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 6 • Issue 1 • January 2019

Contents

Articles

Covariant-Contravariant Refinement Modal Logic
Huili Xing, Zhaohui Zhu and Jinjin Zhang 1

Did the Neo-Babylonians Construct a Symbolic Logic for Legal Proceedings?
Andrew Schumann 31

Weight Predication on Missing Links in Social Networks. A Cross-Entropy-Based Approach
Wilhelm Rödder, Andreas Dellnitz, Ivan Gartner and Sebastian Litzinger 83

AlphaGo's Decision Making
Woosuk Park, Sungyong Kim, Keunhyoung Luke Kim and Jeounghoon Kim 105

Strengthening Gossip Protocols using Protocol-Dependent Knowledge
Hans van Ditmarsch, Malvin Gattinger, Louwe B. Kuijer and Pere Pardo 157

Sponsored by
ELSEVIER

ISBN 978-1-84890-299-2
90000
9 781848 902992

Published by
IFCoLog

Volume 6 • Issue 1 • January 2019

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

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

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

Volume 6, Number 1

January 2019

Disclaimer

ISBN 978-1-84890-299-2
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 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.spurr@kcl.ac.uk) as a pdf file, preferably compiled in LaTeX using the IFCoLog class file.

Contents

ARTICLES

Covariant-Contravariant Refinement Modal Logic 1
Huili Xing, Zhaohui Zhu and Jinjin Zhang

Did the Neo-Babylonians Construct a Symbolic Logic for Legal Proceedings? 31
Andrew Schumann

Weight Predication on Missing Links in Social Networks. A Cross-Entropy-Based Approach 83
Wilhelm Rödder, Andreas Dellnitz, Ivan Gartner and Sebastian Litzinger

AlphaGo's Decision Making 105
Woosuk Park, Sungyong Kim, Keunhyoung Luke Kim and Jeounghoon Kim

Strengthening Gossip Protocols using Protocol-Dependent Knowledge . . . 157
Hans van Ditmarsch, Malvin Gattinger, Louwe B. Kuijer and Pere Pardo

Covariant-Contravariant Refinement Modal Logic

Huili Xing
College of Computer Science and Technology
Nanjing University of Aeronautics and Astronautics, China

Zhaohui Zhu
College of Computer Science and Technology
Nanjing University of Aeronautics and Astronautics, China
zhaohui@nuaa.edu.cn

Jinjin Zhang
College of Information Science
Nanjing Audit University, China

Abstract

The notion of covariant-contravariant refinement (CC-refinement, for short) is a generalization of the notions of bisimulation and refinement. This paper introduces CC-refinement modal logic (CCRML) obtained from the modal system $\boldsymbol{K}$ by adding CC-refinement quantifiers, and provides a sound and complete axiom system for CCRML.

Keywords: modal logic, covariant-contravariant refinement, axiom system

1 Introduction

Recently, Laura Bozzelli et al. presented and explored refinement modal logic (RML) [8, 9], which provides a more abstract perspective of future event logic [20, 21]

The authors are grateful to the anonymous referees for their valuable suggestions, which have helped us to improve the presentation of the paper. In particular, the decidability of CCRML is pointed out by one of referees. This work received financial support of the National Natural Science Foundation of China (No. 61602249) and Postgraduate Research and Practice Innovation Program of Jiangsu Province (No. KYCX17_0288).

and arbitrary public announcement logic [5]. RML is obtained from the modal system $\boldsymbol{K}$ by adding a refinement operator $\exists_B$. The semantics of the refinement operator $\exists_B$ is given in terms of the notion of B-refinement. Such notion captures refinement preorders between Kripke models and the operator $\exists_B$ acts as a quantifier over the set of all B-refinements of a given pointed model. In addition to the refinement operator $\exists_B$, a kind of non-standard propositional quantifier so-called bisimulation quantifier $\widetilde{\exists}_X$ appeared earlier in the literature (see, e.g., [11, 12, 16]), where X is a proposition letter. A formula $\widetilde{\exists}_X \alpha$ is true in a pointed model M_s if there is a pointed model N_t satisfying α and an $(Atom - \{X\})$-bisimulation linking M_s and N_t. Here an $(Atom - \{X\})$-bisimulation is a relation between two models in which related states satisfy the same proposition letters in $Atom - \{X\}$ and have matching transition possibilities. It has been shown that refinement quantification is bisimulation quantification plus relativization [8].

