m h'$, we have $\mathsf{w}_{mh'} = R \circ \mathsf{w}_{mk}$. Thus,

$$\mathsf{w}_{kh'} = \mathsf{w}_{km} \circ \mathsf{w}_{mh'} = \mathsf{w}_{km} \circ R \circ \mathsf{w}_{mk} \tag{6.12}$$

and

$$\mathsf{w}_{h'k} = (\mathsf{w}_{km} \circ R \circ \mathsf{w}_{mk})^{-1} = \mathsf{w}_{km} \circ R^{-1} \circ \mathsf{w}_{mk}$$

whence (as $R^{-1} = R$)

$$\mathsf{w}_{h'k} = \mathsf{w}_{kh'}, \text{ and thus } \mathsf{wl}_k(h') = \mathsf{wl}_{h'}(k). \tag{6.13}$$

[12] We can define ρ in the usual way. Given any $\vec{s}$ we decompose it into a sum $\vec{s} = \vec{s}_\parallel + \vec{s}_\perp$ of components parallel and perpendicular to ℓ, respectively, and then $\rho(\vec{s}) = \vec{s}_\parallel - \vec{s}_\perp$.

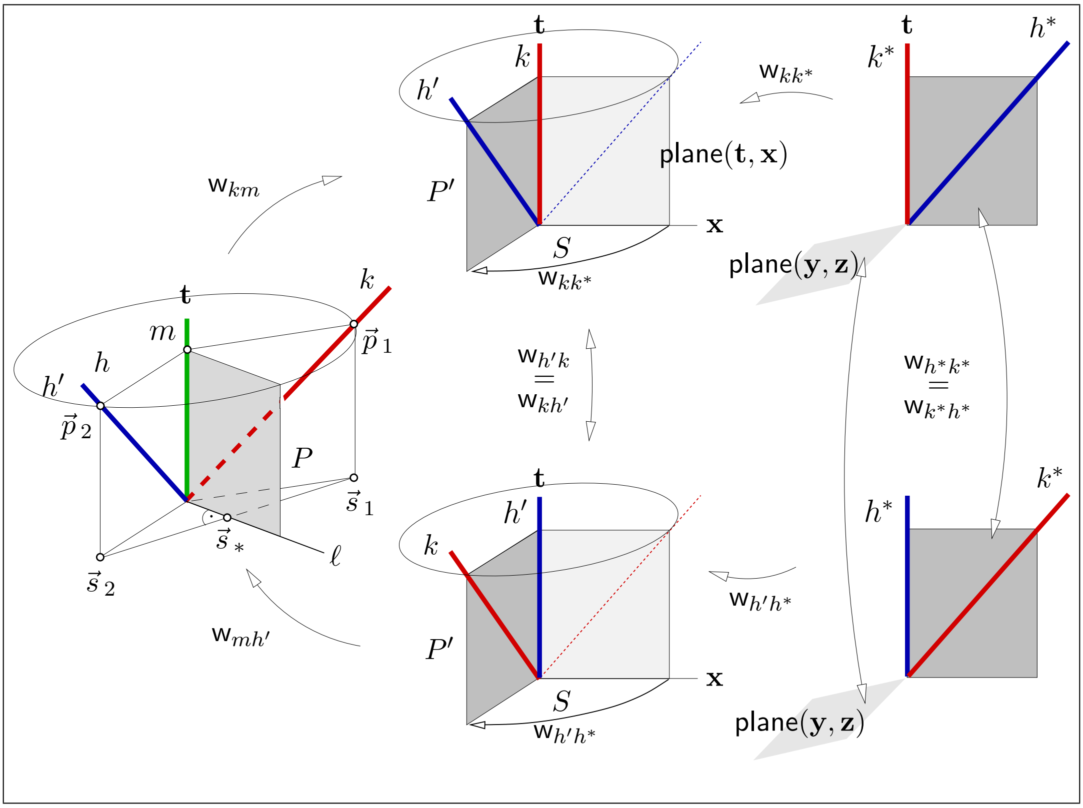


Figure 15: Illustration for the proof of Theorem 6.4 (Same-Speed Lemma)

Let P be the plane containing $(0, \vec{s}^{\,*})$ and $\mathbf{t}$. Since $(0, \vec{s}^{\,*})$ and $\mathbf{t}$ are pointwise fixed by R, it follows that the whole of P is likewise fixed pointwise by R; see Figure 15.

We claim that $\mathsf{w}_{h'k}$ $(= \mathsf{w}_{kh'})$ leaves the plane $\mathsf{w}_{km}[P]$ pointwise fixed. To see this, choose any $\vec{p} \in \mathsf{w}_{km}[P]$. By (6.12), $\mathsf{w}_{kh'}(\vec{p}\,) = (\mathsf{w}_{km} \circ R \circ \mathsf{w}_{mk})(\vec{p}\,)$. But $\mathsf{w}_{mk}(\vec{p}\,) \in \mathsf{w}_{mk}[\mathsf{w}_{km}[P]] = P$, so $R(\mathsf{w}_{mk}(\vec{p}\,)) = \mathsf{w}_{mk}(\vec{p}\,)$. It follows that

$$\mathsf{w}_{kh'}(\vec{p}\,) = (\mathsf{w}_{km} \circ \mathsf{w}_{mk})(\vec{p}\,) = \vec{p}$$

as stated.

We know that $\mathsf{w}_{h'k}$ is a bijective collineation by Theorem 6.2 (Line-to-Line Lemma) and that it leaves $\vec{\mathsf{o}}$ fixed by Lemma 6.4.1 (Translation to IOb$_o$) because $h', k \in IOb_o$. So, by Lemma 6.3.4 (Affine), $\mathsf{w}_{h'k}$ is a linear transformation composed with a map induced by a field automorphism. But since $\mathsf{w}_{h'k}$ leaves the plane $\mathsf{w}_{km}[P]$ pointwise fixed, the automorphism component must be the identity, and we deduce that $\mathsf{w}_{h'k}$ is a linear transformation.

By $\mathsf{wl}_m(k) \neq \mathsf{wl}_m(h) = \mathsf{wl}_m(h')$ and Lemma 6.3.5 (Equal Worldlines), we have that $\mathsf{wl}_{h'}(k) \neq \mathsf{wl}_{h'}(h') = \mathbf{t}$. By Lemma 6.3.3 ($\mathrm{IOb}_o$), we have that $\vec{\mathsf{o}} \in \mathsf{wl}_k(h')$. Let P' be the plane determined by the time-axis and $\mathsf{wl}_k(h')$ $(= \mathsf{wl}_{h'}(k))$ and let S be a spatial rotation that takes the tx-plane to P', see Figure 15. Such a rotation exists by Lemma 6.4.2 (Vertical Plane Rotation). Choose k^*, h^* such that $\mathsf{w}_{kk^*} = \mathsf{w}_{h'h^*} = S$ (these exist by $\mathsf{AxRelocate}$). Then

$$\mathsf{w}_{h^*k^*} = \mathsf{w}_{h^*h'} \circ \mathsf{w}_{h'k} \circ \mathsf{w}_{kk^*} = S^{-1} \circ \mathsf{w}_{h'k} \circ S \tag{6.14}$$

and hence

$$\mathsf{w}_{k^*h^*} = (S^{-1} \circ \mathsf{w}_{h'k} \circ S)^{-1} = S^{-1} \circ \mathsf{w}_{h'k} \circ S$$

because $\mathsf{w}_{h'k} = \mathsf{w}_{kh'}$. Therefore, $\mathsf{w}_{h^*k^*} = \mathsf{w}_{k^*h^*}$ and $\mathsf{w}_{h^*k^*}$ is a linear transformation since S^{-1}, $\mathsf{w}_{h'k}$, and S are linear.

To prove that there is κ such that $\mathsf{w}_{h'k} \in {}_\kappa\mathsf{Iso}$, it is therefore enough to show that there is κ such that $\mathsf{w}_{h^*k^*} \in {}_\kappa\mathsf{Iso}$, because spatial rotations $S, S^{-1} \in {}_\kappa\mathsf{Iso}$ for every κ.

The worldview transformation $\mathsf{w}_{h'k}$ leaves plane P' fixed because it takes $\mathbf{t}$ and $\mathsf{wl}_k(h')$ to $\mathsf{wl}_{h'}(k)$ and $\mathbf{t}$, respectively, and P' is the unique plane that contains $\mathbf{t}$ and $\mathsf{wl}_k(h') = \mathsf{wl}_{h'}(k)$. By this and (6.14), we have that $\mathsf{w}_{h^*k^*}$ maps the tx-plane to itself. Hence, by Theorem 6.3 (tx-Plane Lemma) $\mathsf{w}_{h^*k^*}$ also takes the yz-plane to itself and there is $\lambda > 0$ such that for every $\vec{p} \in \mathsf{plane}(\mathbf{y}, \mathbf{z})$, $|\mathsf{w}_{h^*k^*}(\vec{p}\,)| = \lambda\,|\vec{p}\,|$. But now, for every $\vec{p} \in \mathsf{plane}(\mathbf{y}, \mathbf{z})$, we have

$$|\vec{p}\,| = |(\mathsf{w}_{k^*h^*} \circ \mathsf{w}_{h^*k^*})(\vec{p}\,)| = |(\mathsf{w}_{h^*k^*} \circ \mathsf{w}_{h^*k^*})(\vec{p}\,)| = \lambda^2\,|\vec{p}\,|\,.$$

Thus, $\lambda^2 = 1$, whence $\lambda = 1$ (as $\lambda > 0$).

This means that $\mathsf{w}_{h^*k^*}$ preserves Euclidean length in $\mathsf{plane}(\mathbf{y}, \mathbf{z})$.

We have proven so far that $\mathsf{w}_{h^*k^*} = \mathsf{w}_{k^*h^*}$, that $\mathsf{w}_{h^*k^*}$ is a linear transformation taking $\mathsf{plane}(\mathbf{t}, \mathbf{x})$ to $\mathsf{plane}(\mathbf{t}, \mathbf{x})$ and $\mathsf{plane}(\mathbf{y}, \mathbf{z})$ to $\mathsf{plane}(\mathbf{y}, \mathbf{z})$, and that it preserves Euclidean length in $\mathsf{plane}(\mathbf{y}, \mathbf{z})$. It remains to show that $\mathsf{w}_{h^*k^*} \in {}_\kappa\mathsf{Iso}$.

We have already seen that $\vec{\mathsf{o}} \in \mathsf{wl}_{h'}(k) \neq \mathbf{t}$. Thus, $\mathsf{speed}_{h'}(k) \neq 0$. By Lemma 6.4.3 (LinTriv $\Rightarrow$ Same Speed) and the fact that $\mathsf{w}_{h'h^*}$ and w_{kk^*} are spatial rotations (hence linear trivial transformations), we have that $\mathsf{speed}_{h^*}(k^*) = \mathsf{speed}_{h'}(k^*) = \mathsf{speed}_{h'}(k)$. Thus, $\mathsf{speed}_{h^*}(k^*) \neq 0$.

We will choose κ so that

$$\left\|\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}})\right\|_\kappa^2 = 1.$$

We can do this because we know that $\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}}) \in \mathsf{plane}(\mathbf{t}, \mathbf{x})$, so we can write $\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}}) = (t_e, x_e, 0, 0)$ for some t_e and x_e, and we know that $x_e \neq 0$ because

$\mathsf{speed}_{h^*}(k^*) \neq 0$ and $\vec{\mathsf{o}}, \mathsf{w}_{h^*k^*}(\mathbf{t}) \in \mathsf{wl}_{h^*}(k^*)$. So we can take $\kappa := (t_e^2 - 1)/x_e^2$, because then

$$\|\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}})\|_\kappa^2 = t_e^2 - \kappa x_e^2 = t_e^2 - \frac{(t_e^2-1)}{x_e^2} x_e^2 = 1,$$

as required.

It follows that $\|\mathsf{w}_{h^*k^*}(\vec{p}\,)\|_\kappa^2 = \|\vec{p}\,\|_\kappa^2$ for every $\vec{p} \in \mathsf{plane}(\mathbf{t}, \mathbf{x})$, i.e. $\mathsf{w}_{h^*k^*}$ preserves κ-length in the tx-plane. To see why, let $\vec{p} \in \mathsf{plane}(\mathbf{t}, \mathbf{x})$. Notice that $\vec{p}$ can be written as some linear combination $\vec{p} = \lambda\vec{\mathsf{t}} + \mu\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}})$. From this and the fact that $\mathsf{w}_{h^*k^*} = \mathsf{w}_{k^*h^*}$ is a linear transformation, we have

$$\mathsf{w}_{h^*k^*}(\vec{p}\,) = \mathsf{w}_{h^*k^*}(\lambda\vec{\mathsf{t}} + \mu\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}})) = \lambda\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}}) + \mu\vec{\mathsf{t}}.$$

Writing $\vec{p}_\dagger = \mathsf{w}_{h^*k^*}(\vec{p}\,)$ and recalling that $\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}}) = (t_e, x_e, 0, 0)$, we have

$$\vec{p} = \lambda(1,0,0,0) + \mu(t_e, x_e, 0, 0) \qquad \text{and} \qquad \vec{p}_\dagger = \lambda(t_e, x_e, 0, 0) + \mu(1,0,0,0)$$

and now direct calculation (using $\kappa = (t_e^2 - 1)/x_e^2$) shows that

$$\|\vec{p}\,\|_\kappa^2 = (\lambda + \mu t_e)^2 - \frac{(t_e^2-1)}{x_e^2}\mu^2 x_e^2 = \lambda^2 + 2t_e\lambda\mu + \mu^2$$

and likewise

$$\|\vec{p}_\dagger\|_\kappa^2 = (\lambda t_e + \mu)^2 - \frac{(t_e^2-1)}{x_e^2}\lambda^2 x_e^2 = \lambda^2 + 2t_e\lambda\mu + \mu^2,$$

whence $\left\|\vec{p}^{\,2}\right\|_\kappa = \|\vec{p}_\dagger\|_\kappa^2 = \|\mathsf{w}_{h^*k^*}(\vec{p}\,)\|_\kappa$ as claimed.

Next, we are going to prove that $\mathsf{w}_{k^*h^*}$ preserves the κ-length. To prove this, let $\vec{p} = (t, x, y, z)$ be an arbitrary point in Q^4 and let $(\hat{t}, \hat{x}, \hat{y}, \hat{z}) = \mathsf{w}_{h^*k^*}(\vec{p}\,)$. By linearity, we have

$$(\hat{t}, \hat{x}, \hat{y}, \hat{z}) = \mathsf{w}_{h^*k^*}(t, x, y, z) = \mathsf{w}_{h^*k^*}(t, x, 0, 0) + \mathsf{w}_{h^*k^*}(0, 0, y, z),$$

whence $(\hat{t}, \hat{x}, 0, 0) = \mathsf{w}_{h^*k^*}(t, x, 0, 0)$ and $(0, 0, \hat{y}, \hat{z}) = \mathsf{w}_{h^*k^*}(0, 0, y, z)$, because $\mathsf{w}_{h^*k^*}$ preserves both the tx- and yz-planes. We also have that

$$\hat{t}^2 - \kappa\hat{x}^2 = t^2 - \kappa x^2 \qquad \text{and} \qquad \hat{y}^2 + \hat{z}^2 = y^2 + z^2$$

because $\mathsf{w}_{h^*k^*}$ preserves the κ-length in the tx-plane and preserves the Euclidean length in the yz-plane. It follows immediately that

$$(\hat{t}^2 - \kappa\hat{x}^2) - \kappa(\hat{y}^2 + \hat{z}^2) = (t^2 - \kappa x^2) - \kappa(y^2 + z^2),$$

or in other words, $\|\vec{p}\,\|_\kappa^2 = \|\mathsf{w}_{h^*k}(\vec{p}\,)\|_\kappa$, and so $\mathsf{w}_{h^*k^*}$ preserves the κ-length.

Therefore, if $\kappa \neq 0$, then $\mathsf{w}_{h^*k^*}$ is a linear κ-isometry, so $\mathsf{w}_{h^*k^*} \in {}_\kappa\mathsf{Iso}$, and we are done.

Suppose, finally, that $\kappa = 0$. We will prove that $\mathsf{w}_{h^*k^*}$ is a linear 0-isometry. Recall that $\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}}) = (t_e, x_e, 0, 0)$ and $\kappa = (t_e^2 - 1)/x_e^2$. Since $\kappa = 0$, we have $t_e = \pm 1$, and hence $\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}}) = (\pm 1, x_e, 0, 0)$. Thus, $(0, x_e, 0, 0) = \mathsf{w}_{h^*k^*}(\vec{\mathsf{t}}) \mp \vec{\mathsf{t}}$. This and the fact that $\mathsf{w}_{h^*k^*}$ is both linear and self-inverse now yields

$$
\begin{aligned}
\mathsf{w}_{h^*k^*}(0, x_e, 0, 0) &= \mathsf{w}_{h^*k^*}(\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}}) \mp \vec{\mathsf{t}}) \\
&= \mathsf{w}_{h^*k^*}(\mathsf{w}_{h^*k^*}(\vec{\mathsf{t}})) \mp \mathsf{w}_{h^*k^*}(\vec{\mathsf{t}}) \\
&= \vec{\mathsf{t}} \mp \mathsf{w}_{h^*k^*}(\vec{\mathsf{t}}) \\
&= \mp (0, x_e, 0, 0).
\end{aligned} \tag{6.15}
$$

Writing $f := \mathsf{w}_{h^*k^*}$ we have already shown that f preserves κ-length, so for $\kappa = 0$ we have $f(\vec{p}\,)_t^2 = \|f(\vec{p}\,)\|_0^2 = \|\vec{p}\,\|_0^2 = \vec{p}_t^{\,2}$ for every $\vec{p} \in Q^4$. By (5.1), it only remains to show that $|f(\vec{p}\,)_s|^2 = |\vec{p}_s|^2$ when $\vec{p}_t = 0$. However, we know that f maps the yz-plane to itself and preserves Euclidean length in that plane, and that it simply reverses or preserves x-coordinates by (6.15). Hence, f also preserves Euclidean length in the xyz-hyperplane. Thus, $\mathsf{w}_{h^*k^*}$ is a linear 0-isometry.

This completes the proof of (a).

Proof of (b). By (6.11) (which says that $\mathsf{w}_{hh'}$ is a linear trivial transformation) and by Lemma 6.4.3 (LinTriv ⇒ Same Speed), for every $j \in IOb$, we have that

$$
\begin{aligned}
\mathsf{speed}_j(h) &= \mathsf{speed}_j(h') \text{ and} \qquad (6.16)\\
\mathsf{speed}_h(j) &= \mathsf{speed}_{h'}(j), \qquad (6.17)
\end{aligned}
$$

and so

$$\mathsf{speed}_k(h) \overset{(6.16)}{=\!=} \mathsf{speed}_k(h') \overset{(6.13)}{=\!=} \mathsf{speed}_{h'}(k) \overset{(6.17)}{=\!=} \mathsf{speed}_h(k)$$

as required.

Proof of (c). First we show that

$$\mathsf{wl}_k(m) = \mathsf{wl}_{h'}(m). \tag{6.18}$$

To do so, recall that $\mathsf{w}_{mh'} = R \circ \mathsf{w}_{mk}$ (by $k \overset{R}{\rightsquigarrow}_m h'$). It follows that $\mathsf{w}_{h'm} = \mathsf{w}_{km} \circ R^{-1}$, and hence (because the time-axis $\mathbf{t}$ is fixed under spatial rotations),

$$\mathsf{wl}_{h'}(m) = \mathsf{w}_{h'm}[\mathbf{t}] = (\mathsf{w}_{km} \circ R^{-1})[\mathbf{t}] = \mathsf{w}_{km}[\mathbf{t}] = \mathsf{wl}_k(m)$$

as claimed. Consequently,

$$\mathsf{speed}_k(m) \overset{(6.18)}{=\!=} \mathsf{speed}_{h'}(m) \overset{(6.17)}{=\!=} \mathsf{speed}_h(m).$$

This completes the proof. □

6.5 Fundamental Lemma

Theorem 6.5 (Fundamental Lemma)**.** Assume $\mathsf{KIN}+\mathsf{AxIsotropy}+\neg\exists\infty\mathsf{Speed}$. Then for every $k, m \in IOb_o$ with $\mathsf{speed}_k(m) > 0$, there is a positive $\varepsilon \in Q$ such that for every non-negative $v \leq \mathsf{speed}_k(m) + \varepsilon$, there is some $h \in IOb_o$ with $\mathsf{speed}_k(h) = v$ and $\mathsf{speed}_m(k) = \mathsf{speed}_m(h)$.

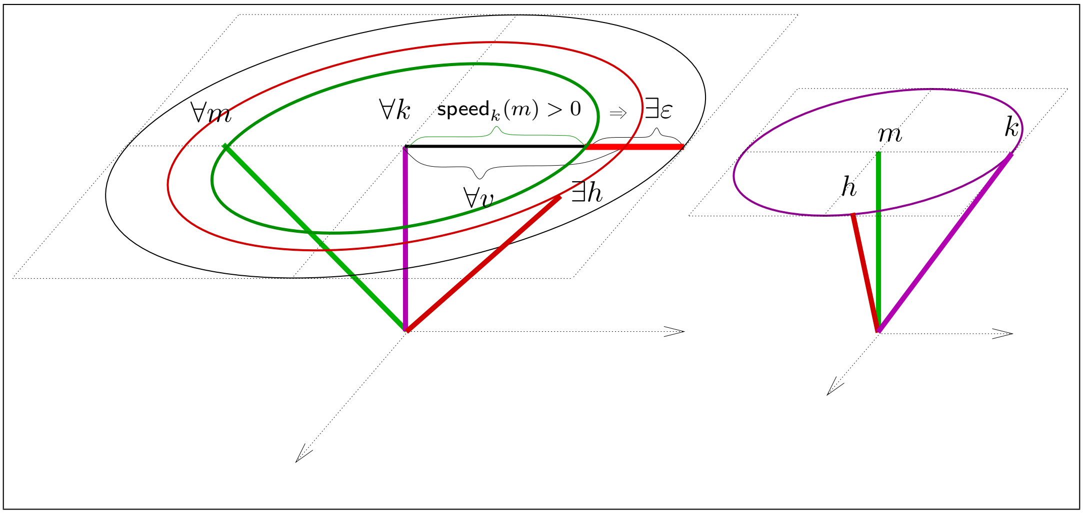


Figure 16: Figure illustrating Theorem 6.5 (Fundamental Lemma).

We first show that observers can be found which satisfy certain standard configurations; see Figure 17.