In addition to refinement and bisimulation, there exist a number of different compatible relations between Kripke models. This paper will focus on the notion of covariant-contravariant refinement (CC-refinement, for short). The notion of CC-refinement (simulation), introduced in [13], is a behavioural preorder over labelled transitions systems (LTSs). The standard definition of simulation (see, e.g., [7, 22]) is related to reactive systems whose actions are passive, that is, their execution must be triggered by the environment. Any correct implementation of a reactive system is required to simulate the actions of its specification, in other words, the former should respond to the user's requests at least as well as the latter. Clearly, the standard notion of simulation provides a reasonable mathematical description of the refinement relations between reactive systems. However, for the systems with generative (or, active) actions, e.g., input/output (I/O) automata, such notion isn't adequate because, for these generative actions (e.g., output), it is the system that produces the output and the environment is obliged to accept the output produced by the system. One of the motivations behind introducing the notion of CC-refinement is to provide an adequate behavioural preorder for the systems referring to the generative actions. The definition of CC-refinement depends on partitioning all actions into three sorts: covariant, contravariant and bivariant actions. The covariant actions denote the passive actions of a system, whose execution is under the control of the environment. The transitions labelled with these actions in a given specification should be simulated by any correct implementation, which is exactly the requirement suggested by the standard notion of simulation. The contravariant actions represent the generative actions being under the control of a system itself. The transitions of these actions in an implementation must be simulated by its specification, in other words, the transitions induced by the generative actions must be allowed by its specification. The bivariant actions are treated as in the usual notion of bisimulation. In a word, by the standards suggested by the notion of CC-refinement, the covariant behaviour of any correct implementation should respect the liveness required by its specification, the contravariant behaviour must be prohibited from going beyond the safeness limit posed by its specification and their bivariant behaviour should match each other. More motivations behind this notion and work on it may be found in [3, 13–15].

This paper, following Laura Bozzelli et al's work, considers CC-refinement modal logic (CCRML, for short). Its language $\mathcal{L}_{CC}$ is obtained from the standard modal language $\mathcal{L}_K$ by adding CC-refinement operators (or, quantifiers) $\exists_{(A_1,A_2)}$, where A_1 (A_2) is the set of all covariant (contravariant, resp.) actions. Intuitively, the formula $\exists_{(A_1,A_2)}\varphi$ says that we can refine the current model so that φ is satisfied. This operator may be used to formalize some interesting problems in the field of formal method. For example, given a specification presented as an LTS M which refers to the set A_1 (A_2) of passive (generative, resp.) actions, the problem whether this specification has an implementation satisfying a given property φ may be boiled down to the model checking problem: whether M satisfies $\exists_{(A_1,A_2)}\varphi$. Thus, based on the CC-refinement quantifiers, the problem whether there exists a special implementation of a given specification may be solved using model checking methods.

CCRML focuses on reasoning and formalizing of the properties such as "there exists a CC-refinement model satisfying φ", which empowers the modal language with the quantifiers over the set of all the models related with CC-refinement to the current model. Fábregas et al. considered a logic for CC-refinement with very different motivation. Following Hennessy and Milner's well-known work [17], they introduced a modal logic to match the distinguishing ability of CC-refinement, and hence established its modal characterization [14].

In this paper we provide an axiom system for CCRML and explore its important properties: its soundness is established based on the $\oplus$-construction over models, and its completeness and decidability are obtained through transforming $\mathcal{L}_{CC}$-formulas into $\mathcal{L}_K$-formulas.

This paper is organized as follows. Section 2 recalls the notion of CC-refinement. Section 3 introduces CC-refinement modal logic. Section 4 presents a sound and complete axiomatization of CC-refinement modal logic. Finally we end the paper with a brief discussion in Section 5.

2 CC-refinement

Given a finite set A of actions and a set *Atom* of proposition letters, a model M is a tripe $\langle S^M, R^M, V^M\rangle$, where S^M is a non-empty set of states, R^M is an accessibility function from A to $2^{S^M \times S^M}$ assigning to each action a in A a binary relation $R_a^M \subseteq S^M \times S^M$, and $V^M : Atom \to 2^{S^M}$ is a valuation function. A pair (M, s) with $s \in S^M$ is said to be a pointed model, which is often written as M_s. For any binary relation R and s, $R(s) \triangleq \{t \mid sRt\}$, $\pi_1(R) \triangleq \{u \mid \exists v(uRv)\}$ and $\pi_2(R) \triangleq \{v \mid \exists u(uRv)\}$. For any sets B and C, if $B \subseteq C$ then $\mathfrak{i}_{B,C}$ is used to denote the graph of the inclusion function from B to C, that is $\mathfrak{i}_{B,C} = \{\langle a, a\rangle \mid a \in B\}$. We use $\circ$ to denote the composition operator of relations.

Given a model $M = \langle S, R, V\rangle$ and $R' : A \to 2^{S\times S}$, the model $M \mid R'$ is obtained from M by replacing R by R'. As usual, we write $M \uplus N$ for the disjoint union of two models M and N with $S^M \cap S^N = \emptyset$, which is defined as $S^{M\uplus N} \triangleq S^M \cup S^N$, $R_a^{M\uplus N} \triangleq R_a^M \cup R_a^N$ for each $a \in A$ and $V^{M\uplus N}(p) \triangleq V^M(p) \cup V^N(p)$ for each $p \in Atom$.

Definition 2.1 (CC-refinement [13])**.** *Let $A_1,\ A_2 \subseteq A$ such that $A_1 \cap A_2 = \emptyset$. Given two models $M = \langle S, R, V\rangle$ and $M' = \langle S', R', V'\rangle$, a binary relation $\mathcal{Z} \subseteq S \times S'$ is an (A_1, A_2)-refinement relation between M and M' if, for each pair $\langle s, s'\rangle$ in $\mathcal{Z}$,*

(atoms) *$s \in V(p)$ iff $s' \in V'(p)$ for each $p \in Atom$;*

(forth) *for each $a \in A - A_2$ and $t \in S$, $sR_a t$ implies $s'R'_a t'$ and $t\mathcal{Z}t'$ for some $t' \in S'$;*

(back) *for each $a \in A - A_1$ and $t' \in S', s'R'_a t'$ implies $sR_a t$ and $t\mathcal{Z}t'$ for some $t \in S$.*

*Here A_1 and A_2 are said to be **covariant** and **contravariant set** respectively. We say that $M'_{s'}$ (A_1, A_2)-refines M_s (or, M_s (A_1, A_2)-simulates $M'_{s'}$), in symbols $M_s \succeq_{(A_1,A_2)} M'_{s'}$, if there exists an ($A_1$, A_2)-refinement relation between M and M' linking s and s'. We also write $\mathcal{Z} : M_s \succeq_{(A_1,A_2)} M'_{s'}$ to indicate that $\mathcal{Z}$ is an (A_1, A_2)-refinement relation such that $s\mathcal{Z}s'$.*

The above notion generalizes the notions of ***bisimulation*** and ***refinement*** considered in [8]. Formally, a bisimulation relation is exactly an $(\emptyset, \emptyset)$-refinement, and a B-refinement relation an $(\emptyset, B)$-refinement. We write $\mathcal{Z} : M_s \underline{\leftrightarrow} M'_{s'}$ to indicate that $\mathcal{Z}$ is a bisimulation which witnesses that M_s is bisimilar to $M'_{s'}$.

Example 2.1. *Consider the models M and N depicted in Figure 1, where $A_1 = \{b, c\}$, $A_2 = \{a\}$, and $V^M(q) = V^N(q) = \emptyset$ for each $q \in Atom$. It is not difficult to see that the relation represented by the dash arrows is indeed an (A_1, A_2)-refinement relation between M_{s_1} and N_{t_1}, but M_{s_1} is not bisimilar to N_{t_1}.*

Roughly speaking, the notion of CC-refinement captures the idea that we should distinguish the different roles played by different kinds of actions when considering refinement relations between models. Analogous notions appear in the literature.