6.5.1 Supporting lemmas

The supporting lemmas can be informally described as:

Lemma 6.5.1 (Configuration)

If two observers k and m are moving at any speed $u > 0$ relative to one another, there are 'rotated versions' k^* and m^* of those observers which agree with each other as to where the tx-plane and the y-axis are located. Moreover, if u is finite, then m^* considers k^* to be moving in the positive direction of the x-axis.

Lemma 6.5.2 (Quadratic IVT)

This is a purely technical lemma stating that the Intermediate Value Theorem holds for functions of the form $f(x) = \sqrt{F(x)/G(x)}$ where F and G are quadratic polynomials over Q.

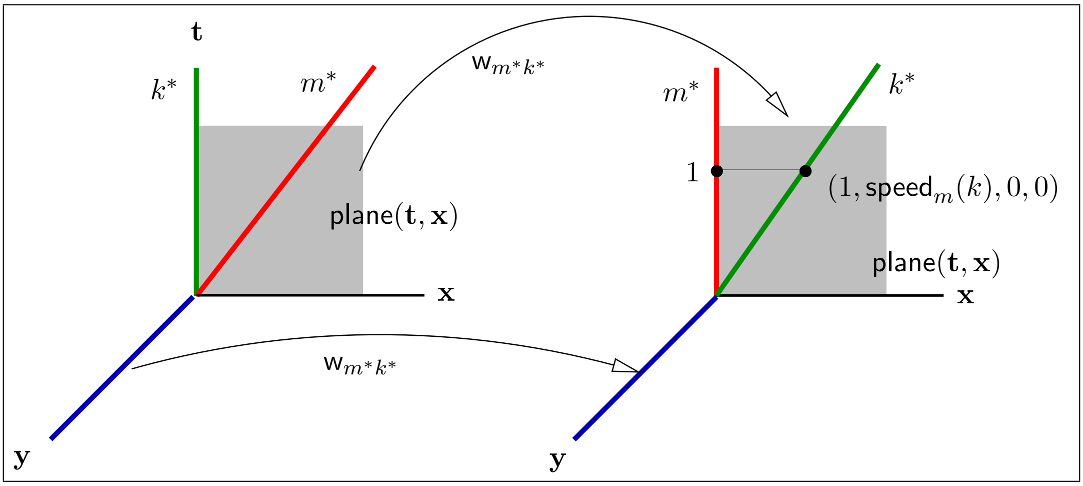


Figure 17: Illustration for Lemma 6.5.1 (Configuration).

6.5.2 Proofs of the supporting lemmas

Lemma 6.5.1 (Configuration)**.** Assume $\mathsf{KIN} + \mathsf{AxIsotropy}$. Given any $k, m \in IOb_o$ satisfying $\mathsf{speed}_m(k) \neq 0$, there exist $k^*, m^* \in IOb_o$ such that

(a) w_{k^*k} and w_{m^*m} are spatial rotations, hence[13]
$\mathsf{speed}_{m^*}(k^*) = \mathsf{speed}_m(k)$,
$\mathsf{speed}_{k^*}(h) = \mathsf{speed}_k(h)$ and $\mathsf{speed}_{m^*}(h) = \mathsf{speed}_m(h)$ for every $h \in IOb$;

(b) $\mathsf{w}_{k^*m^*}[\mathsf{plane}(\mathbf{t}, \mathbf{x})] = \mathsf{plane}(\mathbf{t}, \mathbf{x})$;

(c) $\mathsf{w}_{k^*m^*}[\mathbf{y}] = \mathbf{y}$;

(d) k^* moves in the positive direction of the x-axis according to m^*, i.e. $(1, \mathsf{speed}_{m^*}(k^*), 0, 0) \in \mathsf{wl}_{m^*}(k^*)$ and $\vec{\mathsf{o}} \in \mathsf{wl}_{m^*}(k^*)$ if $\mathsf{speed}_{m^*}(k^*) \neq \infty$.

Proof. Let us recall that, by Theorem 6.2 (Line-to-Line Lemma), worldview transformations are bijections taking lines to lines and planes to planes.

We know that $\mathsf{wl}_k(m)$ and $\mathbf{t}$ are distinct lines, because $\mathsf{speed}_k(m) \neq 0$. Since, by Lemma 6.3.3 (IOb_o), they meet at the origin, we know that $\mathsf{plane}(\mathbf{t}, \mathsf{wl}_k(m))$ is a well-defined plane, and because this plane contains the time-axis, by Lemma 6.4.2 (Vertical Plane Rotation) there must exist a spatial rotation about $\mathbf{t}$ which takes $\mathsf{plane}(\mathbf{t}, \mathbf{x})$ to $\mathsf{plane}(\mathbf{t}, \mathsf{wl}_k(m))$. By $\mathsf{AxRelocate}$ and (b) of Lemma 6.3.3 (IOb_o), there

[13] by Lemma 6.4.3 (LinTriv $\Rightarrow$ Same Speed)

is some $k^* \in IOb_o$ for which this rotation equals w_{kk^*}, so that

$$\mathsf{w}_{kk^*}[\mathsf{plane}(\mathbf{t},\mathbf{x})] = \mathsf{plane}(\mathbf{t},\mathsf{wl}_k(m)), \tag{6.19}$$

see the left-top of Figure 18.

According to Lemma 6.4.2 (Vertical Plane Rotation) there is also a spatial rotation R that takes $\mathsf{plane}(\mathbf{t},\mathsf{wl}_m(k))$ to $\mathsf{plane}(\mathbf{t},\mathbf{x})$; moreover, if $\mathsf{speed}_m(k) \neq \infty$, we can choose $\vec{p} \in \mathsf{wl}_m(k)$ such that $\vec{p}_t = 1$ and require of R that $R(\vec{p}\,) = (1, |\vec{p}_s|, 0, 0)$. In this case, because $\vec{\mathsf{o}}, \vec{p} \in \mathsf{wl}_m(k)$ and $\vec{p}_t = 1$, we have $\mathsf{speed}_m(k) = \mathsf{slope}(\mathsf{wl}_m(k)) = |\vec{p}_s|$, and so

$$R(\vec{p}\,) = (1, \mathsf{speed}_m(k), 0, 0). \tag{6.20}$$

Now let $m' \in IOb_o$ be such that $\mathsf{w}_{m'm} = R$ (such an m' exists by $\mathsf{AxRelocate}$ and (b) of Lemma 6.3.3 (IOb_o)). We will show that $\mathsf{w}_{m'k^*}$ fixes both the tx-plane and the yz-plane. By definition,

$$\mathsf{w}_{m'm}[\mathsf{plane}(\mathbf{t},\mathsf{wl}_m(k))] = \mathsf{plane}(\mathbf{t},\mathbf{x}) \tag{6.21}$$

see the left-bottom of Figure 18. If $\mathsf{speed}_m(k) \neq \infty$, by $\vec{p} \in \mathsf{wl}_m(k)$, we have that $\mathsf{w}_{m'm}(\vec{p}\,) \in \mathsf{wl}_{m'}(k)$. Combining this with (6.20) tells us that

$$(1, \mathsf{speed}_m(k), 0, 0) \in \mathsf{wl}_{m'}(k) \text{ if } \mathsf{speed}_m(k) \neq \infty. \tag{6.22}$$

Notice next that the world-view transformation w_{mk} takes $\mathbf{t}$ to $\mathsf{wl}_m(k)$ and $\mathsf{wl}_k(m)$ to $\mathbf{t}$, respectively. Therefore,

$$\mathsf{w}_{mk}[\mathsf{plane}(\mathbf{t},\mathsf{wl}_k(m))] = \mathsf{plane}(\mathbf{t},\mathsf{wl}_m(k)), \tag{6.23}$$

see the left-hand side of Figure 18. By (6.19), (6.23), (6.21), and the fact that $\mathsf{w}_{m'k^*} = \mathsf{w}_{m'm} \circ \mathsf{w}_{mk} \circ \mathsf{w}_{kk^*}$, we have that

$$\mathsf{w}_{m'k^*}[\mathsf{plane}(\mathbf{t},\mathbf{x})] = \mathsf{plane}(\mathbf{t},\mathbf{x}).$$

By Theorem 6.3 (tx-Plane Lemma), it follows that $\mathsf{w}_{m'k^*}[\mathsf{plane}(\mathbf{y},\mathbf{z})] = \mathsf{plane}(\mathbf{y},\mathbf{z})$. Thus, $\mathsf{w}_{m'k^*}$ fixes both the tx-plane and the yz-plane, as claimed.

Now write $\hat{\mathbf{y}} := \mathsf{w}_{m'k^*}[\mathbf{y}]$, and note that $\hat{\mathbf{y}} \subseteq \mathsf{plane}(\mathbf{y},\mathbf{z})$ because $\mathsf{w}_{m'k^*}$ preserves this plane. We can find a spatial rotation which fixes the tx-plane pointwise and takes $\hat{\mathbf{y}}$ to $\mathbf{y}$ because of the following. Let $\vec{q} \in \hat{\mathbf{y}}$ and $\vec{q}\,' \in \mathbf{y}$ be such that $|\vec{q}\,| = |\vec{q}\,'| \neq 0$. Then $\vec{q} \cdot \vec{\mathsf{t}} = \vec{\mathsf{x}} \cdot \vec{q} = \vec{q}\,' \cdot \vec{\mathsf{t}} = \vec{\mathsf{x}} \cdot \vec{q}\,' = 0$ because $\vec{q}, \vec{q}\,' \in \mathsf{plane}(\mathbf{y},\mathbf{z})$. Therefore, by Lemma 6.1.7 (Horizontal Rotation) there is a spatial rotation that takes $\vec{q}$ to $\vec{q}\,'$ and

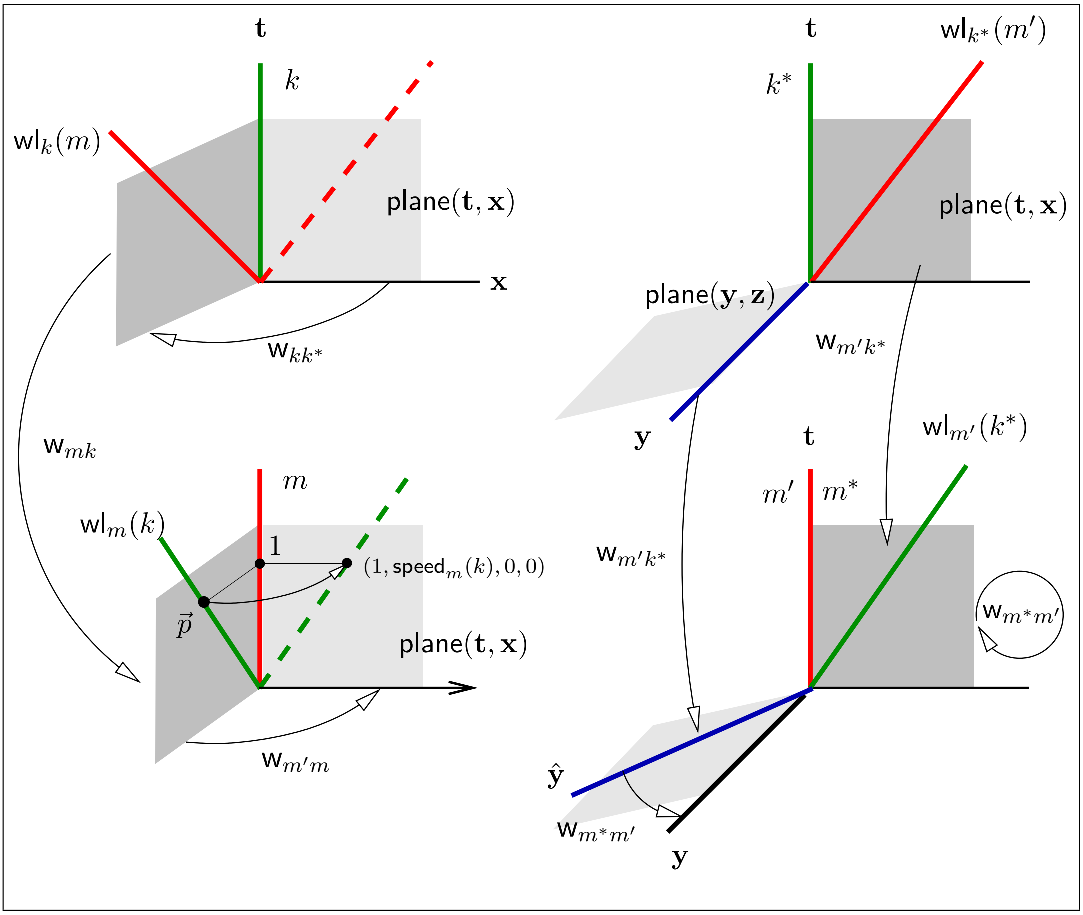


Figure 18: Illustration for the proof of Lemma 6.5.1 (Configuration)

$\vec{\mathsf{x}}$ to itself. By $\mathsf{AxRelocate}$ and Lemma 6.3.3 (IOb_o), there is some $m^* \in IOb_o$ such that $\mathsf{w}_{m^*m'}$ is this spatial rotation, see the right-bottom of Figure 18.

Notice that $\mathsf{w}_{m^*m'}$ maps $\hat{\mathbf{y}}$ to $\mathbf{y}$ (because it fixes $\vec{\mathsf{o}}$ and maps $\vec{q} \in \hat{\mathbf{y}}$ to $\vec{q}\,' \in \mathbf{y}$) and fixes $\mathsf{plane}(\mathbf{t},\mathbf{x})$ pointwise because it fixes $\vec{\mathsf{t}}$ and $\vec{\mathsf{x}}$.

In summary, we have so far shown that $\mathsf{w}_{m^*m'}$ and $\mathsf{w}_{m'm}$ are spatial rotations; and that $\mathsf{w}_{m^*m'}$ and $\mathsf{w}_{m'k^*}$ both fix the tx-plane and the yz-plane.

Proof of (a). The transformation w_{k^*k} is a spatial rotation by definition. Since $\mathsf{w}_{m^*m} = \mathsf{w}_{m^*m'} \circ \mathsf{w}_{m'm}$ is a composition of two spatial rotations, it is also a spatial rotation. By Lemma 6.4.3 (LinTriv $\Rightarrow$ Same Speed), $\mathsf{speed}_{m^*}(k^*) = \mathsf{speed}_m(k^*) = \mathsf{speed}_m(k)$, $\mathsf{speed}_{k^*}(h) = \mathsf{speed}_k(h)$ and $\mathsf{speed}_{m^*}(h) = \mathsf{speed}_m(h)$ for every $h \in IOb$.

Proof of (b). Since $\mathsf{w}_{m^*k^*} = \mathsf{w}_{m^*m'} \circ \mathsf{w}_{m'k^*}$ and both $\mathsf{w}_{m^*m'}$ and $\mathsf{w}_{m'k^*}$ fix the tx-plane, $\mathsf{w}_{m^*k^*}$ and its inverse $\mathsf{w}_{k^*m^*}$ also fix the tx-plane.

Proof of (c). We have $\mathbf{y} = \mathsf{w}_{m^*m'}[\hat{\mathbf{y}}] = \mathsf{w}_{m^*m'}[\mathsf{w}_{m'k^*}[\mathbf{y}]] = \mathsf{w}_{m^*k^*}[\mathbf{y}]$, so $\mathsf{w}_{m^*k^*}$ and its inverse $\mathsf{w}_{k^*m^*}$ fix the y-axis.

Proof of (d). It is already clear that $\vec{\mathsf{o}} \in \mathsf{wl}_{m^*}(k^*)$, by Lemma 6.3.3 (IOb$_o$). We need to show that $(1, \mathsf{speed}_{m^*}(k^*), 0, 0) \in \mathsf{wl}_{m^*}(k^*)$ as well.

By (6.21) and $\mathsf{wl}_{m'}(k) = \mathsf{w}_{m'm}[\mathsf{wl}_m(k)]$, we have that $\mathsf{wl}_{m'}(k) \subseteq \mathsf{plane}(\mathbf{t}, \mathbf{x})$. Because $\mathsf{w}_{m^*m'}$ fixes $\mathsf{plane}(\mathbf{t}, \mathbf{x})$ pointwise and takes $\mathsf{wl}_{m'}(k)$ to $\mathsf{wl}_{m^*}(k)$, we therefore have $\mathsf{wl}_{m^*}(k) = \mathsf{wl}_{m'}(k)$. By Lemma 6.4.3 (LinTriv $\Rightarrow$ Same Speed), $\mathsf{wl}_{m^*}(k^*) = \mathsf{wl}_{m^*}(k)$ because $\mathsf{w}_{k^*k} \in \mathsf{SRot}$ is a linear trivial transformation. Consequently, $\mathsf{wl}_{m^*}(k^*) = \mathsf{wl}_{m^*}(k) = \mathsf{wl}_{m'}(k)$. Now assume that $\mathsf{speed}_m(k) \neq \infty$. Then (6.22) tells us that $(1, \mathsf{speed}_m(k), 0, 0) \in \mathsf{wl}_{m'}(k) = \mathsf{wl}_{m^*}(k^*)$. By (a), $\mathsf{speed}_m(k) = \mathsf{speed}_{m^*}(k^*)$. Therefore, $(1, \mathsf{speed}_{m^*}(k^*), 0, 0) \in \mathsf{wl}_{m^*}(k^*)$, as required.

This completes the proof. $\square$

Remark 6.2. Using the fact that any real-closed field is elementarily equivalent to the field of real numbers (i.e. they satisfy the same first-order logic formulas), it is easy to show that an ordered field is real-closed iff it satisfies the Intermediate Value Theorem for every polynomial function. However, for arbitrary ordered fields (e.g., the field $\mathbb{Q}$ of rationals) the Intermediate Value Theorem can fail even for quadratic functions: if $F(x) = x^2 - 2$, then despite the fact that $F(0) < 0 < F(2)$ there is no $c \in \mathbb{Q}$ for which $F(c) = 0$.

In the proof of Theorem 6.5 (Fundamental Lemma) below, we will need the following lemma stating that the Intermediate Value Theorem holds for a specific class of algebraic functions defined over Euclidean fields.

Lemma 6.5.2 (Quadratic IVT)**.** Assume $\mathsf{AxEField}$, and let F and G be quadratic functions on Q.[14] Let $a < b$ be values in Q and suppose $F(x) \geq 0$ and $G(x) > 0$ for all $x \in [a, b]$. Let $g : [a, b] \to Q$ be the function $g(x) := \sqrt{F(x)/G(x)}$. Then given any y between $g(a)$ and $g(b)$, there exists $c \in [a, b]$ such that $g(c) = y$.

Proof. If $g(a) = y$ or $g(b) = y$ the proof is trivial, so suppose that y lies strictly between $g(a)$ and $g(b)$ and consider the quadratic function $p(x) = F(x) - y^2 G(x) \equiv [g(x)^2 - y^2]G(x)$. Because y^2 lies strictly between $g(a)^2$ and $g(b)^2$, the values $p(a) = [g(a)^2 - y^2]G(a)$ and $p(b) = [g(b)^2 - y^2]G(b)$ are both non-zero and have opposite signs.

[14] $F : Q \to Q$ is called a ***quadratic function*** if there are $p, q, r \in Q$ such that $F(x) = px^2 + qx + r$ for every $x \in Q$.

We will show that there exists some $c \in (a, b)$ for which $p(c) = 0$. Because p is quadratic, it can be written in the form $p(x) = \alpha x^2 + \beta x + \gamma$. We know that p is not constant because $p(a) \neq p(b)$, so α and β cannot both be zero. If $\alpha = 0$, then $\beta \neq 0$ and $p(x) = \beta x + \gamma$ is a linear function for which a suitable c can trivially be found. Suppose, then, that $\alpha \neq 0$. Then we can rewrite p as $p(x) = \alpha \left[(x + \beta/2\alpha)^2 - (\beta^2 - 4\alpha\gamma)/4\alpha^2)\right]$, and now the fact that $p(x)$ can be both positive and negative implies immediately that the discriminant $\Delta := (\beta^2 - 4\alpha\gamma)$ is positive, whence p can be factorised over Q with the usual quadratic roots $x_1 := (-\beta + \sqrt{\Delta})/2\alpha$ and $x_2 := (-\beta - \sqrt{\Delta})/2\alpha$. Writing $p(x) = \alpha(x - x_1)(x - x_2)$ it is now easy to see from $p(a)p(b) < 0$ that at least one of these roots must lie strictly between a and b, and we set c equal to this root.

Given the definition of p it now follows from $p(c) = 0$ that $0 = [g(c)^2 - y^2]G(c)$. Because G is positive on $[a, b]$ we can divide through by $G(c)$, whence $g(c)^2 = y^2$. By construction, however, we know that $g(x) \geq 0$ for all $x \in [a, b]$, so both $g(c)$ and y (which lies between $g(a)$ and $g(b)$) are non-negative. We have therefore found a value $c \in (a, b)$ satisfying $g(c) = y$, as required.

□

6.5.3 Main proof

We now complete the proof of Theorem 6.5 (Fundamental Lemma).

Proof of Theorem 6.5 (Fundamental Lemma). Choose any $k, m \in IOb_o$ satisfying $\mathsf{speed}_k(m) > 0$. Then $\mathbf{t}$ and $\mathsf{wl}_k(m)$ are distinct lines intersecting in $\vec{\mathsf{o}}$. Therefore, their w_{mk}-images, $\mathsf{wl}_m(k)$ and $\mathbf{t}$, are distinct intersecting lines. Hence, $\mathsf{speed}_m(k) > 0$. By Lemma 6.5.1 (Configuration) and $\neg\exists\infty\mathsf{Speed}$, we can assume that

- $\mathsf{w}_{km}[\mathsf{plane}(\mathbf{t}, \mathbf{x})] = \mathsf{plane}(\mathbf{t}, \mathbf{x})$;
- $\mathsf{w}_{km}[\mathbf{y}] = \mathbf{y}$; and
- k moves in the positive direction of the x-axis according to m, i.e. $(1, \mathsf{speed}_m(k), 0, 0) \in \mathsf{wl}_m(k)$ and $\vec{\mathsf{o}} \in \mathsf{wl}_m(k)$.

Let $r := \mathsf{speed}_m(k)$, and note that $r \neq \infty$ by $\neg\exists\infty\mathsf{Speed}$. Then $(1, r, 0, 0) \in \mathsf{wl}_m(k)$.

For each $x \in [0, r]$, let ℓ_x be the line containing $\vec{\mathsf{o}}$ and the point $(1, x, \sqrt{r^2 - x^2}, 0)$. Observe that $\mathsf{slope}(\ell_x) = r$ for all such x, and that $\ell_r = \mathsf{wl}_m(k)$; see Figure 19. Since $\mathsf{wl}_m(k)$ is an m-observer line, by Theorem 6.1 (Observer Lines Lemma) every ℓ_x is an m-observer line, hence by Lemma 6.1.5 (Transformed Observer Lines) every $\mathsf{w}_{km}[\ell_x]$

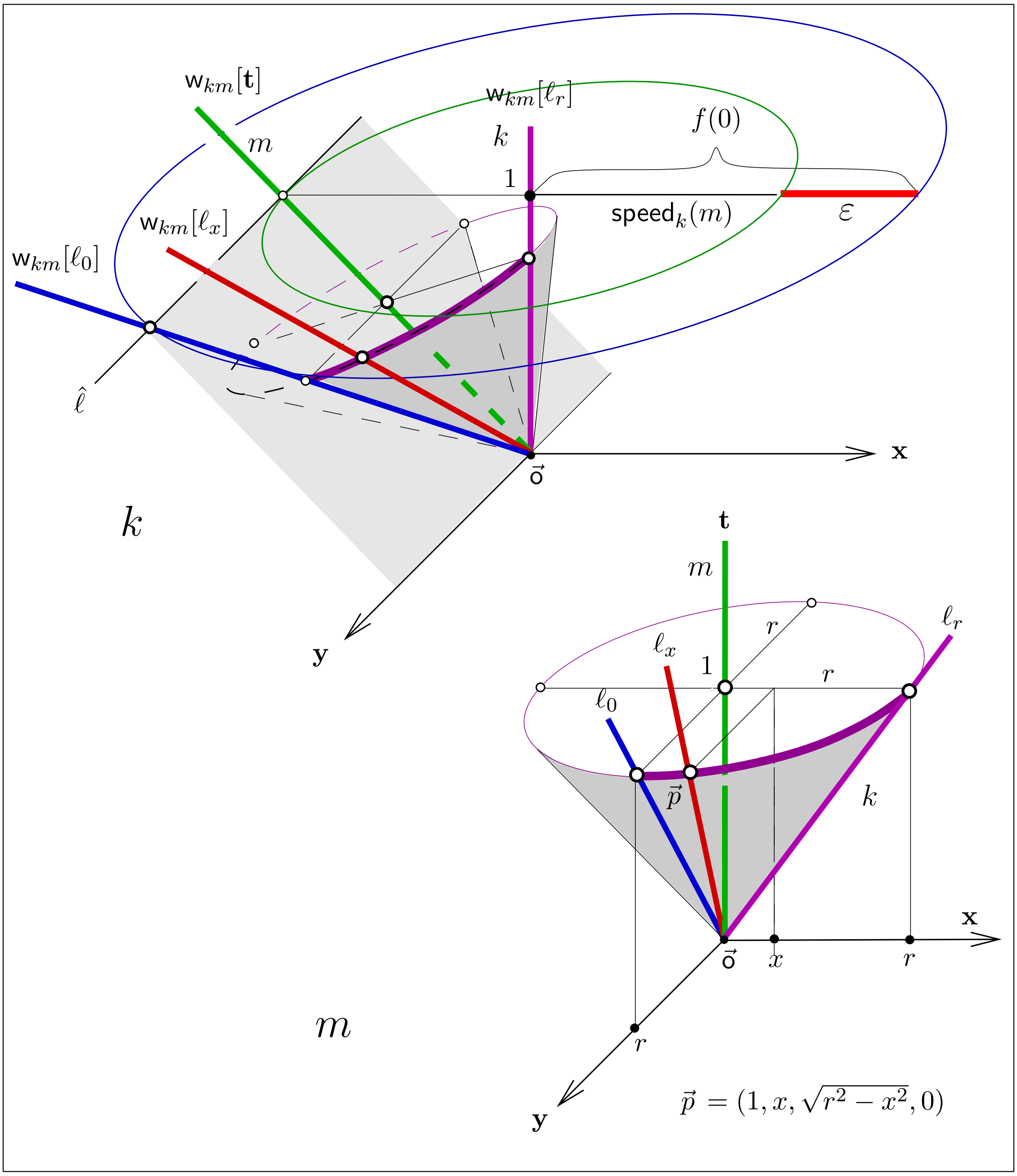


Figure 19: Illustration for the proof of Theorem 6.5 (Fundamental Lemma)

is a k-observer line. It follows from $\neg\exists\infty\mathsf{Speed}$ that the function $f : [0, r] \to Q$ given by

$$f(x) := \mathsf{slope}(\mathsf{w}_{km}[\ell_x])$$

is well-defined, and it is easy to see that

$$f(r) = \mathsf{slope}(\mathsf{w}_{km}[\ell_r]) = \mathsf{slope}(\mathsf{w}_{km}[\mathsf{wl}_m(k)]) = \mathsf{slope}[\mathbf{t}] = 0.$$

We will prove that $f(0) > \mathsf{speed}_k(m)$.

Recall that w_{km} is a bijection taking planes to planes by Theorem 6.2 (Line-to-Line Lemma). Since $\ell_0 \subseteq \mathsf{plane}(\mathbf{t}, \mathbf{y})$ and w_{km} fixes the y-axis, we have

$$\mathsf{w}_{km}[\ell_0] \subseteq \mathsf{plane}(\mathsf{w}_{km}[\mathbf{t}], \mathsf{w}_{km}[\mathbf{y}]) = \mathsf{plane}(\mathsf{w}_{km}[\mathbf{t}], \mathbf{y}).$$

Let us write $\hat{P} := \mathsf{plane}(\mathsf{w}_{km}[\mathbf{t}], \mathbf{y})$.

Because $\mathsf{slope}(\mathsf{w}_{km}[\mathbf{t}]) = \mathsf{speed}_k(m)$ cannot be infinite (by $\neg\exists\infty\mathsf{Speed}$), there exists some $\vec{s} \in Q^3$ and such that $(1, \vec{s}\,) \in \mathsf{w}_{km}[\mathbf{t}]$. And because $\mathbf{t}$ is a subset of the tx-plane (which is fixed by w_{km}), we know that $\mathsf{w}_{km}[\mathbf{t}] \subset \mathsf{plane}(\mathbf{t}, \mathbf{x})$. Thus, the y- and z-components of $\vec{s}$ must both be zero, and there exists some $\hat{x} \in Q$ with $(1, \hat{x}, 0, 0) \in \mathsf{w}_{km}[\mathbf{t}]$.

By Lemma 6.3.3 (IOb$_o$), we have $\mathsf{w}_{km}\,(\vec{\mathsf{o}}\,) = \vec{\mathsf{o}}\,$, so we know that $\vec{\mathsf{o}} \in \mathsf{w}_{km}[\mathbf{t}] \subseteq \hat{P}$. It follows that $\hat{P} = \mathsf{plane}(\mathsf{w}_{km}[\mathbf{t}], \mathbf{y})$ is the unique plane containing both the origin and the line $\hat{\ell} := \{(1, \hat{x}, y, 0) : y \in Q\}$, and every line in this plane which has finite slope and passes through the origin must intersect $\hat{\ell}$ at some point $(1, \hat{x}, y, 0)$ where $y \in Q$. The line of this form with the smallest slope is the one which minimises the value of $\hat{x}^2 + y^2$, and since this is minimal precisely when $y = 0$ the line in this plane through the origin which has the least slope is $\mathsf{w}_{km}[\mathbf{t}]$. At the same time, we know that $\mathsf{w}_{km}[\ell_0]$ is a line in this plane, and that $\mathsf{w}_{km}[\ell_0] \neq \mathsf{w}_{km}[\mathbf{t}]$ because w_{km} is a bijection and $\ell_0 \neq \mathbf{t}$. Hence, $\mathsf{slope}(\mathsf{w}_{km}[\ell_0]) > \mathsf{slope}(\mathsf{w}_{km}[\mathbf{t}])$. Therefore, we have

$$f(0) = \mathsf{slope}(\mathsf{w}_{km}[\ell_0]) > \mathsf{slope}(\mathsf{w}_{km}[\mathbf{t}]) = \mathsf{speed}_k(m).$$

Thus, $f(0) > \mathsf{speed}_k(m)$ as claimed.

Let $\varepsilon = f(0) - \mathsf{speed}_k(m)$. We will prove that for this choice of ε the conclusion of the lemma holds, i.e. that for every non-negative $v \leq \mathsf{speed}_k(m) + \varepsilon$ there is $h \in IOb_o$ such that $\mathsf{speed}_k(h) = v$ and $\mathsf{speed}_m(k) = \mathsf{speed}_m(h)$.

To prove this, choose any $v \in Q$ satisfying $0 \leq v \leq \mathsf{speed}_k(m) + \varepsilon = f(0)$, and recall that $f(r) = 0$. Thus,

$$f(0) \geq v \geq f(r). \tag{6.24}$$

We will use Lemma 6.5.2 (Quadratic IVT) to prove that

$$\text{there is } x \in [0, r] \text{ such that } f(x) = v. \tag{6.25}$$

We know from Theorem 6.2 (Line-to-Line Lemma) that w_{km} is a bijection taking lines to lines. It also preserves the origin since $m, k \in IOb_o$. Hence, by Lemma 6.3.4 (Affine), there exists some linear transformation L and automorphism φ of $(Q, +, \cdot, 0, 1, \leq)$ for which $\mathsf{w}_{km} = L \circ \widetilde{\varphi}$.

By construction, $\widetilde{\varphi}$ maps each coordinate axis to itself, so it takes $\mathsf{plane}(\mathbf{t}, \mathbf{x})$ to $\mathsf{plane}(\mathbf{t}, \mathbf{x})$ and $\mathbf{y}$ to $\mathbf{y}$. We have already seen that w_{km} does likewise, and so the same must be true of L.

We can therefore find $a, b, c, d, \lambda \in Q$ with $\lambda \neq 0$ such that, for every $t, x, y \in Q$,

$$\mathsf{w}_{km}(t, x, y, 0) = (a\varphi(t) + b\varphi(x), c\varphi(t) + d\varphi(x), \lambda\varphi(y), 0).$$

As φ is an automorphism of $(Q, +, \cdot, 0, 1, \leq)$, it follows that $\varphi(1) = 1$; that for every $x \in [0, r]$ we have $\varphi(x) \leq \varphi(r)$; and that

$$\mathsf{w}_{km}\left(1, x, \sqrt{r^2 - x^2}, 0\right) = \left(a + b\varphi(x), c + d\varphi(x), \lambda\sqrt{\varphi(r)^2 - \varphi(x)^2}, 0\right).$$

By definition, for every $x \in [0, r]$, ℓ_x is the line containing $\vec{\mathsf{o}}$ and $\left(1, x, \sqrt{r^2 - x^2}, 0\right)$; therefore, $\mathsf{w}_{km}[\ell_x]$ is the line containing $\vec{\mathsf{o}}$ and $\mathsf{w}_{km}\left(1, x, \sqrt{r^2 - x^2}, 0\right)$, and $f(x) \in Q$ is the slope of this line. Since this slope cannot be infinite we have, for all $x \in [0, r]$, that

$$a + b\varphi(x) \neq 0 \tag{6.26}$$

and hence

$$f(x) = \sqrt{\frac{(c + d\varphi(x))^2 + \lambda^2\left(\varphi(r)^2 - \varphi(x)^2\right)}{(a + b\varphi(x))^2}}.$$

Let $F : [0, \varphi(r)] \to Q$ and $G : [0, \varphi(r)] \to Q$ be the quadratic functions defined by

$$\begin{aligned} F(y) &:= (c + dy)^2 + \lambda^2\left(\varphi(r)^2 - y^2\right) \\ G(y) &:= (a + by)^2, \end{aligned}$$

and consider any $y \in [0, \varphi(r)]$. Because $(\varphi(r)^2 - y^2) \geq 0$, it follows immediately that $F(y) \geq 0$. Moreover, $G(y) > 0$, because φ is an ordered-field automorphism, whence $\varphi^{-1}(y) \in [0, r]$, and so by (6.26) we have $a + by = a + b\varphi(\varphi^{-1}(y)) \neq 0$. So, if we now define $g(y) = \sqrt{F(y)/G(y)}$, then g is of the correct form for Lemma 6.5.2 (Quadratic IVT) to be applied over the interval $[0, \varphi(r)]$.

Because $f = g \circ \varphi$, it follows from (6.24) and $\varphi(0) = 0$ that

$$g(0) \geq v \geq g(\varphi(r)).$$

By Lemma 6.5.2 (Quadratic IVT), there therefore exists some $y \in [0, \varphi(r)]$ with $g(y) = v$. Taking $x = \varphi^{-1}(y)$ now shows that there exists $x \in [0, r]$ satisfying $f(x) = v$, and (6.25) holds as claimed.

Accordingly, let $\tilde{x} \in [0, r]$ be such that $f(\tilde{x}) = \mathsf{slope}(\mathsf{w}_{km}[\ell_{\tilde{x}}]) = v$. Then $\ell_{\tilde{x}}$ is a line satisfying $\mathsf{slope}(\ell_{\tilde{x}}) = r = \mathsf{speed}_m(k)$ and $\mathsf{slope}(\mathsf{w}_{km}[\ell_{\tilde{x}}]) = f(\tilde{x}) = v$. Since $\ell_{\tilde{x}}$ is an m-observer line, there exists $h \in IOb_o$ with $\mathsf{wl}_m(h) = \ell_{\tilde{x}}$, and hence

- $\mathsf{wl}_k(h) = \mathsf{w}_{km}[\ell_{\tilde{x}}]$,
- $\mathsf{speed}_m(h) = \mathsf{slope}(\mathsf{wl}_m(h)) = \mathsf{slope}(\ell_{\tilde{x}}) = r = \mathsf{speed}_m(k)$, and
- $\mathsf{speed}_k(h) = \mathsf{slope}(\mathsf{wl}_k(h)) = \mathsf{slope}(\mathsf{w}_{km}[\ell_{\tilde{x}}]) = v$.

This is exactly what we had to prove, viz. there exists some h with $\mathsf{speed}_k(h) = v$ and $\mathsf{speed}_m(k) = \mathsf{speed}_m(h)$. □

6.6 Main Lemma

Theorem 6.6 (Main Lemma)**.** Assume $\mathsf{KIN} + \mathsf{AxIsotropy}$. Then there is $k \in IOb_o$ and $\kappa \in Q$ such that

$$\{\mathsf{w}_{mk} \;:\; m \in IOb_o\} \subseteq {}_\kappa\mathsf{Iso}. \tag{6.27}$$

6.6.1 Supporting lemmas

The supporting lemmas can be informally described as:

Lemma 6.6.1 (Same Speed Easy)
If m considers k and h to be moving at the same speed and w_{mk} is a κ-isometry, then so is w_{mh}.

Lemma 6.6.2 (Rest)
Two observers are at rest with respect to one another if and only if the transformation between them is trivial.

Lemma 6.6.3 (Observer Origin)
Given any point on an observer's worldline, we can find an observer with the same worldline which regards that point as its origin.

Lemma 6.6.4 (Median Observer)
Given any two observers, there is a third observer which sees them both moving with the same speed.

Lemma 6.6.5 (κ is unique)
If two observers are moving relative to one another, there exists a unique value κ for which the transformation between them is a κ-isometry.

6.6.2 Proofs of the supporting lemmas

Lemma 6.6.1 (Same Speed Easy)**.** Assume $\mathsf{KIN}+\mathsf{AxIsotropy}$, and let $k, h, m \in IOb_o$. If $\mathsf{speed}_m(k) = \mathsf{speed}_m(h)$ and $\mathsf{w}_{mk} \in {}_\kappa\mathsf{Iso}$, then $\mathsf{w}_{mh} \in {}_\kappa\mathsf{Iso}$.

Proof. By Lemma 6.1.8 (Same-Slope Rotation), there exists a spatial rotation R taking $\mathsf{wl}_m(k)$ to $\mathsf{wl}_m(h)$, and by Lemma 6.1.4 (Observer Rotation) there is some observer k^* satisfying $k \overset{R}{\rightsquigarrow}_m k^*$. Since $\mathsf{w}_{mk^*} = R \circ \mathsf{w}_{mk}$ and $R[\mathsf{wl}_m(k)] = \mathsf{wl}_m(h)$, it follows that $\mathsf{wl}_m(k^*) = R[\mathsf{wl}_m(k)] = \mathsf{wl}_m(h)$, so that k^* and h share the same worldline. By Lemma 6.3.7 (Colocate), w_{k^*h} is therefore trivial, and hence a κ-isometry. It now follows that $\mathsf{w}_{mh} = \mathsf{w}_{mk^*} \circ \mathsf{w}_{k^*h} = R \circ \mathsf{w}_{mk} \circ \mathsf{w}_{k^*h}$ is a composition of κ-isometries, so $\mathsf{w}_{mh} \in {}_\kappa\mathsf{Iso}$ as claimed. ☐

Lemma 6.6.2 (Rest)**.** Assume KIN. For all observers $k, m \in IOb$, we have

$$k \text{ is at rest according to } m \quad \text{iff} \quad \mathsf{w}_{mk} \in \mathsf{Triv}.$$

Proof. ($\Rightarrow$) Suppose first that k is at rest according to m, i.e. $\mathsf{w}_{mk}(\vec{\mathsf{o}})_s = \mathsf{w}_{mk}(\vec{\mathsf{t}})_s$. We will show that $\mathsf{w}_{mk} \in \mathsf{Triv}$.

Recall that $\mathsf{wl}_m(k)$ is a line (by AxLine) and notice that $\mathsf{w}_{mk}(\vec{\mathsf{o}}), \mathsf{w}_{mk}(\vec{\mathsf{t}}) \in \mathsf{w}_{mk}[\mathbf{t}] = \mathsf{wl}_m(k)$. Hence, $\mathsf{wl}_m(k)$ is parallel to $\mathbf{t}$ (because it is a line containing two distinct points, $\mathsf{w}_{mk}(\vec{\mathsf{o}})$ and $\mathsf{w}_{mk}(\vec{\mathsf{t}})$, whose spatial components are identical), and it passes through $\mathsf{w}_{mk}(\vec{\mathsf{o}})$.

Next, according to $\mathsf{AxRelocate}$ we can find an observer $m' \in IOb$ for which $\mathsf{w}_{mm'}$ is the translation taking $\vec{\mathsf{o}}$ to $\mathsf{w}_{mk}(\vec{\mathsf{o}})$. Because it is a translation, $\mathsf{w}_{mm'}$ necessarily takes $\mathbf{t}$ to a line parallel to $\mathbf{t}$; and because this line is $\mathsf{w}_{mm'}[\mathbf{t}] = \mathsf{wl}_m(m')$, we see that $\mathsf{wl}_m(m')$ is parallel to $\mathbf{t}$. Moreover, because $\mathbf{t}$ contains $\vec{\mathsf{o}}$, we know that $\mathsf{w}_{mk}(\vec{\mathsf{o}}) = \mathsf{w}_{mm'}(\vec{\mathsf{o}}) \in \mathsf{wl}_m(m')$, whence $\mathsf{wl}_m(m')$ is also a line parallel to $\mathbf{t}$ that passes through $\mathsf{w}_{mk}(\vec{\mathsf{o}})$.

Since $\mathsf{wl}_m(k)$ and $\mathsf{wl}_m(m')$ are parallel lines which share a common point, they must be the same (world)line, so $\mathsf{w}_{m'k} \in \mathsf{Triv}$ by Lemma 6.3.7 (Colocate). At the same time we know that $\mathsf{w}_{mm'} \in \mathsf{Triv}$, because it is a translation. It therefore follows by composition that $\mathsf{w}_{mk} = \mathsf{w}_{mm'} \circ \mathsf{w}_{m'k} \in \mathsf{Triv}$, as claimed.

($\Leftarrow$) To prove the converse, suppose that $\mathsf{w}_{mk} \in \mathsf{Triv}$. We need to show that k is at rest according to m, i.e. $\mathsf{w}_{mk}(\vec{\mathsf{t}})_s = \mathsf{w}_{mk}(\vec{\mathsf{o}})_s$. But this is obvious because every trivial transformation maps $\mathbf{t}$ to a line parallel to $\mathbf{t}$. $\square$

Remark 6.3. It follows easily from Lemma 6.6.2 (Rest) and the fact that Triv is a group under composition that "being at rest according to" is an equivalence relation on observers, and "moving according to" is a symmetric relation.

Lemma 6.6.3 (Observer Origin)**.** Assume $\mathsf{AxEField}$, AxWvt and $\mathsf{AxRelocate}$. If $\ell \in \mathsf{ObLines}(k)$ and $\vec{p} \in \ell$, then there exists some $h \in IOb$ for which $\mathsf{w}_{kh}(\vec{\mathsf{o}}) = \vec{p}$ and $\mathsf{wl}_k(h) = \ell$.

Proof. Choose $h' \in IOb$ such that $\mathsf{wl}_k(h') = \ell$. By $\vec{p} \in \mathsf{wl}_k(h')$, we have $\mathsf{w}_{h'k}(\vec{p}) \in \mathsf{w}_{h'k}[\mathsf{wl}_k(h')] = \mathsf{wl}_{h'}(h') = \mathbf{t}$. Let $h \in IOb$ be such that $\mathsf{w}_{h'h}$ is the translation by vector $\mathsf{w}_{h'k}(\vec{p})$. Such h exists by $\mathsf{AxRelocate}$. Translation $\mathsf{w}_{h'h}$ fixes $\mathbf{t}$ because $\mathsf{w}_{h'k}(\vec{p}) \in \mathbf{t}$. Then $\mathsf{wl}_k(h) = \mathsf{w}_{kh}[\mathbf{t}] = \mathsf{w}_{kh'}[\mathsf{w}_{h'h}[\mathbf{t}]] = \mathsf{w}_{kh'}[\mathbf{t}] = \mathsf{wl}_k(h') = \ell$ and $\mathsf{w}_{kh}(\vec{\mathsf{o}}) = \mathsf{w}_{kh'}(\mathsf{w}_{h'h}(\vec{\mathsf{o}})) = \mathsf{w}_{kh'}(\mathsf{w}_{h'k}(\vec{p})) = \vec{p}$ as claimed. $\square$

Lemma 6.6.4 (Median Observer)**.** Assume KIN, $\mathsf{AxIsotropy}$, and $\neg\exists\infty\mathsf{Speed}$. Then given any $k, m \in IOb_o$, there exists some $h \in IOb_o$ for which $\mathsf{speed}_h(k) = \mathsf{speed}_h(m)$.