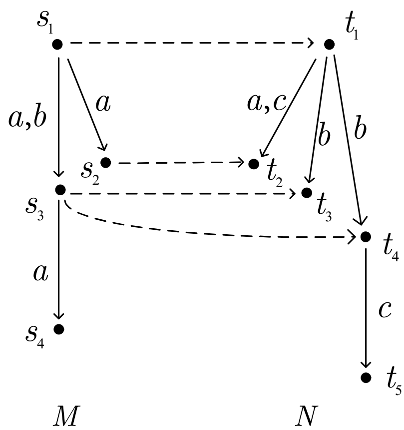


Figure 1: ($\{b, c\}$, $\{a\}$)-refinement

In the framework of modal transition systems (MTSs) [18, 19], a MTS contains two kinds of transitions: the must transitions and the may transitions, which denote the transitions required by a specification and the transitions allowed by a specification respectively. A modal refinement relation between MTSs is a binary relation satisfying the condition (**forth**) for the must transitions and (**back**) for the may transitions. In order to study topics in the field of supervisory control in the process-algebraic style, the notion of partial bisimulation is developed [4] in which the collection of actions is divided into two parts B and $Act - B$ and the transitions labelled with the actions in B ($Act - B$) are treated as in the classic bisimulation (simulation preorder, resp.). In the framework of interface automata [1, 2], the notion of XY-simulation is introduced in which the X-labelled (Y-labelled) transitions are required to fulfil the condition (**forth**) ((**back**), resp.). The relationships between the refinement over MTSs, and the CC-refinement and the partial bisimulation over LTSs are explored in the framework of Institutions [3].

Given $B \subseteq A$, a binary relation $\mathcal{Z}$ is said to be a ***B-restricted bisimulation***, in symbols $\mathcal{Z} : M \underline{\leftrightarrow}_B N$, if the bisimulation condition (**atoms**) holds, and (**forth**) and (**back**) are satisfied for each $b \in B$. It is trivial to see that $\mathcal{Z} : M \succeq_{(A_1, A_2)} M'$ iff for each pair $\langle s, s' \rangle$ in $\mathcal{Z}$,

(**atoms**) $s \in V(p)$ iff $s' \in V'(p)$ for each $p \in Atom$;

(**bis**) $\mathcal{Z} : M_s \underline{\leftrightarrow}_{A-(A_1 \cup A_2)} M'_{s'}$;

(**forth-$\mathbf{A_1}$**) for each $a \in A_1$ and $t \in S$, $sR_a t$ implies $s' R'_a t'$ and $t \mathcal{Z} t'$ for some $t' \in S'$;

(**back-$\mathbf{A_2}$**) for each $a \in A_2$ and $t' \in S'$, $s' R'_a t'$ implies $sR_a t$ and $t \mathcal{Z} t'$ for some $t \in S$.

Proposition 2.2. *The relation $\succeq_{(A_1,A_2)}$ is transitive.*

Proof. Assume $\mathcal{Z}_1 : M_s \succeq_{(A_1,A_2)} M'_t$ and $\mathcal{Z}_2 : M'_t \succeq_{(A_1,A_2)} M''_u$. It suffices to show that the composition relation $\mathcal{Z}_1 \circ \mathcal{Z}_2$ satisfies the conditions (**atoms**), (**forth**) and (**back**) in Definition 2.1, which is straightforward. □

Proposition 2.3. *The relation $\succeq_{(A_1,A_2)}$ is a pre-order satisfying the Church-Rosser property (i.e., if $M_s \succeq_{(A_1,A_2)} N_t$ and $M_s \succeq_{(A_1,A_2)} N'_{t'}$ then $N_t \succeq_{(A_1,A_2)} M'_{s'}$ and $N'_{t'} \succeq_{(A_1,A_2)} M'_{s'}$ for some $M'_{s'}$).*

Proof. It is straightforward to check that $\succeq_{(A_1,A_2)}$ is reflexive and transitive. In the following, we prove that $\succeq_{(A_1,A_2)}$ satisfies the Church-Rosser property. Assume that $\mathcal{Z}_1 : M_s \succeq_{(A_1,A_2)} N_t$ and $\mathcal{Z}_2 : M_s \succeq_{(A_1,A_2)} N'_{t'}$. We intend to provide a pointed model $M'_{s'}$ such that $N_t \succeq_{(A_1,A_2)} M'_{s'}$ and $N'_{t'} \succeq_{(A_1,A_2)} M'_{s'}$.