Proof. If $\mathsf{speed}_k(m) = 0$, the result follows trivially by choosing $h = k$, so suppose $\mathsf{speed}_k(m) > 0$. By applying Theorem 6.5 (Fundamental Lemma) choosing $v = \mathsf{speed}_k(m)$, there exists $h \in IOb_o$ such that

$$\mathsf{speed}_k(h) = \mathsf{speed}_k(m) \tag{6.28}$$
$$\mathsf{speed}_m(k) = \mathsf{speed}_m(h). \tag{6.29}$$

Applying Theorem 6.4 (Same-Speed Lemma) to (6.29) tells us that

$$\mathsf{speed}_k(h) = \mathsf{speed}_h(k) \tag{6.30}$$
$$\mathsf{speed}_h(m) = \mathsf{speed}_k(m) \tag{6.31}$$

and so

$$\mathsf{speed}_h(k) \overset{(6.30)}{=\!=} \mathsf{speed}_k(h) \overset{(6.28)}{=\!=} \mathsf{speed}_k(m) \overset{(6.31)}{=\!=} \mathsf{speed}_h(m)$$

as claimed. $\square$

Lemma 6.6.5 (κ is unique)**.** Assume $\mathsf{AxEField}$ and let $m, k \in IOb$ be observers such that k is moving according to m and $\mathsf{w}_{mk} \in {}_\kappa\mathsf{Iso}$. Then κ is uniquely determined by:

$$\kappa = \frac{\left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_t - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_t\right|^2 - 1}{\left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_s - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_s\right|^2}. \tag{6.32}$$

Proof. Let $f : Q^4 \to Q^4$ be the linear part of w_{mk}, i.e. $f(\vec{p}\,) := \mathsf{w}_{mk}(\vec{p}\,) - \mathsf{w}_{mk}(\vec{\mathsf{o}}\,)$. Then f is a linear κ-isometry, so it preserves κ-length. Hence, $1 = \|\vec{\mathsf{t}}\|_\kappa = \|f(\vec{\mathsf{t}})\|_\kappa = f(\vec{\mathsf{t}})_t^2 - \kappa|f(\vec{\mathsf{t}})_s|^2 = |\mathsf{w}_{mk}(\vec{\mathsf{t}})_t - \mathsf{w}_{mk}(\vec{\mathsf{o}})_t|^2 - \kappa|\mathsf{w}_{mk}(\vec{\mathsf{t}})_s - \mathsf{w}_{mk}(\vec{\mathsf{o}})_s|^2$. We have that $\mathsf{w}_{mk}(\vec{\mathsf{t}})_s \neq \mathsf{w}_{mk}(\vec{\mathsf{o}})_s$ because k is moving according to m. Thus, (6.32) follows by reorganizing the equality above. □

6.6.3 Main proof

We now complete the proof of Theorem 6.6 (Main Lemma).

Proof of Theorem 6.6 (Main Lemma). There are two cases to consider: Case 1: $\neg\exists\infty\mathsf{Speed}$ holds. Case 2: $\exists\infty\mathsf{Speed}$ holds.

Proof of Case 1: Assume $\neg\exists\infty\mathsf{Speed}$.

Suppose $\hat{k}, \hat{m}$ are any observers in IOb_o. According to Lemma 6.6.4 (Median Observer), there is some $\hat{h} \in IOb_o$ such that $\mathsf{speed}_{\hat{h}}(\hat{k}) = \mathsf{speed}_{\hat{h}}(\hat{m})$. By Theorem 6.4 (Same-Speed Lemma), $\mathsf{w}_{\hat{m}\hat{k}}$ is a κ-isometry for some $\kappa \in Q$. This shows that every worldview transformation between two observers in IOb_o is a κ-isometry for some κ, and by Lemma 6.6.5 (κ is unique), this κ is unique if the two observers are moving relative to each other (however, even this unique κ may vary with the choice of the two observers.)

Suppose that $k \in IOb_o$. We will show that κ can be found such that (6.27) holds.

Let us note first that if any observer $m \in IOb_o$ is at rest relative to k, then Lemma 6.6.2 (Rest) tells us that w_{mk} is trivial, thus it is a κ-isometry for every $\kappa \in Q$ by Lemma 6.3.2 ($\mathsf{Triv} = \bigcap_\kappa \mathsf{Iso}$). So we only need to consider observers which are moving relative to k.

Suppose, therefore, that $m_1, m_2 \in IOb_o$ are two observers, and that at least one is moving according to k. Without loss of generality we can assume that $0 < \mathsf{speed}_k(m_1)$ and $\mathsf{speed}_k(m_2) \leq \mathsf{speed}_k(m_1)$. We have already seen that

$$\mathsf{w}_{m_1k} \in {}_{\tilde{\kappa}}\mathsf{Iso} \tag{6.33}$$

for some unique $\tilde{\kappa}$. We will show that $\mathsf{w}_{m_2k} \in {}_{\tilde{\kappa}}\mathsf{Iso}$ as well. We have already seen that this is the case if $\mathsf{speed}_k(m_2) = 0$, so we can assume that $0 < \mathsf{speed}_k(m_2)$.

By Theorem 6.5 (Fundamental Lemma), choosing $v = \mathsf{speed}_k(m_2)$ and $m = m_1$, there exists $h \in IOb_o$ such that

$$\begin{aligned} \mathsf{speed}_k(h) &= \mathsf{speed}_k(m_2) && (6.34)\\ \mathsf{speed}_{m_1}(k) &= \mathsf{speed}_{m_1}(h) && (6.35) \end{aligned}$$

It follows from Lemma 6.6.1 (Same Speed Easy) with (6.33) and (6.35) that

$$\mathsf{w}_{m_1h} \in {}_{\tilde{\kappa}}\mathsf{Iso} \tag{6.36}$$

and hence (by (6.33)) that

$$\mathsf{w}_{kh} = \mathsf{w}_{km_1} \circ \mathsf{w}_{m_1h} = \mathsf{w}_{m_1k}^{-1} \circ \mathsf{w}_{m_1h} \in {}_{\tilde{\kappa}}\mathsf{Iso}. \tag{6.37}$$

Applying Lemma 6.6.1 (Same Speed Easy) with (6.34) and (6.37) now tells us that $\mathsf{w}_{km_2} \in {}_{\tilde{\kappa}}\mathsf{Iso}$. But then

$$\mathsf{w}_{m_2k} \in {}_{\tilde{\kappa}}\mathsf{Iso}$$

as claimed.

Finally, let $m \in IOb_o$ be arbitrary. As we have shown, no matter whether m is at rest or in motion relative to k, there is some κ_m such that w_{m_1k} and w_{mk} are both in ${}_{\kappa_m}\mathsf{Iso}$. But because m_1 is moving relative to k this κ_m is unique for m_1, so we must have $\kappa_m = \tilde{\kappa}$. Thus, taking $\kappa := \tilde{\kappa}$ ensures that (6.27) holds as claimed.

Proof of Case 2: Assume $\exists\infty\mathsf{Speed}$.

By Lemma 6.2.7 (Infinite Speeds ⇒ Lines are Observer Lines), every observer considers every line to be the worldline of an observer, so in particular any 'horizontal' line through $\vec{\mathsf{o}}$ is an observer line. By Lemma 6.6.3 (Observer Origin), therefore, there exists $h \in IOb_o$ satisfying $\mathsf{speed}_o(h) = \infty$.

Recall that $\mathbf{S}$ is the spatial hyperplane $\{(0, x, y, z) : x, y, z \in Q\}$; let us consider $\mathsf{w}_{oh}[\mathbf{S}]$. By Theorem 6.2 (Line-to-Line Lemma), this is a 3-dimensional subspace of Q^4 which contains $\mathsf{w}_{oh}(\vec{\mathsf{o}}) = \vec{\mathsf{o}}$ (because $h \in IOb_o$). It follows that the subspace formed by the intersection of $\mathbf{S}$ with $\mathsf{w}_{oh}[\mathbf{S}]$ must be at least 1-dimensional and so there is some line ℓ such that $\vec{\mathsf{o}} \in \ell \subseteq \mathsf{w}_{oh}[\mathbf{S}] \cap \mathbf{S}$. See Figure 20.

Because every observer considers every line to be an observer line, o considers ℓ to be an observer line, so there exists some k such that $\ell = \mathsf{wl}_o(k)$. By Lemma 6.6.3 (Observer Origin), we can choose this k to be in IOb_o. Since $\ell \subseteq \mathbf{S}$, we have $\mathsf{speed}_o(k) = \infty$. It follows that $\mathsf{speed}_o(k) = \mathsf{speed}_o(h)$ (both are infinite), whence Theorem 6.4 (Same-Speed Lemma) tells us that $\mathsf{w}_{hk} \in {}_{\kappa}\mathsf{Iso}$ for some κ. Let us fix such a κ. We will prove that (6.27) holds for this κ.

To do this, we first switch from o's worldview to h's. By construction, we know that $\mathsf{wl}_o(k) = \ell \subseteq \mathsf{w}_{oh}[\mathbf{S}]$, so by applying w_{ho}, we have

$$\mathsf{wl}_h(k) \subseteq \mathbf{S}, \tag{6.38}$$

and hence $\mathsf{speed}_h(k) = \infty$.

Now let m be any observer $m \in IOb_o$.

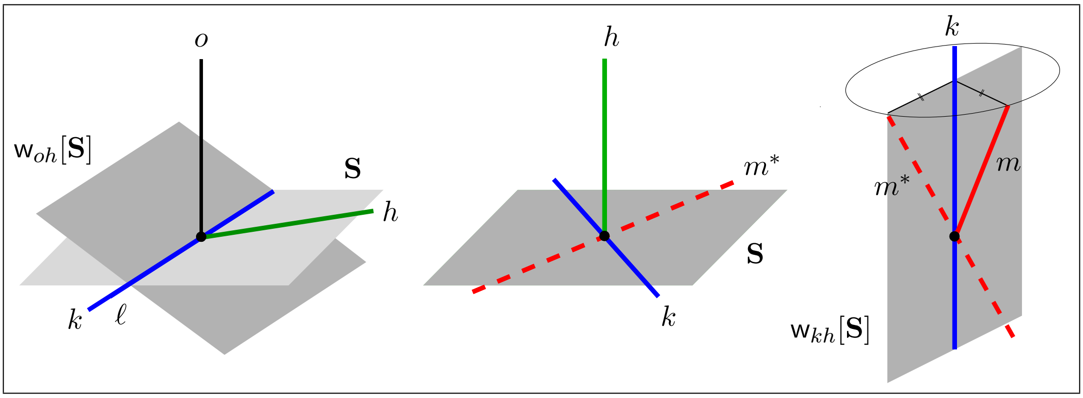


Figure 20: Illustration for the proof of Theorem 6.6 (Main Lemma) if $\exists\infty\mathsf{Speed}$ is assumed.

In the particular case when $\mathsf{wl}_h(m) \subseteq \mathbf{S}$, we must have $\mathsf{speed}_h(m) = \infty$ because all points in $\mathbf{S}$ have the same time coordinate. In this case, we have $\mathsf{speed}_h(k) = \mathsf{speed}_h(m)$, and since we know that $\mathsf{w}_{hk} \in {}_\kappa\mathsf{Iso}$, Lemma 6.6.1 (Same Speed Easy) tells us that w_{hm} (hence also w_{mh}) is a κ-isometry as well. It now follows by composition, in this special case, that $\mathsf{w}_{mk} = \mathsf{w}_{mh} \circ \mathsf{w}_{hk}$ is a κ-isometry, as required.

Now consider things more generally from k's point of view. As before, $\mathsf{w}_{kh}[\mathbf{S}]$ is a hyperplane, and we know from (6.38) that $\mathsf{wl}_h(k) \subseteq \mathbf{S}$. It follows that

$$\mathbf{t} = \mathsf{wl}_k(k) = \mathsf{w}_{kh}[\mathsf{wl}_k(h)] \subseteq \mathsf{w}_{kh}[\mathbf{S}]$$

so $\mathsf{w}_{kh}[\mathbf{S}]$ contains the time-axis $\mathbf{t}$.

We can therefore find a line ℓ such that $\vec{\mathsf{o}} \in \ell \subseteq \mathsf{w}_{kh}[\mathbf{S}]$ and $\mathsf{slope}(\ell) = \mathsf{speed}_k(m)$. For if $\mathsf{speed}_k(m) = \infty$ we can choose the line through $\vec{\mathsf{o}}$ in $\mathsf{w}_{kh}[\mathbf{S}]$ that is perpendicular to $\mathbf{t}$, and if $\mathsf{speed}_k(m) = 0$ we can take $\ell = \mathbf{t}$. For the remaining case, where $0 < \mathsf{speed}_k(m) < \infty$, choose any point $\vec{p} \in \mathsf{w}_{kh}[\mathbf{S}] \setminus \mathbf{t}$. By Lemma 6.1.10 (Triangulation), we can find a line of slope $\mathsf{speed}_k(m)$ in $\mathsf{w}_{kh}[\mathbf{S}]$ which meets $\mathbf{t}$, and a translation along $\mathbf{t}$ can then be applied to find a parallel line (also in $\mathsf{w}_{kh}[\mathbf{S}]$) that passes through $\vec{\mathsf{o}}$.

Because all lines are observer lines, ℓ is an observer line; and by Lemma 6.6.3 (Observer Origin) there is some $m^* \in IOb_o$ for which $\mathsf{wl}_k(m^*) = \ell \subseteq \mathsf{w}_{kh}[\mathbf{S}]$. But this means that $\mathsf{wl}_h(m^*) = \mathsf{w}_{hk}[\mathsf{wl}_k(m^*)] \subseteq \mathsf{w}_{hk}[\mathsf{w}_{kh}[\mathbf{S}]] = \mathbf{S}$ and hence, as we saw in the special case above, $\mathsf{w}_{m^*k} \in {}_\kappa\mathsf{Iso}$. But now Lemma 6.6.1 (Same Speed Easy) tells us that from $\mathsf{speed}_k(m) = \mathsf{slope}(\ell) = \mathsf{speed}_k(m^*)$ and $\mathsf{w}_{km^*} \in {}_\kappa\mathsf{Iso}$ we can deduce $\mathsf{w}_{km} \in {}_\kappa\mathsf{Iso}$. Therefore, for arbitrary $m \in IOb_o$, $\mathsf{w}_{mk} \in {}_\kappa\mathsf{Iso}$, i.e. (6.27) holds. $\square$

7 Proofs of the main theorems

Proof of Theorem 5.1 (Characterisation). If $\neg\exists\mathsf{MovingIOb}$ is assumed, then $\mathbb{W} \subseteq \mathsf{Triv}$ by Lemma 6.6.2 (Rest), hence $\mathbb{W} \subseteq {}_\kappa\mathsf{Iso}$ for every κ by Lemma 6.3.2 ($\mathsf{Triv} = \bigcap_\kappa \mathsf{Iso}$).

Assume $\exists\mathsf{MovingIOb}$. Let $k \in IOb_o$ and κ be such that (6.27) in Theorem 6.6 (Main Lemma) holds, i.e. $\{\mathsf{w}_{mk} : m \in IOb_o\} \subseteq {}_\kappa\mathsf{Iso}$. Then by Lemma 6.6.5 ($\kappa$ is unique) it is enough to prove that the worldview transformations are κ-isometries.

To prove that worldview transformations are κ-isometries, choose any observers $m_1, m_2 \in IOb$. By Lemma 6.4.1 (Translation to IOb$_o$), we can find $m_1^o, m_2^o \in IOb_o$ for which $\mathsf{w}_{m_1 m_1^o}$ and $\mathsf{w}_{m_2^o m_2}$ are translations and hence κ-isometries. As $\mathsf{w}_{m_1^o k}$ and $\mathsf{w}_{m_2^o k}$ are also κ-isometries, so it follows that

$$\mathsf{w}_{m_1 m_2} = \mathsf{w}_{m_1 m_1^o} \circ \mathsf{w}_{m_1^o k} \circ \mathsf{w}_{k m_2^o} \circ \mathsf{w}_{m_2^o m_2} = \mathsf{w}_{m_1 m_1^o} \circ \mathsf{w}_{m_1^o k} \circ \mathsf{w}_{m_2^o k}^{-1} \circ \mathsf{w}_{m_2^o m_2} \tag{7.1}$$

is a κ-isometry. □

Proof of Theorem 5.2 (Satisfaction). Let us first prove that

$$\mathbb{W}_k = \mathsf{G}, \text{ for every } k \in IOb. \tag{7.2}$$

To do so, let $k \in IOb$. Then, by the definition of $\mathbb{W}_k$ and the construction of $\mathcal{M}_G$,

$$\mathbb{W}_k = \{\mathsf{w}_{kh} : h \in IOb\} = \{k \circ h^{-1} : h \in \mathsf{G}\} = k \circ \mathsf{G}^{-1} = \mathsf{G}$$

because G is a group. Thus, (7.2) holds.

(a) By construction of $\mathcal{M}_G$, we have $\mathsf{w}_{kk} = k \circ k^{-1} = Id$ and $\mathsf{w}_{mh} \circ \mathsf{w}_{hk} = m \circ h^{-1} \circ h \circ k^{-1} = m \circ k^{-1} = \mathsf{w}_{mk}$ for every $m, k, h \in IOb = \mathsf{G}$. Thus, AxWvt holds. By (7.2), we have that $\mathbb{W}_k = \mathbb{W}_h$ for every $k, h \in IOb$, which is a trivial reformulation of AxSPR. Finally, also by (7.2), we have $\mathbb{W} = \bigcup_{k \in IOb} \mathbb{W}_k = \mathsf{G}$.

(b) A trivial reformulation of $\mathsf{AxRelocate}$ is that $\mathsf{SRot} \cup \mathsf{Trans} \subseteq \mathbb{W}_k$ for all $k \in IOb$, which, by (7.2), is equivalent to $\mathsf{SRot} \cup \mathsf{Trans} \subseteq \mathsf{G}$ in $\mathcal{M}_\mathsf{G}$.

(c) By definition of worldline, a trivial reformulation of AxLine is that $g[\mathbf{t}]$ is a line for every $g \in \mathbb{W}$. We know from (a) that $\mathbb{W} = \mathsf{G}$, hence the statement holds.

(d) We know from (a) that AxWvt holds, hence by Lemma 6.1.2 (WVT), $\mathsf{wl}_k(k) = \mathbf{t}$ for every $k \in IOb$. Recall that by definition of worldline $\mathsf{wl}_k(k') := \mathsf{w}_{kk'}[\mathbf{t}]$ for every $k, k' \in IOb$. Hence, for every $k, k' \in IOb$, $\mathsf{wl}_k(k) = \mathsf{wl}_k(k')$ is equivalent to $\mathsf{w}_{kk'}[\mathbf{t}] = \mathbf{t}$ in $\mathcal{M}_\mathsf{G}$. Therefore, $\mathsf{AxColocate}$ holds in $\mathcal{M}_\mathsf{G}$ iff $g \in \mathsf{Triv}$ whenever $g \in \mathbb{W}$ and $g[\mathbf{t}] = \mathbf{t}$. We know from (a) that $\mathbb{W} = \mathsf{G}$, hence the statement holds. □

Proof of Theorem 5.3 (Model Construction). From Lemma 5.2 (Satisfaction)(a-c), it is clear that AxWvt, AxSPR, AxLine and $\mathsf{AxRelocate}$ all hold, and that $\mathbb{W} = \mathsf{G}$. To see that $\mathsf{AxColocate}$ also holds, suppose $g \in c\mathsf{Poi} \cup c\mathsf{Eucl} \cup \mathsf{Gal}$ satisfies $g[\mathbf{t}] = \mathbf{t}$. We will show that $g \in \mathsf{Triv}$, whence the result follows by Lemma 5.2 (Satisfaction)(d).

To this end, write $g = T \circ L$ as a composition of a translation T and linear κ-isometry L, and recall that a linear map is trivial if and only if it fixes (setwise) both the time-axis and the present simultaneity, and preserves squared lengths in both. We will show that L has these properties.

To see that $L[\mathbf{t}] = \mathbf{t}$, note that $T(\vec{o}) = T(L(\vec{o})) = g(\vec{o}) \in \mathbf{t}$, whence T must be a translation along the t-axis. Thus, g and T both fix $\mathbf{t}$ setwise, whence so does $L = T^{-1} \circ g$.

To see that L preserves squared length in $\mathbf{t}$, choose arbitrary $t \in Q$. Since $L[\mathbf{t}] = \mathbf{t}$ there is some $t' \in Q$ such that $L(t, \vec{0}) = (t', \vec{0})$, and now $\left\|L(t, \vec{0})\right\|_\kappa^2 = \left\|(t, \vec{0})\right\|_\kappa^2$ forces $t' = \pm t$. Thus, L preserves squared lengths in $\mathbf{t}$.

If $\kappa = 0$, then L fixes the present simultaneity $\mathbf{S}$ and preserves the square lengths in it by definition. To see that the same statement holds if $\kappa \neq 0$, choose arbitrary $\vec{s} \in Q^3$ and define $t^* \in Q$ and $\vec{s}^* \in Q^3$ by $(t^*, \vec{s}^*) := L(0, \vec{s})$. Then by linearity

$$L(1, \vec{s}) = (\pm 1 + t^*, \vec{s}^*) \qquad \text{and} \qquad L(1, -\vec{s}) = (\pm 1 - t^*, -\vec{s}^*).$$

Since $\|(1, \vec{s})\|_\kappa^2 = \|(1, -\vec{s})\|_\kappa^2$ and L is a linear κ-isometry, we have that $\|L(1, \vec{s})\|_\kappa^2 = \|L(1, -\vec{s})\|_\kappa^2$, which implies that $(1 + t^*)^2 = (1 - t^*)^2$ and hence $t^* = 0$. Thus, $L(0, \vec{s}) = (0, \vec{s}^*)$, i.e. L maps $\mathbf{S}$ to itself. If $\kappa \neq 0$, $\|(0, \vec{s})\|_\kappa^2 = \|L(0, \vec{s})\|_\kappa^2 = \|(0, \vec{s}^*)\|_\kappa$ implies that $|\vec{s}|^2 = |\vec{s}^*|^2$. Hence, L preserves the square lengths in $\mathbf{S}$.

As claimed, therefore, L is a linear map which fixes both the time-axis and the present simultaneity, and preserves squared lengths in both, whence it is linear trivial and $g = T \circ L$ is trivial. As outlined above, it now follows that $\mathsf{AxColocate}$ also holds, and that hence $\mathcal{M}_\mathsf{G}$ is a model in which $\mathsf{KIN} + \mathsf{AxSPR}$ holds and $\mathbb{W} = \mathsf{G}$. □

Proof of Theorem 5.4 (Determination). Assume that G is a group satisfying the conditions. We will prove that statements (i) and (ii) are equivalent.

Assume that (i) holds. By Theorem 5.2 (Satisfaction), $\mathcal{M}_\mathsf{G}$ is a model of $\mathsf{KIN} + \mathsf{AxSPR}$ (and hence also $\mathsf{KIN} + \mathsf{AxIsotropy}$) for which $\mathbb{W} = \mathsf{G}$. Then (ii) follows by Theorem 5.1 (Characterisation).

Assume that (ii) holds. Then by Theorem 5.3 (Model Construction) $\mathcal{M}_\mathsf{G}$ is a model of $\mathsf{KIN} + \mathsf{AxSPR}$ for which $\mathbb{W} = \mathsf{G}$. Then (i) follows by Theorem 5.2 (Satisfaction). □

Proof of Theorem 5.5 (Classification). Assume $\mathsf{KIN} + \mathsf{AxIsotropy}$. It is clear that at least one of cases (1)-(4) holds. First we show the consequences of the cases and then from those we show that they are mutually exclusive.

(Cases 1-3) If $k, m \in IOb$ are at rest relative to each other, then because w_{mk} is trivial by Lemma 6.6.2 (Rest), it is also a Euclidean isometry by Lemma 6.3.2 ($\mathsf{Triv} = \bigcap_\kappa \mathsf{Iso}$). Thus, for all observers k and m we have

$$\text{if } \mathsf{w}_{mk}(\vec{\mathsf{t}})_s = \mathsf{w}_{mk}(\vec{\mathsf{o}})_s, \text{ then } \left|\mathsf{w}_{mk}(\vec{\mathsf{t}})_t - \mathsf{w}_{mk}(\vec{\mathsf{o}})_t\right| = 1. \tag{7.3}$$

We claim we can choose k^* and m^* such that $\mathsf{w}_{m^*k^*}(\vec{\mathsf{t}})_s \neq \mathsf{w}_{m^*k^*}(\vec{\mathsf{o}})_s$. This is true by definition if $\exists\mathsf{MovingAccurateClock}$ holds, and follows from (7.3) if either $\exists\mathsf{SlowClock}$ or $\exists\mathsf{FastClock}$ holds because in each of these cases we can choose m^*, k^* such that $\left|\mathsf{w}_{m^*k^*}(\vec{\mathsf{t}})_t - \mathsf{w}_{m^*k^*}(\vec{\mathsf{o}})_t\right| \neq 1$.

It follows that $\exists\mathsf{MovingIOb}$ holds in all three cases, and so by Theorem 5.1 (Characterisation), there is a unique κ such that $\mathbb{W} \subseteq {}_\kappa\mathsf{Iso}$. Recall from (6.32) that κ can be determined from the motion of any two observers moving relative to one another by

$$\kappa = \frac{\left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_t - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_t\right|^2 - 1}{\left|\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_s - \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_s\right|^2}.$$

So, given our choice of m^*, k^* (and the definitions of $\exists\mathsf{FastClock}$, $\exists\mathsf{SlowClock}$ and $\exists\mathsf{MovingAccurateClock}$) we have

$$\begin{aligned}
\exists\mathsf{SlowClock} &\Rightarrow \left|\mathsf{w}_{m^*k^*}(\vec{\mathsf{t}})_t - \mathsf{w}_{m^*k^*}(\vec{\mathsf{o}})_t\right|^2 > 1 &&\Rightarrow \kappa > 0\\
\exists\mathsf{FastClock} &\Rightarrow \left|\mathsf{w}_{m^*k^*}(\vec{\mathsf{t}})_t - \mathsf{w}_{m^*k^*}(\vec{\mathsf{o}})_t\right|^2 < 1 &&\Rightarrow \kappa < 0\\
\exists\mathsf{MovingAccurateClock} &\Rightarrow \left|\mathsf{w}_{m^*k^*}(\vec{\mathsf{t}})_t - \mathsf{w}_{m^*k^*}(\vec{\mathsf{o}})_t\right|^2 = 1 &&\Rightarrow \kappa = 0.
\end{aligned}$$

Because (6.32) holds for any two relatively moving observers it now follows from the uniqueness of κ that $\exists\mathsf{SlowClock} \Rightarrow \forall\mathsf{MovingClockSlow}$, $\exists\mathsf{FastClock} \Rightarrow \forall\mathsf{MovingClockFast}$ and $\exists\mathsf{MovingAccurateClock} \Rightarrow \forall\mathsf{ClockAccurate}$.

Finally, to complete the proof of cases (1-3) it is enough to note that

$$\begin{aligned}
\kappa > 0 &\quad\Rightarrow\quad {}_\kappa\mathsf{Iso} = c\mathsf{Poi} \text{ where } c = \sqrt{1/\kappa};\\
\kappa < 0 &\quad\Rightarrow\quad {}_\kappa\mathsf{Iso} = c\mathsf{Eucl} \text{ where } c = \sqrt{-1/\kappa};\\
\kappa = 0 &\quad\Rightarrow\quad {}_\kappa\mathsf{Iso} = \mathsf{Gal}.
\end{aligned}$$

(Case 4). If $\neg\exists\mathsf{MovingIOb}$ holds, then all worldview transformations are trivial by Lemma 6.6.2 (Rest), so $\mathbb{W} \subseteq \mathsf{Triv}$ as claimed.

The four cases are clearly mutually exclusive, because the situations

$$(\forall\mathsf{MovingClockSlow} + \exists\mathsf{MovingIOb}), \quad (\forall\mathsf{MovingClockFast} + \exists\mathsf{MovingIOb}),$$

$$\exists\mathsf{MovingAccurateClock} \quad \text{and} \quad \neg\exists\mathsf{MovingIOb}$$

are mutually exclusive. $\square$

Proof of Theorem 5.6 (Consistency). (Cases 1-3) By Theorem 5.3 (Model Construction) and (5.2), there are models $\mathcal{M}_P$, $\mathcal{M}_E$ and $\mathcal{M}_G$ of $\mathsf{KIN} + \mathsf{AxSPR}$ such that the set of worldview transformations are respectively Poi, Eucl and Gal. In all three models, there are $m, k \in IOb$ such that $\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_s \neq \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_s$ because if $\mathbb{W} = \mathsf{Poi}$ or $\mathbb{W} = \mathsf{Eucl}$ or $\mathbb{W} = \mathsf{Gal}$, then it can be easily seen that there is $f \in \mathbb{W}$ such that $f\left(\vec{\mathsf{t}}\right)_s \neq f\left(\vec{\mathsf{o}}\right)_s$. Let such m and k be fixed. Then $\exists\mathsf{MovingIOb}$ holds. Thus, by Theorem 5.1 (Characterisation), there is a unique κ such that the set of worldview transformations is a subset of ${}_{\kappa}\mathsf{Iso}$. This κ is positive ($\kappa = 1$) in $\mathcal{M}_P$, negative ($\kappa = -1$) in $\mathcal{M}_E$ and 0 in $\mathcal{M}_G$. Then by equation (6.32) in Lemma 6.6.5 (κ is unique) it can be seen that $\exists\mathsf{SlowClock}$ holds in $\mathcal{M}_P$, $\exists\mathsf{FastClock}$ holds in $\mathcal{M}_E$ and $\exists\mathsf{MovingAccurateClock}$ holds in $\mathcal{M}_G$.

(Case 4) It remains to prove that $\mathsf{KIN} + \mathsf{AxSPR} + \neg\exists\mathsf{MovingIOb}$ has a model. Let $\mathcal{M}_T$ be a model of $\mathsf{KIN} + \mathsf{AxSPR}$ such that $\mathbb{W} = \mathsf{Triv}$. Such $\mathcal{M}_T$ exists by Theorem 5.3 (Model Construction) and (5.2). Let us notice that for any $f \in \mathsf{Triv}$, $f\left(\vec{\mathsf{t}}\right)_s = f\left(\vec{\mathsf{o}}\right)_s$. Therefore, for every $m, k \in IOb$, $\mathsf{w}_{mk}\left(\vec{\mathsf{t}}\right)_s = \mathsf{w}_{mk}\left(\vec{\mathsf{o}}\right)_s$, and this means that $\neg\exists\mathsf{MovingIOb}$ holds in $\mathcal{M}_T$. $\square$

8 Discussion

In this paper, we have presented an essentially elementary description of what can be deduced about the geometry of (1 + 3)-dimensional spacetime from isotropy if we restrict ourselves to first-order logic and make as few background assumptions as reasonably possible. Nonetheless, there is potential to go further, as even our own very simple assumptions can potentially be weakened while still providing a physically relevant description. The history of the field has shown repeatedly that authors have inadvertently made unconscious, and sometimes unnecessary, assumptions, and it would be foolish to assume that we are necessarily immune to this problem. We have accordingly started a programme of painstakingly machine-verifying our results using interactive theorem provers [31], but this programme remains very much in its infancy. In the meantime, therefore, we have been as explicit as possible at all stages of our proofs.

We began by noting that, in the elementary framework advocated in this paper there are reasons why it is no longer appropriate to assume that the ordered field Q of numbers used when recording physical measurements is the field $\mathbb{R}$ of real numbers. Partly this is because practical measurements can never achieve more than a few decimal points of accuracy, and partly because the field $\mathbb{R}$ cannot be uniquely characterised in terms of the first-order sentences it satisfies. But as we have also shown, it is simply not necessary to make the assumption. As long as Q allows the taking of square roots of non-negative values, all of our results hold.

Our results tell us, subject to a small number of very basic axioms, that the worldview transformations that characterise kinematics in isotropic spacetime form a group $\mathbb{W}$ of κ-isometries for some κ. In contrast to earlier studies, we have not needed to assume the full special principle of relativity, but have shown instead that the strictly weaker assumption that space is isotropic is already enough to entail these results. We accordingly obtain four basic possibilities: the universe is not static (there are moving observers) and $\mathbb{W}$ is a subgroup of either Poi, Eucl or Gal, or the universe is static (all observers are at rest with respect to one another) and $\mathbb{W} \subseteq \mathsf{Triv}$.

As usual (if moving observers exist) we can identify which kind of spacetime we are in by considering whether moving clocks run slow or fast or remain accurate. But because we have not restricted ourselves to $Q = \mathbb{R}$, we have allowed for the possibility that the structure of Q may be somewhat more complicated than usually assumed (for example, there is no reason why Q should not contain infinite or infinitesimal values). This in turn means that the topological structure of Q^4 need not satisfy the usual theorems of $\mathbb{R}^4$, nor the symmetry group $\mathsf{Sym}(Q^4)$ has to satisfy those of $\mathsf{Sym}(\mathbb{R}^4)$. Even so, we have shown that all 'reasonable' subgroups G of $\mathsf{Sym}(Q^4)$ can occur as the transformation group $\mathbb{W}$ in some associated model $\mathcal{M}_\mathsf{G}$. In other words, assuming that $Q = \mathbb{R}$ has inadvertently imposed severe and unnecessary limitations on the set of models investigated in earlier papers.

Nonetheless, many questions remain to be answered. Which of our results still hold, for example, if we remove the requirement for Q to be Euclidean? Are square roots essential, and if not, how can this be interpreted physically? For example, when $\kappa > 0$ the value κ corresponds to a model in which the speed of light is given by $c = \sqrt{1/\kappa}$, but what happens if κ has no square root? Presumably this would be a model in which light signals cannot exist, since they would need to travel with non-existent speed. Some familiar expressions might still be meaningful, for example $\sqrt{1 - v^2/c^2}$ can be rewritten as $\sqrt{1 - \kappa v^2}$, but even so, how does time dilation 'work' if v is a value for which $\sqrt{1 - \kappa v^2}$ is undefined?

There is also the issue of dimensionality. Our initial investigations suggest that all of the proofs presented here go through for dimensions $d \geq (1 + 3)$, but can fail for $d = (1+1)$. But do they hold for $d = (1+2)$? The answer appears to be yes if we allow trivial transformations to reverse the direction of time — but is this inclusion of reflections essential? We simply do not know.

References

[1] H. Andréka, J. X. Madarász, and I. Németi, *Logical axiomatizations of space-time. Samples from the literature*, Non-Euclidean Geometries: János Bolyai Memorial Volume (A. Prékopa and E. Molnár, eds.), Springer Verlag, 2006, pp. 155–185.

[2] ———, *Logic of space-time and relativity theory*, Handbook of Spatial Logics (M. Aiello, I. Pratt-Hartmann, and J. van Benthem, eds.), Springer Verlag, 2007, pp. 607–711.

[3] H. Andréka, J. X. Madarász, I. Németi, and G. Székely, *A logic road from special relativity to general relativity*, Synthese **186,3** (2012), 633–649.

[4] H. Andréka and I. Németi, *Comparing theories: the dynamics of changing vocabulary*, Johan van Benthem on Logic and Information Dynamics (A. Baltag and S. Smets, eds.), Springer Verlag, 2014, pp. 143–172.

[5] J. Ax, *The elementary foundations of spacetime*, Found. Phys. **8** (1978), no. 7–8, 507–546.

[6] M. Berger, *Geometry I*, Universitext, Springer Berlin Heidelberg, 1987.

[7] Yu. F. Borisov, *Axiomatic definition of the Galilean and Lorentz groups*, Siberian Mathematical Journal **19** (1978), no. 6, 870–882.

[8] L. Cocco and J. Babic, *A system of axioms for Minkowski spacetime*, Journal of Philosophical Logic (2020).

[9] Gy. Dávid, *Special relativity based on group theory*, 1990, Talk at Summer School on Special Relativity, Galyatető, Hungary, 1990.

[10] P. Frank and H. Rothe, *Über die Transformation der Raumzeitkoordinaten von ruhenden auf bewegte Systeme*, Annalen der Physik **339** (1911), no. 5, 825–855, English translation (transl. Morris D. Friedman Inc.): `https://archive.org/details/nasa_techdoc_19880069066/page/n18`.

[11] M. Friend, *On the epistemological significance of the Hungarian project*, Synthese **192,7** (2015), 2035–2051.

[12] M. Friend and D. Molinini, *Using mathematics to explain a scientific theory*, Philosophia Mathematica **24** (2015), no. 2, 185–213.

[13] R. Goldblatt, *Orthogonality and spacetime geometry*, Springer–Verlag, Berlin, 1987.

[14] M. Gömöri, *The Principle of Relativity—An Empiricist Analysis*, Ph.D. thesis, Eötvös University, Budapest, 2015.

[15] M. Gömöri and L. E. Szabó, *Formal statement of the special principle of relativity*, Synthese **192** (2015), no. 7, 2053–2076.

[16] N. S. Govindarajalulu, S. Bringsjord, and J. Taylor, *Proof verification and proof discovery for relativity*, Synthese **192** (2015), no. 7, 2077–2094.

[17] A. K. Guts, *The axiomatic theory of relativity*, Russ. Math. Surv. **37** (1982), no. 2, 41–89.

[18] W. v. Ignatowsky, *Das Relativitätsprinzip*, Archiv der Mathematik und Physik **17** (1910), 1–24.

[19] ———, *Einige allgemeine Bemerkungen über das Relativitätsprinzip*, Physikalische Zeitschrift **11** (1910), 972–976, English translation: https://en.wikisource.org/wiki/Translation:Some_General_Remarks_on_the_Relativity_Principle.

[20] ———, *Das Relativitätsprinzip*, Archiv der Mathematik und Physik **18** (1911), 17–40.

[21] J. X. Madarász and G. Székely, *Special relativity over the field of rational numbers*, International Journal of Theoretical Physics **52,5** (2013), 1706–1718.

[22] J. X. Madarász, G. Székely, and M. Stannett, *Three different formalisations of Einstein's relativity principle*, The Review of Symbolic Logic **10** (2017), no. 3, 530–548.

[23] ———, *Groups of worldview transformations implied by Einstein's special principle of relativity over arbitrary ordered fields*, 2019, submitted.

[24] M. Mamone-Capria, *On the fundamental theorem of the theory of relativity*, Foundations of Physics **46** (2016), no. 12, 1680–1712.

[25] Encyclopaedia of Mathematics, *Euclidean field*, Accessed 11 Feb 2020, http://www.encyclopediaofmath.org/index.php?title=Euclidean_field&oldid=39810.

[26] B. Odom, D. Hanneke, B. D'Urso, and G. Gabrielse, *New measurement of the electron magnetic moment using a one-electron quantum cyclotron*, Phys. Rev. Lett. **97** (2006), 030801.

[27] V. Pambuccian, *Alexandrov–Zeeman type theorems expressed in terms of definability*, Aequationes mathematicae **74** (2007), no. 3, 249–261.

[28] V. Pambuccian, H. Struve, and R. Struve, *Metric geometries in an axiomatic perspective*, From Riemann to Differential Geometry and Relativity (L. Ji, A. Papadopoulos, and S. Yamada, eds.), Springer International Publishing, Cham, 2017, pp. 413–455.

[29] J. W. Schutz, *Independent axioms for Minkowski space-time*, Longoman, London, 1997.

[30] E. Snapper and R. J. Troyer, *Metric affine geometry*, Academic Press, 1971.

[31] M. Stannett and I. Németi, *Using Isabelle/HOL to verify first-order relativity theory*, Journal of Automated Reasoning **52** (2014), no. 4, 361–378.

[32] R. Struve, *An axiomatic foundation of Cayley-Klein geometries*, Journal of Geometry **107** (2016), no. 2, 225–248.

[33] P. Suppes, *Axioms for relativistic kinematics with or without parity*, Symposium on the axiomatic method with special reference to physics, 1959, pp. 291–307.

[34] L. E. Szabó, *Empirical foundation of space and time*, EPSA07: Launch of the European Philosophy of Science Association (M. Suárez, M. Dorato, and M. Rédei, eds.), Springer Verlag, 2009.

 Received 12 November 2020

Events as Located Situations: Ontological Commitments in the Problem of Individuating Events

B.O. Akinkunmi
Dept of Computer Science, University of Ibadan, Ibadan, Nigeria.
ope34648@yahoo.com

Abstract

One of the sub-problems of the event individuation problem is: if A does x by doing y, is A doing x the same event as A doing y? Although it has been argued in the literature that doing x is a trying and as such, a mental event, the basis for defining the physicality of such events when they succeed, needs to be defined.

This paper argues that actions that must be done by doing something else are Vendlerian accomplishments and that treating events as located situations enables one to commit to the option in which doing y is a sub-situation of doing x. This is done by defining a sufficient condition for inferring that x has been accomplished from the fact that y has been done, which, is when a situation in which y has been done causes another situation in which the culmination of the accomplishment x holds.

By extending the situation theory with intentional predicates, it becomes possible to model x as an intentional or strategic accomplishment and consequently define the sufficient condition for inferring that a situation in which y is done is a plan that resulted in a situation for accomplishing x. That condition holds when it is known that agent's intention in carrying out y is to accomplish x and the agent's plan is rational, and the situation characterized by the agent's doing of y causes a situation in which the culmination of x holds.

1 Introduction

The problem of event individuation according to Bennett and Galton [7] is that of determining the criteria sufficient for distinguishing one event from every other. Various criteria proposed in the literature for event individuation include the specific details given about the event otherwise known as *event particulars* [20], the event's causes and effects, which are also events [10] and spatiotemporal regions [21, 25]. Unwin [30] identified three aspects (which he actually called instances) of the event individuation problem, which we regard as some kind of refinement of the event individuation problem. These aspects identified in Unwin's refinement of the event individuation problem are stated thus:

1. "Will the addition of adverbial modifiers to an event designator alter its reference?"

2. "If A does x by doing y, is A's doing of x the same [event] as A's doing of y?"

3. "Can two events completely occupy the same time-space zone?"

Unwin concluded that Quine' s criteria for event individuation, which are based on spatio-temporal regions, support the intuition that the addition of an adverbial modifier to an event designator will not alter its identity (just as Sebastian's stroll down streets of Bologna, is the same as Sebastian's leisurely stroll down streets of Bologna if both are known to have taken place at the same time), but not the intuition that two events can completely occupy the same time-space zone (just as the heating of a metal ball and its simultaneous rotation through 35 degrees are two different events). On the other hand, Kim's criteria, which uses all the details stated in the event description, is not against the possibility that two events may completely occupy the same time-space zone while it does not support the intuition that the addition of an adverbial modifier to an event will not alter its identity. Thus a set of criterion that can make an ontological commitment with respect to all these questions is needed.

Our position in this paper is that treating events as *located situation* enables us to make a definite various commitments with respect to each of the above questions. A located situation is defined by a *fluent* that fully characterizes it and its spatiotemporal context. A fluent is the descriptive essence of an event or state, or some other eventuality.

Pietroski [23] adopts the approach that actions must be viewed as *tryings*, such that if an agent wishes to carry out an action x, she tries something even if that thing she tries is not x. Thus for example if an agent tries to do x by doing y, x should be treated *as part of the agent's mental events*. However, there are three potential problems with treating such actions as mental actions. Firstly, there is a need to define the basis for knowing when such a mental action leaves the mental realms and becomes physical reality. Secondly, the mere fact that an agent carried out y that resulted in her doing x does not necessarily imply that the agent conceived a mental action x. The agent may have a different objective than x for carrying out the action y and it only happens that the action y eventually results in x. Thus our motivation in this paper is to develop a logical theory that answers these three questions with respect to an agent doing x by doing y:

- Is doing x the same as doing y?
- What are the bases for determining that x has been done given that y has been done?
- What should be the bases for determining the fact that an agent's action of doing y is part of the plan to accomplish x?

The logical theory proposed in this paper advocates treating events as located situations so that if an agent does x by doing y, then the doing of y should be treated as a sub-situation of doing x. At the same time, as we argue, treating events as located situations enables one to commit to events which identity are not altered by adding adverbials and to events that cannot occupy any spatiotemporal region with any other event that is fully characterized by the same fluent.

All the fluents in the theory presented are categorized along Vendler's aspectual lines. Vendler categorized verbs along aspectual lines as *activities, states, achievements* and *accomplishments*. If in trying to do Φ an agent must carry out some other action Ψ, then every successful incidence of Φ must be an accomplishment. In Vendler's aspectual categorization of verbs, an accomplishment is of non-zero duration and every accomplishment culminates in an achievement which is of (almost) momentary duration [29]. Based on this analysis of fluents, the sufficient condition for inferring that the action Φ has been carried out by an agent that has done Ψ is the following: *that the action* Ψ *causes the culmination of the doing* Φ. When that happens, the agent is deemed to have done Φ by doing Ψ and the situation in which the agent did Ψ is deemed in our logical theory to be a sub-situation of the situation in which she did Φ. However, if in addition we can somehow show that doing Φ was the agent's reason before for undertaking the action, Ψ, and that the agent is rational with respect to that plan, then, we should be able to infer that the situation in which Ψ was done was a part of the plan to accomplish a situation in which the agent accomplishes Φ. In this regard, our logical theory includes a ternary logical relation that defines an agent's intention or goal in carrying out an action. In addition to this there is also a ternary relation that describes the fact that an agent's plan is rational. Rationality is defined in terms of ceteris paribus causation relation between fluents.