Without loss of generality, we assume that N and N' are disjoint. Before giving the desired pointed model, we explain the ideas behind the construction. Since the models N and N' are generated submodels of $N \uplus N'$, it holds trivially that $N_t \succeq_{(A_1,A_2)} (N \uplus N')_t$ and $N'_{t'} \succeq_{(A_1,A_2)} (N \uplus N')_{t'}$. Inspired by this observation, we intend to construct the desired model M' through modifying the model $N \uplus N'$. It is not difficult to see that the modification will occur on the accessibility function $R^{N \uplus N'}$. The modification will be guided by the following strategy: forcing M'_t to CC-refine $N'_{t'}$ and ensuring that all changes are safe (that is, M'_t still CC-refines N_t). We will illustrate the modification of the accessibility function based on the sorts of actions in turn.

For $a \in A_1$, since it is related to the condition (**forth**), we will modify the accessibility function $R^{N \uplus N'}$ so that it can afford the matching transitions for the a-labelled transitions in N'. For each transition $t' \xrightarrow{a} u'$ in $N'_{t'}$, we will add the transition depicted by the dash arrow in Figure 2. That is, we intend to define $R^{M'}$ so that the inequality $\mathcal{Z}_1^{-1} \circ \mathcal{Z}_2 \circ R_a^{N'} \subseteq R_a^{M'}$ holds.

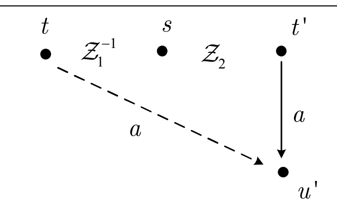


Figure 2: The construction of the model M'

For $a \in A_2$, in order to force t (in M') to CC-refine t' (in N'), for each a-labelled transition $t \xrightarrow{a} u$ in M', due to the condition (**back**), we need an a-labelled transition outgoing from t' in N' which matches the transition $t \stackrel{a}{\to} u$. Since N' is given and fixed, to meet this requirement, we are obliged to remove all the a-labelled transitions in $(N \uplus N')$ which come from N. After this modification, the resulted model M' will contain no a-labelled transition outgoing from the points in N, and then the condition (**back**) will hold trivially. Fortunately, it is easy to see that t (in M') still CC-refines t (in N).

Now we construct the desired pointed model M' formally. For each $a \in A$, set

$$R_a^{M'} \triangleq \begin{cases} R_a^N \cup R_a^{N'} \cup (\mathcal{Z}_1^{-1} \circ \mathcal{Z}_2 \circ R_a^{N'}) & \text{if } a \in A_1 \\ R_a^{N'} & \text{if } a \in A_2 \\ R_a^N \cup R_a^{N'} & \text{otherwise} \end{cases}$$

Put $M' \triangleq (N \uplus N') \mid R^{M'}$. It is straightforward to check $\mathfrak{i}_{S^N, S^{M'}} : N_t \succeq_{(A_1,A_2)} M'_t$ (Note that $R_a^{M'}(w) = \emptyset$ for each $w \in S^N$ and $a \in A_2$). Set

$$\mathcal{Z} \triangleq \mathfrak{i}_{S^{N'}, S^{M'}} \cup (\mathcal{Z}_2^{-1} \circ \mathcal{Z}_1).$$

Clearly, $t'\mathcal{Z}t$. In the following we verify $\mathcal{Z} : N'_{t'} \succeq_{(A_1,A_2)} M'_t$. Let $\langle u', u\rangle \in \mathcal{Z}$. If $\langle u', u\rangle \in \mathfrak{i}_{S^{N'}, S^{M'}}$, it is not difficult to see that the conditions (**atoms**), (**bis**), (**forth-$\mathbf{A_1}$**) and (**back-$\mathbf{A_2}$**) hold. Next we consider the case when $\langle u', u\rangle \in \mathcal{Z}_2^{-1} \circ \mathcal{Z}_1$.

(**atoms**) Trivially.

(**bis**) Since $\langle u', u\rangle \in \mathcal{Z}_2^{-1} \circ \mathcal{Z}_1$, there exists w such that $u'\mathcal{Z}_2^{-1}w$ and $w\mathcal{Z}_1 u$. Then it follows from $\mathcal{Z}_1 : M_w \succeq_{(A_1,A_2)} N_u$ and $\mathcal{Z}_2 : M_w \succeq_{(A_1,A_2)} N'_{u'}$ that $\mathcal{Z}_1 : M_w \underline{\leftrightarrow}_{A-(A_1\cup A_2)} N_u$ and $\mathcal{Z}_2 : M_w \underline{\leftrightarrow}_{A-(A_1\cup A_2)} N'_{u'}$, which implies

$$(\mathcal{Z}_2^{-1} \circ \mathcal{Z}_1) : N'_{u'} \underline{\leftrightarrow}_{A-(A_1\cup A_2)} N_u.$$