The rest of this paper is organized as follows: Section 2 discusses the background literature on the meaning of event types, event instances and their individuation, situations in the Artificial Intelligence (AI) literature and Vendler's aspectual categorization of verbs. Section 3 presents a first-order logic treating events as situations in which the fluents defining the situations are categorized along Vendler's aspectual lines and an intentionality relation is defined between an agent's participation in an event and the intended outcome on the part of that agent. In section 4 we demonstrate how such a theory allows us to define the nature of intentional events such as A does x with the intention of doing y, as well as strategic events such as A does x with the intention of getting B to do y.

2 Literature Background

This section begins in subsection 2.1 with a discussion of our perception of the nature of events as drawn from philosophical and artificial intelligence (AI) literature. That sets the stage for a discussion on the event individuation debate from the literature in subsection 2.2. Subsection 2.3 discusses the notion of situations from the AI literature. Finally subsection 2.4 discusses the literature based on Zeno Vendler's aspectual classification of verbs which becomes relevant in analyzing the actions involved in the events.

2.1 Events as Instances of Event Types

An issue that arose from the discourse in modelling events is whether events should be treated as event types or as event instances [15]. Event instances are individual events. Davidson [12] and his descendants in the knowledge representation literature such as Bennett and Galton [7] and Galton [15] define event instances around verbs and the relationship they create between the agent of the verb and its object(s). Davidson created an event form that associates with each verb, an "existentially quantified event token variable". Thus to represent the event described as "Pat cooked spaghetti yesterday" he would write:

$$\exists e.\ \mathrm{Cook}(e, \mathrm{pat}, \mathrm{spaghetti}) \wedge \mathrm{Happen}(e, \mathrm{yesterday}).$$

Bennett and Galton [7] would have represented the statement "Pat is cooking spaghetti late" as:

$$\exists e.\ \mathrm{Cook}(e, \mathrm{pat\ spaghetti}) \wedge \mathrm{Prog}(e) \wedge \mathrm{Late}(e)$$

In the above formula *Prog* represents the progressive property for asserting that an event is on-going and *Late* represents a property for asserting that an event is late. On the other hand an *event type* according to Galton [15] is an abstract or universal entity for which actual events are realizations known as event tokens or instances. For example, the event "Pat cooked spaghetti yesterday" is an instance of the universal event class or event type in which Pat cooks spaghetti. The instance of the event mentioned in the statement happened yesterday. In notational terms, Galton reckoned that Allen's reified logic representation of the statement: "John saw Mary yesterday" as Occurs(see(john, mary), yesterday) implies that he was stating that an event instance of the type defined by John's action of seeing Mary (denoted by see(john, mary)) happened yesterday. Thus the event type defined by John's action of seeing Mary is different from any other event type. For example, it is different from another event type defined by Mary's action of dancing with John. An event type is therefore defined by a kind of action involving specific agents and objects. For example, the event type dance(mary, john) is defined from a *dance-with* action undertaken by an agent called Mary and the object of her action is an person called John. There are two important points to note about event types.

One may obtain a sub-type by adding location and time information to the event type. For example, every dance between John and Mary that takes place at Trafalgar square may be viewed as instances of a subtype of the type that generalizes all John and Mary dances, because it refers to a subset of all John and Mary dances. So may all the John and Mary dances that take place at Trafalgar Square on Sundays between 4 and 5 pm. However, for us, the notion of event types refer to a specific relations between specific domain entities that is unfiltered with time and location information.

Secondly, while action modifiers (such as late, suddenly, vigorously etc.) are captured as properties of events instances in Davidson's representation of event tokens, modifiers cannot be regarded as aspects of event type definitions. For example there is an event type defined by *Mary's dance with John*; but there can be no event type defined by *Mary's vigorous dance with John.* Note that that the event type vigorously-danced –with(mary, john) is a subtype of danced-with(mary, john). We will end up multiplying the number of event types by as many modifiers as can possibly modify each action.

By the way they are defined event types are generics of event instances. Event types have formed the basis for formalizing the notion of event repetition [5] because it takes generics to talk about being repeatable. While event individuation is a problem about event instances, event types may be of help as we see in the next section.

2.2 Event Individuation in the Literature

In the philosophical and AI literature, the various identity criteria proposed for events include causation (i.e. what causes the event and what the event causes) [10] and the event's time and place [25, 21], event's particulars [20] and event type and time [34]. Each of these proposals has its drawbacks. For example, Kim's proposal will reckon that a stroll by Sebastian and a leisurely stroll by the same person should be regarded as two different events even if they both take place at the same time and location. Davidson's proposal [10] has been famously criticized for its circularity by Quine [25]. Quine and Lemmon's proposal also have a drawback: the fact that certain events can share the similar spatiotemporal zones. Consider the example: Sebastian walked around the beach while sun tanning. Going by the proposal of Lemmon [21] and Quine [25] the beach walking event and the sun tanning event will be the same event.

Cleland [9] repeated Davidson's example of two events sharing the same spatio-temporal region. These involve the *rotation of a sphere* which is also *changing colour* at the same time. Cleland claims that both events cannot be identified as separate from one another. However, going by our earlier discussion about event types in subsection 2.1, the rotation of a sphere is of a different event type from the event of its colour changing. Thus in a case such as this, event types can be used to distinguish between two events that occupy the time-space region. Before event types and time-space region can become generally applicable as sufficient event individuation criteria, it must be the case that *it is not possible for more than one event belonging to the same event type to occupy the same time-space zone*. This assumption seems intuitively appealing considering the fact that it is difficult to find an example of an event of the same type sharing the same time-space region. It is important to apply this line of thought to the broader problem of event individuation.

From the foregoing analysis in this paper therefore, we wish to argue that by slightly extending Vila and Reichgelt's criteria [34] for individuating events with spatial location, we can make specific ontological commitments with respect to two of the three sub-problems of the event individuation problem. Firstly, if as we have argued earlier, event types are taken to be invariant to adverbial modifiers, then by combining event type with time and spatial location as event individuation criteria, we are definitely committed to the ontological position that *adding adverbial modifiers to an event designator will not alter its reference.*

Secondly, if considering the difficulty of finding an example of two events of the same type occupying the same exact time-space zone leads us to conclude that *no two events of the same (most specific) event type can completely occupy the same exact time-space region*, then, again by combining event-type with time and spatial location as event individuation criteria we are definitely committed to the ontological position that *no two events of the same type can completely occupy the same time-space zone*.

In a similar vein there is an existing proposal in ontological literature that an event can be individuated by its relational *essence* (which for us is its event type), and the event's place and time (also known as *scenes*) [16]. While we still cannot rule out two different events of the same type taking place in the same *broad* location at the same time, however, one may argue, following Guarino and Guizzardi [16], that by sufficiently *focusing* on the sub-location of the broad location, it will be possible to delineate one event from another of the same type. For example, if two conferences with the same name are holding at the same time at the same Hyatt's hotel, say, the room in which each event is holding is a focus on the broad location that is sufficient for delineating one conference event from another. Thus it is arguable that two different events of the same type cannot occupy the precise space-time zone.

The next section discusses situations in the form in which they will be used in this paper.

2.3 Situations

Schubert [28] identified two approaches for incorporating "situations" into the semantic representation of sentences. Both of these two approaches influenced more contemporary approaches. The first approach which he attributed to Davidson [12] allows only a single event relationship (denoted by verb-predicates or fluents) to be assigned to situations, (such as Stab(brutus, caesar, e)). When a Davidson event token is identified as a stabbing, even though many assertions can be made about the stabbing event token, it must remain only a stabbing and nothing more. It cannot be both a stabbing and a killing. The implication of this is that none of Davidson's event tokens can belong to more than one event type. However, unlike Davidson's notion of event tokens, it is important to note that the kinds of event tokens introduced by Kautz [19], may be associated with more than one distinct event type. For example, an event token may be an instance of both of the event types: *Make-Spaghetti-Dish* and *Make-Marinara-Dish*. That can be the case if the event token is an instance of the event type *Make-Spaghetti-Marinara*.

The second approach credited to Reichenbach [26] allows more than one such sentence, for example, *Brutus stabbed Caesar* and *Brutus killed Caesar* to be assigned to (or characterize) the same situation. Reichenbach's approach influenced the Situational Semantics of Barwise and Perry [6]. This is also the case in the situation calculus by McCarthy and Hayes [22] as well as in episodic logic [28]. On the other hand Davidson's approach influenced the logics of Galton [15], Bennett and

Galton [7], Allen [2] and Allen and Fergusson [1]. Schubert prefers the situational semantics approach for the task of semantic representation of narrative texts.

A Davidsonian event token cannot be associated with two or more statements. So one cannot associate the stabbing and killing of Caesar with the same davidsonian event token thus:

$$\exists e.\ \text{Stab(brutus, Caesar, } e) \wedge \text{Kill(brutus, Caesar, } e)$$

On the other hand, one is allowed to associate both the stabbing and the killing with the same situation thus:

$$\exists s.\ [\text{stab(brutus, caesar)}]^{*}s \wedge\ [\text{kill(brutus, caesar)}]^{*}s$$

Each of stab(brutus, Caesar), denoting the fact of the stabbing of Caesar by Brutus, and kill(brutus, caesar) denoting the fact of the killing of Caesar by Brutus, are examples of *fluents*. Essentially a fluent is a description of what may hold true in a situation. A fluent may refer to a state, event or any other kind of eventuality [14]. However, for the sake of this paper, our focus is on events. This definition of fluents is somewhat more generic than the fluents in situational calculus [22], in which situations only describe the states of the world. Thus in situation calculus, fluents are partial descriptions of states of individuals in the world. On the other hand situations in situational semantics are more generic covering events, states and other kinds of eventualities. Our formalization here is akin to the forms in situation semantics. Associating a fluent with a situation then means the kind of eventuality defined by that fluent characterizes that situation, either partially or fully. A situation may be partially characterized by more than one fluent, while only one fluent can fully characterize a situation. Ordinarily, a situation is (partially) characterized by a potentially infinite number of fluents. However, it is possible to talk about a situation being fully characterized by a fluent if we focus on a certain aspect of a situation and define a new situation which is a sub-situation of the existing situation, thereby.

There is a closed additive operator + defined on fluents (under which the set of fluents are closed) for fluents such that when both fluents f and f_1 partially characterize a situation s, it is the case that the fluent $f + f_1$ also characterizes s either partially or fully. There is also a sub-situation relation between pairs of situations such that if $f_1 + f_2$ fully characterize s, then there is a situation, s_1 say, which is fully characterized by f_1 and which is a sub-situation of s. The sub-situation enables the precise definition of causation relations among situations. If for example situation s_1 is the cause of some situation s_3, then it is not the case that s is the cause of s_3 but its sub-situation s_1.

For Schubert [28], situations can be individuated by two main criteria: the fluent that fully characterizes them and the time of the situation. Thus, in the world of situations as they exist in Schubert [28], there is no possibility of associating two or more situations which are completely described by the same fluent with the same time frame. In other words, two situational variables that are fully characterized by the same fluent and have the same temporal extents are referents to the same situation. Thus, situations as they currently exist in the knowledge representation literature are location invariant. This kind of situation may be sufficient for the tasks of representing and reasoning with narrative texts, which were the tasks for which Schubert and his colleagues invented their FOL**. To see an example of a situation that cannot be location invariant, consider a situation defined a 0-arity raining fluent called *rain*. Two situation variables defined (fully characterized) by the same rain fluent, cannot be referents to the same situation, unless they happen at the same time and over the same spatial location. Thus location invariant situations as they currently exist in the literature are not sufficient for modelling events in general.

The notion of a fluent can be thought of as being roughly equivalent to the notion of an event type. This is because they both refer to universal class or kinds of situations. That is why fluents have been used as a basis for defining repetition within the context of situation calculus by Pinto [24] just as event types have been used similarly by Akinkunmi and Osofisan [5]. Thus, if we let situations that these fluents define be *located situations* such that two situational variables are referring to the same situation if and only if they are fully characterized by exactly the same fluents and they are associated with the same exact spatiotemporal region, then it means that treating events as located situation makes the same definite ontological commitment with respect the first and third sub-problems of the event individuation problem as event types and spatiotemporal regions. Those ontological commitments are the same that are made by using event types with time and spatial location as criteria for event individuation as discussed earlier.

Ultimately, treating events as situations, if an agent does x by doing y, then we wish to let the situation fully characterized by y be a sub-situation of some situation characterized by x. However, a sufficient condition for making that inference is discussed in sub-section 2.4.

2.4 Vendler's Aspectual Theory of Events

An aspect of literature that is relevant for the discourse in this paper is the literature on the nature of events that have their roots in Zeno Vendler's aspectual classification of verbs [33]. These aspectual classes include: *states, activities, achievements* and *accomplishments.* Drawing on Dowty's development of Vendler's classification, Rothstein [27] presents the classes thus: *States* are based on non-dynamic verbs such as *sleeping, activities* are based on open-ended cumulative actions such as *run, achievements* are events that are near instantaneous defined by a culmination, while *accomplishments* are events that take place over a time interval and have a natural end-point or culminations. Trustwell [29] defined the four aspectual classes around processes and their potential culminations. Both states and activities are *atelic* in the sense that they are open ended events without any known culmination. Telic events on the other hand have a known conclusion. We distinguish between two categories of telic events. One category of telic events consist of those events that are identified as quantized by Krifka [29] such as: building a snowman or a house. The other category is the category of cumulative events that have a conclusion. For example: *falling to the ground* (Galton, personal communication). The earlier category of telic events we will refer to as *quantized-telic* events while the other category will be called *cumulative-telic*.

An achievement is an instantaneous culmination such as *finding a dice.* An accomplishment is an extended process followed by a culmination such as *running a mile.* Sometimes, the literature can be confusing about the distinction between achievements and accomplishments. For example, Trustwell [29] categorized crossing a street as an achievement while categorizing building a house as an accomplishment. Perhaps that confusion may arise from differing perceptions about which process is instantaneous and which is not.

Based on Vendlerian analysis, Trustwell [29] identified four classes of events. These are *physical, intentional, strategic* and *analytical* events. Physical events are not defined by the intention of any participant, and as such can be reduced, according to Trustwell, to "configurations of event participants, associated with different forces". Intentional events are those for which a participant in an event has a goal in mind. Strategic events are events for which goals are formed and executed by a party that is not a participant in the event. Analytical events are events that can only be inferred by some analysis, not necessarily by a single observation or by some agent's intention. Examples of an analytical event include: *global warming* or *the event of a species going extinct* or even a national brain drain event.

One way of viewing events categorized as accomplishments by Vendler is taking them to be a composite of two events involved in a causation relationship [29]. In a significant part of the literature, however, (e.g. [29, 8]), the causing event is always believed to be an activity while the caused event is an achievement. Trustwell [29] for example, treats an accomplishment as an activity culminating in some achievement.

One thing is very clear from the literature; and this is the fact that *all accomplishments culminate in an achievement.* The culmination of a killing is the achievement of the victim's death, the culmination of a building process is the achievement of a built house just as the culmination of a shooting is the achievement of getting a victim shot. The departure between these examples and the pattern presented in the literature is that the processes that may possibly lead to the culmination in all these examples hardly qualify to be called Vendlerian activities because they are quantized in nature. The process leading to getting a house built is quantized in nature and is therefore not a Vendlerian activity. In fact, in the *stab/kill* and *trigger-pulling/shoot* examples, those processes are themselves achievements.

Going back to the argument of Pietroski [23] that actions must be viewed as tryings, one can argue that any action for which trying to do it must necessarily involve trying some other action must be a Vendlerian accomplishment. Examples of such accomplishments are building, killing, shooting and door opening. The physical existence of such an accomplishment is only deemed to hold if some predefined culmination of that accomplishment which is an achievement on the part of the agent, results from the agent's action. For example the culmination of a shooting by an agent is for the agent to achieve getting the victim shot. That culmination is an achievement by the agent, which can only result from pulling the trigger of a loaded gun pointed at the desired victim. Every culmination of an accomplishment is an achievement. Every achievement leaves behind the irreversible attainment of a goal, which we may regard as its effect. For example the effect of an agent's achievement of getting a victim shot is to attainment of the goal of putting the victim in a shot state, while the effect of reaching a mountain peak by an agent is the attainment of that goal on the agent's part.

Thus, if an agent conceives an accomplishment Φ, and carries out an action Ψ in order to execute that accomplishment, then if Ψ causes the achievement of a goal which is the culmination of Φ, then it is the case that the agent has carried out her accomplishment. In that case, Φ ceases to be just a mental event as it was at the point of conception, but an event with a physical manifestation. Besides this we must regard Ψ as a plan on the part of the agent, to accomplish Φ.

In section 3, we bring the insight of Vendler [33] and Trustwell [29] into the theory that we present in this paper by categorizing fluents that define situations along Vendlerian lines: states, activities, achievements and accomplishments. The language of located situation introduced in section 3 includes a function for deriving the culmination of an accomplishment, which is an achievement and another one for deriving the effect of an achievement which is state fluent. Similarly, the insight of Pietroski [23] is useful in this way: if a rational agent does x by doing y then if we know that the agent conceived x as a mental event before doing y, then we can infer that the agent's action y is part of the rational agent's plan to accomplish y.

3 Events as Located Situations with Causative Relationships

In what follows, we present a logical theory of located situations, which is complete with the notion of sub-situations and causation. Our notion of a situation is a localized form of Schubert's situation which is an enhancement of Schubert's situation or *episodes* [28] with location information. While the assertions associated with both can be described by fluents, the difference between Schubert's situation and our localized situation is the fact that while Schubert's situation can be individuated by their times of occurrence, localized situations require both time and location to individuate. We have motivated the need for localized situations of that nature by illustrating the example of a rain event treated as a situation, and have argued that one raining situation cannot be distinguished from another unless the location of the situation is known. Thus, location situations will provide a better vehicle for representing events than Schubert's situations.

In this section, we present a logic of localized situations henceforth referred to as situations. Subsection 3.2 discusses an axiomatization around situations and the sub-situation relationships that may exist between them. Subsection 3.3 introduces axiomatizations around two causation relationships. One is a direct causation relationship between pairs of situations named *Cause-of* while the other is *ceteris paribus* causation between pairs of fluents named *Causes*. Subsection 3.4 presents sufficient conditions for the completion of an accomplishment. Finally section 4 develops the axiomatizations needed to formalize examples of intentional and strategic events. In our representation, A *does* y is treated as a sub-situation of A *does* x. In addition we also formalize strategic events of the type: A *does* x *to* B in order to get B *to accomplish* y. In that case A *does* x *to* B is treated as a sub-situation of A *gets* B *to accomplish* y.

3.1 Language and Notation

The logical language is a many-sorted reified first order predicate logic with equality with standard semantics for all the known operators. The sorts are fluents F, situations S, and the domain of objects and individuals, D, time intervals T, time instants TP and location L. A fluent is a relational entity that describes partly or fully, an event or state or similar eventualities. The *Characterize* predicate denotes the idea of a fluent occurring over or partially characterizing a situation and *Characterize*** which denotes the idea that a fluent *defines* or *completely characterizes* a situation and similar to the ** operator which denotes full characterization in Schubert's FOL** [28]. There is the predicate: *Characterizemax*, which denotes the fact that a fluent maximally characterizes a situation in such a way that the eventuality the fluent describes cannot be extended beyond either side of the situation's temporal extent (i.e. the situation's beginning or end); There is also the predicate *Characterizemax*** which denotes the idea that a fluent completely characterizes a situation and the eventuality it describes cannot be extended beyond either side of the situation's temporal extents. The signature of each of these functions is the same and given thus:

Characterize: $F \times S \rightarrow$ Boolean
Characterize** : $F \times S \rightarrow$ Boolean
Characterizemax: $F \times S \rightarrow$ Boolean
Characterizemax** : $F \times S \rightarrow$ Boolean

Our language has a causation relation between pairs of situation denoted by the predicate *Cause-of* and a potential causation (i.e. all things being equal or *ceteris paribus*) relation between a pair of fluents, denoted by the predicate *Causes*. The *Causes* relation means a situation completely characterized by the first fluent can cause some situation that is characterized by the second fluent. Their signatures are given below:

Cause-of: $S \times S \rightarrow$ Boolean
Causes: $F \times F \rightarrow$ Boolean

The infix predicates $\prec$ and $\preceq$, which denote the ideas that one situation is a proper and improper sub-situation of another situation (in the sense that every fluent that is reckoned to occur in the former is also reckoned to occur in the latter and they happened at the same time) as well as the infix predicate symbols $\lhd$ and $\unlhd$ which respectively denote the relations proper and improper *temporal sub-situation of*. One situation is a temporal sub-situation of another if its temporal extents are part of the temporal extents of the other. The notation $\top$ denotes the relation of temporal overlap between two situations which means the time of one situation overlap with the other. Their signatures are:

$\preceq, \prec, \lhd, \unlhd, \top : S \times S \rightarrow$ Boolean

In addition to these, there are three temporal relations denoted as $\sqsubset, \sqsubseteq, \sqcap$ between time intervals, known respectively as proper subinterval, subinterval, and non-disjoint relations respectively.

$\sqsubset, \sqsubseteq, \sqcap : \top \times \top \rightarrow$ Boolean

Finally, we introduce the predicate MinContainSit which denotes a relation between a situation and two fluents. It means that the two fluents completely and maximally describe two situations that are both temporal part of the situation in the relationship, and no smaller situation contains the two situations described by the fluents. The signature is given thus:

MinContainSit: $S \times F \times F \rightarrow$ Boolean

The final predicate we introduce here is the one that describes the intention of an agent in bringing about a situation. The predicate Intent-to-accomplish denotes a ternary relation with the signature:

Intent-to-accomplish: $D \times S \times F \rightarrow$ Boolean

The relation means an agent of the sort D participates in the situation of the sort S with the intention of accomplishing a fluent of the sort F. In other words an agent forms a plan to carry out an accomplishment f by participating or instigating the situation s.

Ordinarily, fluents are structured as function applications. The functions that we introduce at this point are the basic fluent functions. There are fluents defined from unary functions which denote the state of some domain element:

dead, broken, drunk, shot, tired: $D \sim\!\rightarrow F$

The symbol $\sim\!\rightarrow$ in function signatures is used to denote the fact that the function being described is a partial function that is not necessarily defined for every member of the domain.

One class of fluents denote actions that are carried out by one agent from the domain on another domain entity.

shoot, stab, kill, get-tired, get-drunk, shoot, hold-down, fully-turn:
$D \times D \sim\!\rightarrow F$

The other functions are the infix operator $+$ defined as the operator for combining fluents as well as *time* and *context* functions for getting time and location. The signatures of the functions are given thus:

$+ : F \times F \sim\!\rightarrow F$
time: $S \rightarrow T$
place: $S \rightarrow L$
context: $S \rightarrow T \times L$

The additive nature of + implies that it is idempotent, commutative and associative. The *time* function returns the time over which a situation holds, while the *place* function returns the location of the situation. The *context* function returns both the time and location of the situation as pair.

Two other fluent functions introduced here are the sequence function, *seq*, and the maintain function, *maintain*. The sequence function is a temporal sequence of two fluents such that that there is no time gap between their occurrence. The function *seq* has the following signature:

$$\mathit{seq}\colon\ F \times F \rightarrow F$$

The *maintain* function, on the other hand, is a partial function that denotes the maintenance of a certain state (denoted by a fluent) by an agent. The signature of the function is:

$$\mathit{maintain}\colon\ D \times F \sim\!\rightarrow F$$

The *maintain* and *seq* functions are used in Example 2.4.3.

We introduce three functions that operate on time intervals. The *begin* function returns the time instant that corresponds to the beginning of some given time interval, the *end* function returns the time instant that corresponds to the ending instant of a time interval. The *common* function returns the longest time interval that is common to a pair of non-disjoint intervals. Their signatures are:

$$\mathit{begin},\ \mathit{end}\colon T \rightarrow TP$$
$$\mathit{common}\colon\ T \times T \sim\!\rightarrow T$$

The first definition in our formalization clarifies the meaning of the + function.

Definition 1. *The fluent $f + f_1$ characterizes a situation s if and if each of f and f_1 also characterizes s.*

$$\forall f, f_2, s.\ \mathit{Characterize}(f + f_1, s) \equiv \mathit{Characterize}(f, s) \wedge\ \mathit{Characterize}(f_1, s)$$

Definition 2. *A fluent fully characterizes a situation s if and only if no other fluent characterises s.*

$$\forall f, f_2, s.\ \mathit{Characterize}^{**}(f, s) \equiv\ \mathit{Characterize}(f, s) \wedge \forall f_1 . f \neq f_1 \supset \neg \mathit{Characterize}(f_1, s)$$

Definition 3. *The fluent $f + f_1$ Characterizes a situation s if and only if both f and f_1 characterize the situation s.*

$$\forall f, f_1, s.\ \mathit{Characterize}(f + f_1, s) \equiv \\ \mathit{Characterize}(f, s) \wedge\ \mathit{Characterize}(f_1, s)$$

The relations denoted by the predicate Characterize** is defined in terms of Characterize thus:

Definition 4. *A fluent completely characterizes a situation if and only if no new fluent can be added to it in order to also describe the situation.*

$$\forall f, s.\ \mathit{Characterize}^{**}(f, s) \equiv \mathit{Characterize}(f, s) \wedge \\ (\forall f_1.(f \neq f_1 \wedge \neg \exists f_2.f = f_1 + f_2) \supset \neg \mathit{Characterize}(f + f_1, s))$$

The purpose of Characterize** is that it enables us to define a news situation that focuses on certain aspects of what is going on in another situation. If a situation is characterized by several fluents, it is possible to define a new situation from it, with exactly the same temporal extents, which is fully characterized by a finite number of fluents. Note that in classical situation calculus as well as situational semantics (e.g. in FOL** [28]), it is generally assumed that situations are identified by the fluents that define them and the time of the situation i.e.

Two mentioned (Schubert's) situations are the same if exactly the same fluents Occur in them and both situations have exactly the same temporal extents.

$$\forall s.s = s_1 \equiv (\forall f.\ \mathrm{Characterize}(f, s) \equiv \mathrm{Characterize}(f, s_1)) \wedge\ \mathrm{time}(s) = \mathrm{time}(s_1)$$

This can be the case when fluents are defined around clearly identified objects as in a narrative. However, when fluents are defined around indefinite objects, such as: *A lion killed a dog*, as opposed to definite objects, such as: *lion36 killed dog180*, there is the need for a stronger context than just time for individuating situations.

For example, deciding the difference between lion36 and lion32 who are both male lions that look very much alike may be difficult. In that case, one needs more contextual information in order to individuate situations. Such information includes particulars such as time and place. We are assuming here that events are stationary; in other words that, all the temporal parts of events are in one place. We will define the context of a situation as a pair of the time and place of that situation.

$$\mathrm{context} : S \rightarrow T \times P$$

Thus we define context in Definition 2.3 below.

Definition 5. *The context of two located situations is the same if they share the same time, place.*

$$\forall s, s_1.\ \mathit{context(s)} = \mathit{context(s_1)} \equiv \mathit{time(s)} = \mathit{time(s_1)} \wedge \mathit{place(s)} = \mathit{place(s_1)}$$

It is some kind of generalization of contexts that led Akinkunmi [4] to propose the notion of qualification such that an event token e_{token} say is formed form an event type e_{type} by the application of a tokenization function f_T to a pair of the event type and qualification q such that: $e_{\text{token}} = f_T(e_{\text{type}}, q)$, such that q carries all the information about time and location and others.

If we are dealing with situations that are not stationary, i.e. whose spatial extents may vary with time, for example, a situation evolving on a moving ship as discussed by Hacker [17], then we will need to turn place into two-argument functions so that contextual equality can be redefined as two situations that share the same time and for which the values of spatial extents and the other particulars at the same time instants are the same, thus:

$$\forall s, s_1.\ \text{context}(s) = \text{context}(s_1) \equiv \text{time}(s) = \text{time}(s_1) \wedge (\forall tp.tp \in \text{time}(s) \supset \text{place}(s, tp) = \text{place}(s_1, tp)$$

Thus, even when dealing with moving situations, the context of situations can easily be defined. With a clear definition of contexts, we can then individuate located situations using context thus:

Definition 6. *Two situational variables refer to the same situation if exactly the same fluents characterize the both of the situations they represent and their contexts are the same.*

$$\forall s, s_1.s = s_1 \equiv (\forall f.\ \mathit{Characterize}(f, s) \equiv \mathit{Characterize}(f, s_1)) \wedge\ \mathit{context(s)} = \mathit{context(s_1)}$$

Although for most domains, time and place should be sufficient for defining situational contexts, we cannot rule out the possibility that there are domains for which they are not sufficient.

There are different kinds of fluents in this paper according to Vendler's aspectual properties of verbs. These are states, activities, achievements and accomplishments. We will identify them by the kind of function that generates them and the kind of fluent for which they are defined. Therefore, we will present a number of partial functions defining these fluents and the fluents for which they are defined.

While there are activities that can be defined directly as fluents, there are activities that must be defined around other fluents. Firstly, we introduce a function that denotes the fact that an agent is engaged in some activity. This is the function act with the signature:

act: $D \times F \sim\!\!\rightarrow F$

The function *act* is defined for an agent and basic Vendler activity such as running, spinning and mixing. Examples of this are act(jack, mix(x, y)) or act(machine12, spin(fabric134)).

Secondly, we introduce a partial second-order function, *achieve*, for deriving achievement fluents from an agent and state fluents, which denotes the idea that a member of the domain (an agent) brings about some state fluent.

achieve: $D \times F \sim\!\!\rightarrow F$

An example of an achievement is achieve(tola, locked(door24)) which means Tola has achieved putting door24 in a locked state. The achieve function is particularly useful in deriving the culminations of Vendler's accomplishments. For example the achievement of a locked door for an agent is the culmination of a door locking accomplishment by that same agent.

Similarly, have an *accomplish* function for making accomplishment fluents thus:

accomplish: $D \times F \sim\!\!\rightarrow F$

The accomplish function takes a domain agent and another achievement or accomplishment and returns an accomplishment. Primarily, an accomplishment is an activity leading up to an achievement, Thus the application of the accomplish function to an individual and an achievement fluent results in an accomplishment that culminates in the achievement. For example accomplish(tola, achieve(tola, locked(door24))), is a fluent that describes an accomplishment that culminates in the locking of door24 by tola. An alternative way of stating this is to write: accomplish(tola, lock-door(tola, door24)) where it should be noted that lock-door(tola, door24) is an accomplishment.

The latter representation in which the fluent argument for the accomplish function is also an accomplishment is particularly useful when we need to specify strategic events for which an agent gets another agent to accomplish a task such as in this example:

accomplish(kola, lock-door(tola, door24)).

The achieve function distributes over the + operator on fluents thus:

$\forall x, f_1, f_2.\ \text{achieve}(x, f_1 + f_2) = \text{achieve}(x, f_1) + \text{achieve}(x, f_2)$

In general, achievements may have effects which are state fluents. For that purpose, we therefore introduce the partial function, *eff*, with the signature:

eff: $F \rightsquigarrow F$

The eff function is applicable to an achievement fluent and generates a state fluent. For achievement fluents defined in terms of the achieve function, the effect of such fluents can be defined in terms of the goal state thus:

$\forall x, f.\ \text{eff}(\text{achieve}(x, f)) = f$

Another partial function with a similar signature to eff is the culmination function culm that is only defined for fluents that are accomplishments. It returns the culminating achievement of the accomplishment. For example fact that x kills y culminates in x achieving putting x in a dead state is expressed thus:

$\text{culm}(\text{kill}(x, y)) = \text{achieve}(x, \text{dead}(y))$.

In fact $\text{kill}(x, y)$ is also the same fluent as $\text{accomplish}(x, \text{achieve}(x, \text{dead}(y))$. Thus we can further define the culmination of an accomplishment defined in terms of an accomplishment thus:

$\text{culm}(\text{accomplish}(x, \text{achieve}(x, \varphi))) = \text{achieve}(x, \varphi)$.

However if there are accomplishments defined in terms of other accomplishments, such as when an agent gets another agent to act in a certain kind of way, their culmination the first agent's accomplishment is defined in terms of the culmination of the second agent's accomplishment thus:

$\text{culm}(\text{accomplish}(x, \text{accomplish}(y, \varphi))) = \text{culm}(\text{accomplish}(y, \varphi))$.

It is important to note at this point that the effect of a fluent can be a state, such as when a fluent denotes the opening of a door by an agent or it can be another event fluent, such as when an agent instigates another agent to effect the opening of the door. In the former case the effect of the fluent: achieve(x, open(door)) is open(door). In the latter case, the effect of the fluent: achieve(x, achieve(y, open(door)) is achieve(y, open(door)). In the latter case, x got y to open the door. The *achieve* and *eff* functions appear in section 4 and in the examples in subsection 4.1.

3.2 Sub-situations and Temporal Sub-situations

A (localized) situation like an episode can be characterized by many fluents. As such, each of those fluents can be considered as defining sub-situations. There is also the need to represent the notion of a causation relationship between two situations. Thus there is a need for *concurrent sub-situation* relationship between two situations. A sub-situation of a situation is another situation within the same context that describes some relevant aspect of a certain situation.

Definition 7. *A situation s is a sub-situation of s_1 if and only if every fluent that characterizes s, also characterizes s_1 and both s and s_1 share the same context.*

$$\forall s, s_1 . s \preceq s_1 \equiv (\forall f.\ \mathit{Characterize}(f, s) \supset \mathit{Characterize}(f, s_1)) \wedge\ \mathit{context(s)} = \mathit{context(s_1)}$$

Definition 8. *A situation s is a proper sub-situation of s_1 if and only if s is a sub-situation of s_1 that is not equal to s_1.*

$$\forall s, s_1 . s \prec s_1 \equiv s \preceq s_1 \wedge s \neq s_1$$

Another kind of sub-situation relation is one in which the time of one situation is a temporal part of the time of another, while other aspects of the context remains the same. This is regarded as a temporal sub-situation relation defined next.

Definition 9. *Situation s is a proper temporal sub-situation of s_1 if and only if time of s is also within the time of s_1, while the other aspect of the context remains the same.*

$$\forall s, s_1 . s \lhd s_1 \equiv \mathit{time(s)} \sqsubset \mathit{time(s_1)} \wedge \mathit{place(s)} = \mathit{place(s_1)}$$

Definition 10. *A situation s is a temporal sub-situation of another situation s_1 if and only if s is either a proper sub-situation of s_1 or s is the same as s_1.*

$$\forall s, s_1 . s \unlhd s_1 \equiv s \lhd s_1 \vee s = s_1$$

Theorem 11. *If a situation is a sub-situation of another, then it is also a temporal sub-situation of the same situation.*

$$\forall s, s_1 . s \preceq s_1 \supset s \unlhd s$$

3.3 Causation and Consequences

There are certain basic facts about causation relations among situations. The most fundamental of these is the fact that a causation relation can only be defined among two completely defined situations.

Axiom 12. *If the situation s causes some other situation, then s can be characterized fully and maximally by some fluent.*

$$\forall s, s_1\ \mathit{Cause\text{-}of(s, s_1)} \supset \exists f, f_1 .\ \mathit{Characterizemax}^{**}(f, s) \wedge \mathit{Characterize}^{**}(f_1, s_1)$$

The cause of a situation is also the cause of all its sub-situations as well as all its temporal sub-situations. These ideas are expressed as Axioms 13 and 14 below.

Axiom 13. *If the situation s_2 is a sub-situation s_1, then whatever causes s_2 causes s_1.*

$$\forall s_1, s_2. s_2 \preceq s_1 \supset (\forall s.\ \textit{Cause-of}(s, s_1) \equiv \textit{Cause-of}(s, s_2)\ )$$

Axiom 14. *If the situation s_2 is a temporal sub-situation s_1, then whatever causes s_2 causes s_1.*

$$\forall s_1, s_2. s_2 \trianglelefteq s_1 \supset \forall s.\ \textit{Cause-of}(s, s_1) \equiv \textit{Cause-of}(s, s_2)$$

No sub-situation or temporal sub-situation of the cause of any particular situation can be taken to be the cause of that situation. Those ideas are expressed as Axioms 15 and 16.

Axiom 15. *If the situation s causes another situation s_1 then no proper temporal sub-situation of s can be the cause of s_1.*

$$\forall s, s_1.\ \textit{Cause-of}(s, s_1) \supset \forall s_2. s_2 \triangleleft s \supset \neg\ \textit{Cause-of}(s_2, s_1)$$

Axiom 16. *If the situation s causes another situation s_1 then no proper sub-situation of s can be the cause of s_1*

$$\forall s, s_1.\ \textit{Cause-of}(s, s_1) \supset \forall s_2. s_2 \prec s \supset \neg\ \textit{Cause-of}(s_2, s_1)$$

The next sub-section uses the notions of causation defined here and the notion of an accomplishment's culmination to define sufficient conditions for defining the existence of a physical accomplishment.

3.4 Accomplishments and Situations

If it happens that an agent has done x by doing y, irrespective of their intentions in doing y, it is important to be able to determine the conditions for knowing when x has done. As we have argued earlier, x must be an accomplishment with a culmination defined around a specific achievement.

In this sub section, we define a sufficient condition for deciding when such an action, x has been accomplished by an agent doing y. A sufficient condition for completing an accomplishment is when a situation defined by some other action causes another situation fully characterized by the culmination of that accomplishment, which has temporal extents that begins where the causing situation ends.

Axiom 17. *A sufficient ground for an accomplishment to characterize a situation is for another situation to cause another situation characterized by the culmination of that accomplishment and the temporal extent of the causing situation ends at the same time as the caused situation starts. In that case the causing situation is deemed to be a sub-situation of the situation characterized by the accomplishment.*

$$\begin{aligned}&\forall f.\ \mathit{Accomplishment}(f)\wedge\\ &\qquad \exists f_2, s_1, s_2.\ \mathit{eff(culm(f))} = f_2\wedge\ \mathit{Characterize}^{**}(f_2, s_2)\wedge\\ &\qquad\ \mathit{Cause\text{-}of}(s_1, s_2)\wedge\ \mathit{end(time(s_1))} = \mathit{begin(time(s_2))}\\ &\qquad\quad \supset \exists s.\ \mathit{Characterize}(f, s) \wedge s_1 \preceq s).\end{aligned}$$

In the next section, we extend the logical theory to be able to identify a case in which a rational agent conceives doing y as a plan for accomplishing x and does so. The theory will therefore be extended in section 4 to formalize intentional and strategic accomplishments.

4 Intentional and Strategic Accomplishments

Pietroski [23] has argued that actions must be viewed as *tryings*. For example, "trying to [shoot] is doing something even if [that thing] is not [shooting]". In order to go about shooting Abraham Lincoln, Daniel Booth pulled the trigger of a loaded gun that is pointed at Lincoln. When an agent decides to carry out an accomplishment, they must decide what actions to undertake in order to get there. Thus in trying to accomplish x by doing y, an agent first conceives the accomplishment of x. This is what Pietroski [23] argues should be part of an agent's mental events. In the last section we have defined sufficient conditions for determining when x has been accomplished from doing y. That inference was neutral about whether or not x was pre-conceived before y was undertaken as a means of accomplishing it.

In this section, we extend the existing logical theory by presenting a sufficient condition for deciding whether or not doing y was an agent's plan for accomplishing x. Those conditions are based on knowing whether an agent's intention in doing y, was to accomplish x and whether the agent was rational with respect to that intention. We will define the rationality of an agent's plan in terms of a ceteris paribus causation relationship between action fluents.

For example, when Booth decided to shoot Lincoln, he decided to pull the trigger of the loaded gun pointed at Lincoln. As an agent that knows enough about the domain of guns and shooting, Booth must have known that the pulling of a trigger of a loaded gun can lead to a situation fully characterized by the shooting of the victim. In other words, he must have known there is a *ceteris paribus* causation relation between pulling the trigger of a loaded gun pointed at his victim, and the culmination of the accomplishment of the victim's shooting, *all things being equal.* This relation is denoted by the *Causes* predicate between fluents.

Every agent should be able to infer what fluent can cause another by learning from the experience of actual causation experiences thus:

Axiom 18. *When a situation s fully characterized by f causes another situation s1 which is fully characterized by f_1 then a ceteris paribus causation relation has been established between f and f_1.*

$$\forall s, s_1, f, f-1.\ \textit{Cause-of}(s, s1) \wedge\ \textit{Characterize}^{**}(f, s) \wedge \\ \textit{Characterize}^{**}(f_1, s_1) \supset \textit{Causes}(f, f_1)$$

This axiom does not rule out that the possibility of the agent acquiring knowledge about the ceteris paribus relation from other sources apart from learning from experience. If the only means by which an agent can know what can cause what is by experience, then we could strengthen all the axioms like 4.1 above into a definition for ceteris paribus causation. However, we stop short of doing that, so that we can accommodate knowledge of ceteris paribus causation from sources other than experience (e.g. tradition).

Informally, an agent effects or instigates a situation s (described by a fluent φ) with the intention of bringing about an accomplishment ϕ. Thus when an agent desires ϕ and believes (rightly or wrongly) that doing φ will lead to the desired accomplishment, then s/he does φ. The predicate Intent-to-accomplish denotes the ternary relation between an agent, a situation s/he instigates and a fluent s/he wishes to accomplish. The predicate means, an agent instigates a situation s in order to achieve the goal fluent f. The signature is:

Intent-to-accomplish: Agent $\times$ Situation $\times$ Fluent $\rightarrow$ Boolean

Examples that can be formalized using the predicate are given thus:

1. Sam drinks with Neil in order to get Neil drunk.

$$\exists s.\ \text{Intent-to-accomplish(sam, } s\text{, get-drunk(sam, neil))} \wedge \\ \text{Characterize}^{**}\text{(drinks-with(sam, neil), } s\text{)}$$

2. In pointing the gun at Lincoln and pulling the trigger, Booth intended to have him shot.

$$\exists s.\ \text{Intent-to-accomplish(booth, } s\text{, get-shot(booth, lincoln))} \wedge \\ \text{Characterize}^{**}\text{(point-at(booth, gun17, lincoln) +} \\ \text{pull-trigger(gun17), } s\text{)}$$

3. Uju dances with Lagbaja in order to get Lagbaja tired.

$$\exists s.\ \text{Intent-to-accomplish(uju, } s\text{, get-tired(uju, lagbaja))} \wedge \\ \text{Characterize}^{**}\text{(dance-with(uju, lagbaja), } s\text{)}$$

4. Sam stabs Carlos in order to get Carlos killed.

$$\exists s.\ \text{Intent-to-accomplish(sam, } s\text{, kill(sam, carlos))} \wedge \text{Characterize}^{**}\text{(stabs(sam, carlos), } s\text{)}$$

5. Mike strikes the rock with a hammer in order to break it.

$$\exists s.\ \text{Intent-to-accomplish(mike, } s\text{, break(mike, rock))} \wedge \text{Characterize}^{**}\text{(strike-with-hammer(mike, rock), } s\text{)}$$

It is important to note that some achievement fluents of the sort defined by the achieve function given in section 3.1, are related to accomplishment fluents through the applications of the culminations function, culm.

$$\forall x, y.\ \text{culm(get-drunk}(x, y)) \equiv \text{achieve}(x, \text{drunk}(y))$$
$$\forall x, y.\ \text{culm(get-tired}(x, y)) \equiv \text{achieve}(x, \text{tired}(y))$$
$$\forall x, y.\ \text{culm(kill}(x, y)) \equiv \text{achieve}(x, \text{dead}(y))$$
$$\forall x, y.\ \text{culm(break}(x, y)) \equiv \text{achieve}(x, \text{broken}(y))$$
$$\forall x, y.\ \text{culm(shoots}(x, y)) \equiv \text{achieve}(x, \text{shot}(y))$$

It is also important to recall that for every achievement fluent of the sort on the left hand side of the equations above, the application of the effect function, *eff* will yield the goal that the achievement brought about. For example:

$$\text{eff(achieve}(x, \text{drunk}(y)) = \text{drunk}(y).$$
$$\text{eff(achieve}(x, \text{tired}(y) = \text{tired}(y).$$
$$\text{eff(achieve}(x, \text{dead}(y)) = \text{dead}(y).$$
$$\text{eff(achieve}(x, \text{broken}(y)) = \text{broken}(y).$$

We can now define what makes an agent rational with respect to their plan or expectation to accomplish a specific goal by bringing about a situation. According to Davidson [13], "a reason *rationalizes an action* only if it [i.e. that reason] leads us to see something the agent saw, or thought he saw, in his action-some feature, *consequence*, or aspect of the action the agent wanted, desired, prized". Thus every intent-to-accomplish assertion gives a reason to rationalize an action. That reason therefore is rational if it can be shown that the situation the agent participated in can lead to the desired outcome which is the effect of the desired accomplishment's culmination.