Further, due to the definition of M', it is not difficult to see that

$$(\mathcal{Z}_2^{-1} \circ \mathcal{Z}_1) : N'_{u'} \underline{\leftrightarrow}_{A-(A_1\cup A_2)} M'_u.$$

Thus $\mathcal{Z} : N'_{u'} \underline{\leftrightarrow}_{A-(A_1\cup A_2)} M'_u$ because of $\mathfrak{i}_{S^{N'}, S^{M'}} : N' \underline{\leftrightarrow}_{A-(A_1\cup A_2)} M'$.

(**forth-$\mathbf{A_1}$**) Let $a \in A_1$ and $u' R_a^{N'} v'$. By the definition of $R^{M'}$, we obtain $u R_a^{M'} v'$ due to $u' R_a^{N'} v'$ and $\langle u, u'\rangle \in \mathcal{Z}_1^{-1} \circ \mathcal{Z}_2$. Moreover, $v'\mathcal{Z}v'$ because of $\mathfrak{i}_{S^{N'}, S^{M'}} \subseteq \mathcal{Z}$.

(**back-$\mathbf{A_2}$**) Let $a \in A_2$. Since $u \notin S^{N'}$ and $R_a^{M'} = R_a^{N'}$, the state u has no a-labelled transition in M'. Thus (**back-$\mathbf{A_2}$**) holds immediately. □

In the following, we intend to show that, through taking compositions, any CC-refinement may be captured by the CC-refinements with singleton covariant and contravariant sets. To prove this result, Proposition 2.4 is needed to simplify its proof.

Proposition 2.4 (2) reveals that, given $\mathcal{Z} : M_s \succeq_{(A_1,A_2)} N_t$, we can construct two pointed models $M'_{s'}$ and $N'_{t'}$, which are bisimilar to M_s and N_t respectively, moreover, there is an injective partial function between $M'_{s'}$ and $N'_{t'}$ whose graph is an (A_1, A_2)-refinement relation. In the following, we will explain the construction of $M'_{s'}$, and the pointed model $N'_{t'}$ is constructed similarly. At first glance, in order to get an injective (A_1, A_2)-refinement function, the model $M'_{s'}$ $(N'_{t'})$ can be obtained from M_s (N_t, resp.) by adding enough copies of the states involved in $\mathcal{Z}$. That is, we intend to replace each $u \in \pi_1(\mathcal{Z})$ by all the pairs of the form $\langle u, v\rangle$ in $\mathcal{Z}$. Moreover the transitions from these new states $\langle u, v\rangle$ in M' are prescribed according to the ones related to u in M. The detailed construction will be given in the proof of Proposition 2.4, here we only explain an interesting part of the construction. In our construction, the transitions between two new states $\langle u, v\rangle$ and $\langle u', v'\rangle$ are captured by the rule

$$\langle u, v\rangle R_a^{M'} \langle u', v'\rangle \quad \text{iff} \quad uR_a^M u' \text{ and } vR_a^N v'. \qquad (*)$$

Unfortunately, although such simple construction brings us the desired result that there is an (A_1, A_2)-refinement relation $\mathcal{Z}'$ which is an injective partial function such that $\mathcal{Z}' : M'_{\langle s,t\rangle} \succeq_{(A_1,A_2)} N'_{\langle t,s\rangle}$, it does not always hold that $M_s \underline{\leftrightarrow} M'_{\langle s,t\rangle}$ and $N_t \underline{\leftrightarrow} N'_{\langle t,s\rangle}$.

Example 2.2. *Consider the models M and N depicted in Figure 3. Here $b \in A_1$, $a \in A_2$ and the dash arrows represent an (A_1, A_2)-refinement relation. According to the construction mentioned above, the models M' and N' may be depicted as in Figure 3. Clearly, the isomorphism between M' and N' is indeed an injective CC-refinement function between them, but $M_s \not\underline{\leftrightarrow} M'_{\langle s,t\rangle}$.*

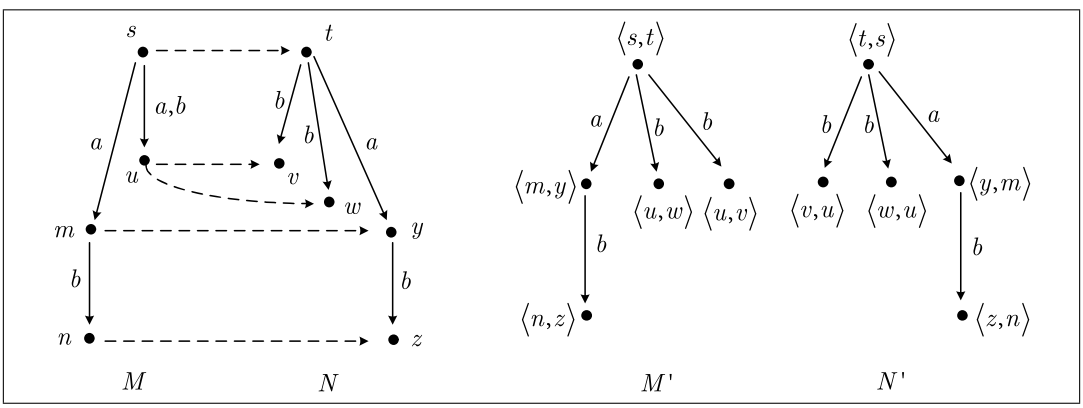


Figure 3: One counterexample for the construction $(*)$

In the above example, due to the rule $(*)$, $t \stackrel{a}{\nrightarrow} v$ and $t \stackrel{a}{\nrightarrow} w$, we have that $\langle s,t\rangle \stackrel{a}{\nrightarrow} \langle u,v\rangle$ and $\langle s,t\rangle \stackrel{a}{\nrightarrow} \langle u,w\rangle$. Thus the a-labelled transition $s \stackrel{a}{\to} u$ in the model M does not reflect in M', which does not affect that $N'_{\langle t,s\rangle}$ CC-refines $M'_{\langle s,t\rangle}$, but causes that $M_s \not\leftrightarrow M'_{\langle s,t\rangle}$. In fact, given $\mathcal{Z} : M \succeq_{(A_1,A_2)} N$ and a transition $u \stackrel{a}{\to} v$ $(a \in A_2)$ in M, if the following conditions hold

(1) $u \in \pi_1(\mathcal{Z})$ and $v \in \pi_1(\mathcal{Z})$, and

(2) $\exists w(u\mathcal{Z}w$ and $\forall w'(v\mathcal{Z}w' \Rightarrow w \stackrel{a}{\nrightarrow} w'))$,

then the transition $u \stackrel{a}{\to} v$ does not reflect in the model M' obtained by the construction mentioned in the preceding paragraph. To remedy such flaw, we will preserve these states v and the a-labelled transitions entering v. For each $a \in A_2$, in the proof of Proposition 2.4, the set of all these states v is denoted by $S^{M,\mathcal{Z}}_{a-}$, which is defined as

$$S^{M,\mathcal{Z}}_{a-} \triangleq \pi_1(\mathcal{Z}) \cap \pi_2(R^M_a \cap (\mathcal{Z} \circ \overline{R^N_a \circ \mathcal{Z}^{-1}}\,)).$$

Here $\overline{R^N_a \circ \mathcal{Z}^{-1}}$ is the complementation of $R^N_a \circ \mathcal{Z}^{-1}$, namely, $\overline{R^N_a \circ \mathcal{Z}^{-1}} = S^N \times S^M - R^N_a \circ \mathcal{Z}^{-1}$. Obviously, the definition of $S^{M,\mathcal{Z}}_{a-}$ is induced by the conditions (1) and (2). Summarizing, $S^{M'}$ is obtained from S^M by adding all the pairs in $\mathcal{Z}$ and, for the states in $\pi_1(\mathcal{Z})$, only keeping the ones in $\bigcup_{a\in A_2} S^{M,\mathcal{Z}}_{a-}$. Applying this remedied construction to the model M in Example 2.2, we get the one depicted in Figure 4.