Thus, an agent is rational with respect to a plan to bring about the accomplishment fluent f by participating in a situation s if the effect of the culmination of f can be caused by the fluent that defines that situation. In other words, if what s/he intends to accomplish by bringing about a particular situation is feasible.

Definition 19. *An agent is rational with respect to using a situation s to bring about accomplishment f, if and only if the effect of that accomplishment's culmination can be caused by the fluent defining the situation s.*

$$\forall x, s, f.\ \textit{Rational-wrts}(x, s, f) \equiv$$
$$(\exists f_1.\ \textit{Intent-to-accomplish}(x, s, f) \wedge$$
$$\textit{Characterize}^{**}(f_1, s) \wedge\ \textit{eff(culm(f))} = f_2 \supset \textit{Causes}(f_1, f_2)\ )$$

With the definition of an agent's rationality, we are now ready to define a plan-oriented relationship between an accomplishment and the action that leads to it. A Plan-for relation exists between two situations if the action that characterizes the first situation was conceived to lead to the accomplishment that characterizes the second situation. The signature for Plan-for is:

Plan-for: $S \times S \rightarrow$ Boolean

A definition for Plan-for is presented below:

Definition 20. *A situation s is deemed to have been a part of a plan for accomplishing a situation s_1 if and only if the intention of an agent x in carrying out s is to accomplish f, and x is rational with respect to the intention to accomplish f by doing s and f characterizes s_1 and s is a temporal situation of s.*

$$\forall s, s_1.\ \textit{Plan-for}(s, s_1) \equiv \exists f.\ \textit{Intent-to-accomplish}(x, s, f) \wedge$$
$$\textit{Characterize}^{**}(f, s_1) \wedge s \trianglelefteq s_1 \wedge\ \textit{Rational-wrts}(x, s, f).$$

The next axiom gives sufficient condition for deciding whether or not a situation is a plan for bringing about another situation.

Axiom 21. *If an agent a participates in or triggers a situation s in order to bring about an accomplishment f and the agent is rational with respect to using the situation s to accomplish f, and a succeeding situation s_1 arises that is characterized by effect of culmination of that accomplishment and (s_1) is also caused by s, then a has brought about the accomplishment f which he set out to bring about, and the situation s is deemed to be a plan for the situation s_2.*

$$\forall a, s, \varphi, f_1, f_2.\ \textit{Intent-to-accomplish}(a, s, f) \wedge\ \textit{Rational}(a, s, f) \wedge$$
$$\textit{Characterize}(f_2, s) \wedge\ \textit{eff(culm}(\varphi)) = f_1 \wedge$$
$$\exists s_1.\ \textit{Characterize}^{**}(f_1, s_1) \wedge\ \textit{Cause-of}(s, s_1) \wedge\ \textit{end(time(s))} =$$
$$\textit{begin(time}(s_1)) \wedge\ \textit{Characterize}(f_1, s_1)$$
$$\supset \exists s_2.\ \textit{Characterize}^{**}(f, s_2) \wedge\ \textit{Plan-for}(s, s_2).$$

It is important to draw attention to the use of the *Intent-to-accomplish* predicate used here and its significance. The predicate allows the specification of both intentional and strategic events in the same sense as contained in Trustwell's analysis [29]. An intentional event is one carried out by an agent with a goal and the agent is involved in the event. On the other hand, a strategic event is one instigated by a party who is not part of the event and with a goal in the mind of the instigator.

The subsection below presents examples of how the ontological commitment made in Axiom 21 above can be used to formalize some examples of intentional events (Example 1 is about intentional events while Example 3 is about intentional non-atomic events). Example 2 focuses on a strategic event instance.

4.1 Examples

We now proceed to see examples for which situations and their causation relationships are useful. For the first example, clearly, a logic based on atomic events is inadequate. The first example takes care of the Brutus stab/kill Caesar problem by making the stabbing of Caesar by Brutus, a sub-situation of another in which Brutus kills Caesar.

Example 1 (Intentional Events)**.** *In this example, we are distinguishing between a stabbing event in which the actor has the intention to kill (as is the case with the Brutus and Caesar case) and other stabbing events in which the actor does not have the intention to kill. If the stabbing agent intends to kill and he succeeds then we treat the stabbing situation as part of the plan for accomplishing the killing situation.*

These treatments of intentional and unintentional events are presented below:

Any stabbing situation with the intention to make the victim dead by an agent who is rational with respect to a situation s, agent that causes the death of its victim is part of a killing situation of the victim by the agent.

$$\begin{aligned}
&\forall x, y, s.\ \text{Characterize}^{**}(\text{stab}(x, y), s)\wedge \\
&\quad \text{Intent-to-accomplish}(x, s,\ \text{kill}(x, y)) \\
&\quad \wedge\ \text{Rational-wrts}(x, s,\ \text{kill}(x, y))\wedge \\
&\quad \exists s_1.\text{Characterize}^{**}(\text{dead}(y), s_1)\wedge\ \text{Cause-of}(s, s_1)\wedge \\
&\qquad \text{end}(\text{time}(s)) = \text{begin}(\text{time}(s_1)) \supset \\
&\qquad\quad \exists s_2.\ \text{Characterize}(\text{kill}(x, y), s_2)\wedge\ \text{Plan-for}(s, s_2).
\end{aligned}$$

The above axiom associated with Example 1, is an instance of Axiom 21. To see this, it is important to note that both of the following are true:

$$\begin{aligned}
&\text{culm}(\text{kill}(x, y)) = \text{achieve}(x,\ \text{dead}(y)) \\
&\text{eff}(\text{achieve}(x,\ \text{dead}(y))) = \text{dead}(y).
\end{aligned}$$

The logic of intent presented here accurately mirrors the intuition of Pietroski [23]. In Pietroski's manner of speaking: *In an attempt to kill Caesar*, Brutus did something, which is stabbing Caesar. Thus Pietroski's thought can be translated into our logic of intent thus: *in trying to accomplish ϕ, an agent x does φ*, is represented in our logic as:

$$\text{Intent-to-accomplish}(x, s, \phi) \wedge \text{ Characterize}^{**}(\text{achieve}(x, \varphi), s).$$

Thus the killing begins as a conceived mental event in the mind of the agent as Pietroski rightly observed. Axiom 21 above gives the condition for the agent to be deemed to have accomplished his plan through a specific line of action. That condition is if we know that the agent has a rational plan to accomplish her goal ϕ through a specific action φ and the action caused the culmination of that accomplishment. The next example is one involving strategic events.

Example 2 (Strategic Events)**.** *Strategic events [29] are events that an agent instigates without being a participant. It is similar to intention event because the instigating agent hope the culmination of that event will result in the achievement of objective. For example, an agent can instigate a stabbing event without being part of it. For example, it is true that Cassius got Brutus to stab Caesar in a particular situation. Such a situation can be represented by using the accomplishment fluent thus:*

$$\exists s.\ \textit{Characterize}^{**}\textit{(accomplish(cassius, stab(brutus, caesar)), s)}$$

The following statement is formalized using the accomplishment fluent below:

a. The intention of Cassius in convincing Brutus to stab Caesar is to get him to stab Caesar.

$$\exists s.\ \textit{Characterize}^{**}\textit{(convince-to(cassius, brutus, stab(brutus, caesar)), s)} \wedge \textit{ Intent-to-accomplish(cassius, s, accomplish(cassius, stab(brutus, caesar)))}$$

It is important to note that strategic events such as the one above are generally accomplishments. The treatment of the relationship between a strategic event of this nature and the event that brings it about is explained in the following axiom schema:

Axiom 22. *If x convinces y to do φ to z and his intent in doing so is to get y to do φ to z and x was rational about his course of action to achieve his objective, then if y eventually does what he was directed to do as a result of x's action, then x is deemed to have succeeded in getting y to do φ to z over a situation towards which the convincing situation is part of the plan.*

$$\begin{aligned}
&\forall\varphi.\\
&\forall x, y, z, s, s_1, s_2.\\
&\quad \mathit{Characterize}^{**}(\mathit{convince}(x, y, \varphi(y, z)), s)\wedge\\
&\quad \mathit{Intent\text{-}to\text{-}accomplish}(x, s, \varphi(y, z))\wedge\ \mathit{Rational\text{-}wrts}(x, s, \varphi(y, z))\wedge\\
&\quad \exists s_1.\ \mathit{Characterize}^{**}(\varphi(y, z), s_1)\wedge\ \mathit{Cause}(s, s_1)\wedge\\
&\quad \mathit{end}(\mathit{time}(s)) \leq \mathit{begin}(\mathit{time}(s_1)) \supset\\
&\qquad \exists s_2.\ \mathit{Characterize}^{**}(\mathit{accomplish}(x, \varphi(y, z)), s_2)\wedge\ \mathit{Plan\text{-}for}(s, s_2)
\end{aligned}$$

Axiom 22 above captures the notion of the strategic agent telling getting or directing the convinced agent the exact thing to do. There are definitely cases in which the strategic agent only gives the convinced agent an accomplishment to carry out and that agent is to decide how to do it as seen in the following axiom.

Axiom 23. *If x convinces y to do ϕ to z, with the intention of getting y to do ϕ to z and x is rational with respect to his objective, then if y does φ to z as a result of the convincing until the action results in putting z in a state τ which is the culmination of doing ϕ to z, then x is deemed to have accomplished the task of getting y to do ϕ to z over a situation for which the directing situation is a part of its plan.*

$$\begin{aligned}
&\forall\phi, \tau, \varphi.\\
&\forall x, y, z, s, s_1, s_2.\\
&\quad \mathit{Characterize}^{**}(\mathit{convince}(x, y, \phi(y, z)), s)\wedge\\
&\quad \mathit{Intent\text{-}to\text{-}accomplish}(x, s, \phi(y, z))\wedge\ \mathit{Rational\text{-}wrts}(x, s, \phi(y, z))\wedge\\
&\quad \exists s_1.\ \mathit{Characterize}^{**}(\varphi(y, z), s_1)\wedge\ \mathit{Cause\text{-}of}(s, s_1)\wedge\\
&\quad \exists s_2.\ \mathit{Characterize}^{**}(\tau(z), s_2)\wedge\ \mathit{Cause\text{-}of}(s_1, s_2)\wedge\\
&\quad \mathit{eff}(\mathit{culm}(\phi(y, z))) = \tau(z)\wedge\\
&\quad \mathit{end}(\mathit{time}(s_1)) = \mathit{begin}(\mathit{time}(s_2)) \supset\\
&\qquad \exists s_3.\ \mathit{Characterize}^{**}(\mathit{accomplish}(x, \phi(y, z)), s_3)\wedge\ \mathit{Plan\text{-}for}(s, s_3).
\end{aligned}$$

Another justification for preferring situations to davidsonian events in representing events is the fact that causation is not necessarily always a relationship between events or situation that can be described by an atomic fluents or so-called atomic events. There are cases in which it takes a combination of events to cause another event or set of events. The next example demonstrates such a case.

Example 3 (A non-atomic Event)**.** *The following is an example from the planning domain arising from a paper of James Allen [2]. In that example, it takes both the full turning of the knob and holding down the latch of a door by an agent, at the same time, in order to open the door.*

Axiom 24. *If a rational agent fully turns the knob of a door while pressing and holding down its latch with the intention of opening the door, and that situation* *causes the door to open, then there exists a situation fully characterized by the opening of the door by the agent for which the situation characterized by knob turning and latch holding is part of the plan.*

$$
\begin{aligned}
&\forall s, x, d.\\
&\textit{Characterize}^{**}(f, s) \wedge \textit{Rational-wrts}(x, s,\ \textit{open-door}(x, d)) \wedge\\
&\textit{Intent-to-accomplish}(x, s,\ \textit{open-door}(x, d)) \wedge f =\\
&\textit{seq(hold-down}(x,\ \textit{latch}(d)),\\
&\textit{maintain}(x,\ \textit{held-down}(\textit{latch}(d))) + \textit{fully-turn}(x,\ \textit{knob}(d))) \wedge\\
&\quad \exists s_1.\ \ \textit{Characterize}^{**}(\textit{opened}(d), s_1) \wedge\ \textit{Causes}(s, s_1) \wedge\ \textit{end}(\textit{time}(s)) =\\
&\textit{begin}(\textit{time}(s_1))\\
&\qquad \supset \exists s_2.\ \textit{Characterize}^{**}(\textit{open-door}(x, d), s_2) \wedge\ \textit{Plan-for}(s, s_2).
\end{aligned}
$$

It is important to note that the culmination of the accomplishment *open-door*(x, d) is the achievement *achieve(x, opened(d))*. Thus:

$$
\begin{aligned}
&\text{culm(open-door}(x, d)) = \text{achieve}(x, \text{opened}(d)) \text{ and}\\
&\text{eff(culm(open-door}(x, d))) = \text{opened}(d)
\end{aligned}
$$

In this section we have seen the internal plan structures of intentional and strategic events. The first kind is intentional events in which A is acting for itself with the structure *A does x by doing y*. In that case the precondition is presented for which the situation in which A does y is a plan for A doing x. If the agent does y with the intention of accomplishing x and that causes the culmination of x in the immediate or long term, then the situation in which A does y is a part of the situation in which A does x.

The second one is strategic events in which an agent A is getting some other agent to act in some kind of way, for example, A directs B to do x. In the case that x is an accomplishment then B is free to choose how to carry out the accomplishment. Supposing B chooses to do x by doing y, then condition for knowing when A has accomplished a strategic goal of directing B to do x is when A directs B to do x and B does something that causes the culmination of x in the immediate or long term. In that case, the situation in which A does whatever he did is a part of the plan for getting B to accomplish x.

5 Summary

Starting from Unwin's refinement of the three aspects of Davidson's event individuation problem, this paper has argued that it is possible to commit to an ontological position with respect to the problem of whether adverbial modifiers can alter an event reference and the problem of whether or not two events can occupy the same spatiotemporal zone, by using identity criteria for events that combines event types and the time-space region. This follows, firstly, from the fact that event types (which are roughly equivalent to the notion of fluents) as they appear in the knowledge representation literature are invariant with respect to adverbial modifiers. Thus given that an event is identified by an event type and the time and space they occupy, "the addition of an adverbial modifier to an event designator", will not "alter its reference". Secondly, we argue that while two events may occupy the same exact time space zone, they may be differentiated from one another by their event types, as in the case of the rotating cylinder changing colour to red at the same time, which must be taken to consist of two events of types: the rotation of the cylinder and the changing of the cylinder's colour. Even when the two events of the same type do occupy the same (broad) time-space zone their exact time-space zones may be delineated by Guarino and Guizzardi [16] style focusing process.

Our argument is that treating events as located situations also helps to capture the same kind of ontological commitment with respect to first and third aspects of the refinement of the event individuation problem, made by using event types and spatio-temporal region. Subsequently, a logical theory of located situations is provided that enables committing to a middle ground between the "identificationist" and "anti-identificationist" positions identified by Pietroski [23], with respect to the second aspect of the refinement which is: *If A does x to B by doing y to B, is the event of A's doing of y to B the same as the event of A's doing of x to B*? That middle ground treats the situation in which A does y to B as a sub-situation in which A does x to B, rather than treating the two events as either the same event or as two different events.

We also argue that if the only way an agent can do x is by doing something other than x, then x must be an accomplishment, and the condition for the existence of a situation characterized by x is when the existence of a situation characterized by A doing y leads to the culmination of the accomplishment A doing x. In that sense we must infer that A doing x characterizes a situation for which A doing y is a sub-situation. If an agent intentionally does y in order to do x, then A doing x started as a mental event as observed by Pietroski [23], we infer that when x is accomplished by doing y then the situation in which A does y is a part of the plan for the situation in which A does x.

It is important to understand clearly the kind of ontological commitment that choosing to represent events with various knowledge representation structures entail. For instance, representing events as pure Schubert's situation commits us to having a domain in which two events described by the same fluents cannot happen at the same time even if those events are happening in two different locations. That is a restriction that may be suitable for the domain of narrative texts for which Episodic Logic was intended, but is definitely not suitable for the wider context of reasoning about events.

We must end this paper with a disclaimer: that the author is not an advocate of killings or violence in any way or form, and that the examples chosen in this paper should not be construed as suggesting so.

Acknowledgements

The author is grateful to Nicola Guarino, Antony Galton and Giles Kassel for insightful and helpful feedbacks on earlier drafts of the paper. He is also grateful to the anonymous referee for their helpful comments. The author is also grateful to Jane Spurr for important editorial help.

References

[1] Allen, J. F. and Fergusson G.M. (1994) Actions and Events in Interval Temporal Logic. *Journal of Logic and Computation*, 4(5), 531-579, 1994

[2] Allen, J.F. (1991) Planning as Temporal Reasoning. In Allen, J.F., Fikes R. and Sandewall, E. (eds) *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, Massachusetts, USA, 1991.

[3] Allen, J.F. (1984) Towards A General Theory of Action and Time. *Artificial Intelligence*, 23(2), 123-154, 1984.

[4] Akinkunmi, B.O. (2010) Temporal Entities – Types, Tokens and Qualifications. In J. Filipe, & J.L.G. Dietz (Eds.) *Proceedings of the International Conference on Knowledge Engineering and Ontology Development*, pp 288-294, Lisbon Scitex, 2010.

[5] Akinkunmi B.O. and Osofisan A.O. (2004) Temporal Properties of Repetitive Entities. *Artificial Intelligence Review*, 22(4), December 2004.

[6] Barwise, J. and Perry J. (1983) *Situations and Attitudes*. MIT Press, Cambridge, Massachusetts, 1983.

[7] Bennett B and Galton A.P. (2004) A Unifying Semantics for Time and Events. *Artificial Intelligence*, 153, 13-48.

[8] Copley B. and Harley, H. (2015) A force-theoretic framework for events structure. *Linguistics and Philosophy* 38, 103-158, 2015.

[9] Cleland, C. (1991) On the Individuation of Events. *Synthese* 86, 229-251, 1991.

[10] Davidson, D. (1969) The Individuation of Events. In N. Rescher (Ed.) *Essays in Honor of Carl G. Hempel* pp 216-234, Dordrecht, Reidel. Reprinted in In D. Davidson (Ed.) Essays in Action and Events pp 163-180, Oxford Clarendon, 2001.

[11] Davidson, D. (2001) *Essays on Actions and Events*. Clarendon Press, Oxford 2001.

[12] Davidson, D. (1967) The Logical Form of Action Sentences. In N. Rescher(Ed.) *The Logic of Decision and Action*, University of Pittsburgh Press. Reprinted In D. Davidson(Ed.) *Essays in Actions and Events*, pp 105 -121, 2001,Oxford Clarendon, 1967.

[13] Davidson, D. (1963) Actions, Reasons and Causes, *Journal of Philosophy*, 60(23), 685-700, 1963.

[14] Galton A. (2005) Eventualities In M. Fisher, D.M. Gabbay & Vila L. (eds) *Handbook of Temporal Reasoning in Artificial Intelligence*, pp. 25-58, Elsevier, Amsterdam, 2005.

[15] Galton, A. (1991) Reified Temporal Theories and How to Unreify Them. In Mylopoulos J. and Reiter R. (eds) *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, 1991, pp 1177- 1182.

[16] Guarino, N and Guizzardi G. (2016) Relationships and Events: Towards A General Theory of Reification and Truthmaking, In G. Adoni, S. Cagnoni & M. Maratea (Eds.) *Advances in Artificial Intelligence: 15th International Conference of the Italian Association for Artificial Intelligence*, LNCS, Springer Verlag, 2016.

[17] Hacker, P.M.S. (1982) Events and Objects in Space and Time. *Mind* 91(361):1-19, 1982.

[18] Kautz, H.A. (1991) A Formal Theory of Plan Recognition and its Implementation. In J.F. Allen, H.A. Kautz, R.N. Pelavin & J.D. Tenenberg (Eds.) *Reasoning about Plans*, (pp 69-126), San Mateo, Morgan Kaufmann Publishers, 1991.

[19] Kautz, H.A. (1987) *A Formal Theory of Plan Recognition*. PhD Thesis, University of Rochester, Rochester, New York, 1987.

[20] Kim, J. (1966) On the Psycho-Physical Identity Theory, *The American Philosophical Quarterly* III, 295-311

[21] Lemmon, J. (1967) Comments on D. Davidson's "The Logical Form of Action Sentences", In N. Rescher (Ed.) *The Logic of Decision and Actions*, pp 96 – 103, University of Pittsburgh Press, Pittsburgh.

[22] McCarthy, J. and Hayes, P. (1969) Some philosophical problems from the standpoint of artificial intelligence. In Meltzer B. and Michie, D. (eds.) *Machine Intelligence, vol.* 4, pages 463–502. Edinburgh University Press, Edinburgh, 1969.

[23] Pietroski, P. (2002) *Causing Actions*. Oxford University Press, New York, 288 pp, 2002.

[24] Pinto J.A. (1994). *Temporal Reasoning in the Situation Calculus*. PhD Thesis, Department of Computer Science, University of Toronto, Canada, 1994.

[25] Quine, W.V. (1985) Events and Reification In E. LePore and B.P. MacLaughlin (Eds.) *Perspectives on the Philosophy of Donald Davidson*, pp 162 -171, Basil Blackwell, New York, 1985.

[26] Reichenbach H. (1947) *Elements of Symbolic Logic*. MacMillan, New York, 1947.

[27] Rothstein, S. (2005) *Structuring Events : A Study in the Semantics of Lexical Aspects*. Cambridge University Press, 2005.

[28] Schubert, L.K. (2000) The Situations We talk About. In J. Minker (Ed.) *Logic-based Artificial Intelligence*, pp 407-439, Kluwer, Dordrecht, 2000.

[29] Trustwell, R. (2016) Event Composition and Event Individuation In Trustwell, R. (Ed.) *Oxford Handbook of Events*, (pp. 1 – 32), Oxford University Press, Oxford, 2016.

[30] Unwin, N. (1996) The Individuation of Events. *Mind* 105 Issue 418, 315-330.

[31] van Lambalgen, M. and Hamm F. (2005) *The Proper Treatment of Events*. Blackwell Publishers, Oxford, 243 pp

[32] Varzi, A. C. (2006) The Talk I was Supposed to Give. In Bottani A and Davis, R. (eds) *Modes of Existence: Papers in Ontology and Philosophical Logic*, pp. 131-152, Ontos Verlag, Frankfurt, 2006

[33] Vendler, Z. (1957) Verbs and Times. *The Philosophical Review* 66(2), 143-160, April 1957

[34] Vila, Ll and Reichgelt (1996) The Token-Reification Approach to Temporal Reasoning, *Artificial Intelligence* 83(1), 59-74.

 Received 14 November 2020