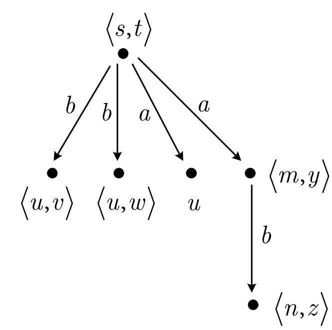


Figure 4: The model obtained by applying the remedied construction

Similarly, the construction of N' should also be remedied. In this case, we will concern ourselves with the A_1-sort actions instead of the A_2-sort actions. This is the motivation behind introducing the set $S^{N,\mathcal{Z}}_{a+}$ in the proof of Proposition 2.4.

Proposition 2.4. (1) $M_{s_1} \underline{\leftrightarrow} M'_{s_2} \succeq_{(A_1,A_2)} N'_{t_2} \underline{\leftrightarrow} N_{t_1}$ *implies* $M_{s_1} \succeq_{(A_1,A_2)} N_{t_1}$.
(2) *If* $M_s \succeq_{(A_1,A_2)} N_t$ *then there exist* $M'_{s'}$, $N'_{t'}$ *and* $\mathcal{Z}$ *such that* $M_s \underline{\leftrightarrow} M'_{s'}$, $N_t \underline{\leftrightarrow} N'_{t'}$,

and $\mathcal{Z} : M'_{s'} \succeq_{(A_1,A_2)} N'_{t'}$ *that is an injective partial function from* $S^{M'}$ *to* $S^{N'}$, *namely,* $\mathcal{Z}$ *satisfies*
(2.1) $\forall w \in S^{M'}\ \forall v_1, v_2 \in S^{N'}(w\mathcal{Z}v_1$ *and* $w\mathcal{Z}v_2 \Rightarrow v_1 = v_2)$, *and*
(2.2) $\forall v \in S^{N'}\ \forall w_1, w_2 \in S^{M'}(w_1\mathcal{Z}v$ *and* $w_2\mathcal{Z}v \Rightarrow w_1 = w_2)$.

Proof. (1) Straightforward.

(2) Assume $\mathcal{Z} : M_s \succeq_{(A_1,A_2)} N_t$. For each $a \in A_2$, set

$$S^{M,\mathcal{Z}}_{a-} \triangleq \pi_1(\mathcal{Z}) \cap \pi_2(R^M_a \cap (\mathcal{Z} \circ \overline{R^N_a \circ \mathcal{Z}^{-1}}\,)).$$

The model M' is defined as follows.

(M'_1) $S^{M'} \triangleq (S^M - \pi_1(\mathcal{Z})) \cup \mathcal{Z} \cup \bigcup_{a\in A_2} S^{M,\mathcal{Z}}_{a-}$ (Here we assume $S^M \cap \mathcal{Z} = \emptyset$).

(M'_2) For each $a \in A$, $R^{M'}_a \subseteq S^{M'} \times S^{M'}$ is obtained from R^M_a by preserving the transitions between the states in $S^M \cap S^{M'}$, and prescribing the behaviour of a new state $\langle u, v\rangle$ according to the behaviour of u in M and the rule $(*)$. Formally,

$$R^{M'}_a \triangleq \begin{cases} R^{\blacktriangle}_a \cup \{\langle\langle u, v\rangle, w\rangle \mid u\mathcal{Z}v, w \in S^M \cap S^{M'} \text{ and } uR^M_a w\} & \text{if } a \in A_2 \\ R^{\blacktriangle}_a & \text{otherwise} \end{cases}$$

Here

$$R^{\blacktriangle}_a \triangleq (R^M_a \cap (S^{M'})^2) \cup \{\langle w, \langle u, v\rangle\rangle \mid u\mathcal{Z}v \text{ and } wR^M_a u\} \\ \cup \{\langle\langle u, v\rangle, \langle u', v'\rangle\rangle \mid uR^M_a u', vR^N_a v', u\mathcal{Z}v \text{ and } u'\mathcal{Z}v'\}.$$

Note that there is no transition like $\langle u, v\rangle \xrightarrow{a} w$ with $a \notin A_2$ and $w \in S^M$ even if $uR^M_a w$. Since $a \notin A_2$ and $u\mathcal{Z}v$, by the condition (**forth**), it follows from $uR^M_a w$ that $vR^N_a w'$ and $w\mathcal{Z}w'$ for some $w' \in S^N$. Thus, the transition $\langle u, v\rangle \xrightarrow{a} \langle w, w'\rangle$ exists in M' due to the definition of $R^{\blacktriangle}_a$, and hence the transition $\langle u, v\rangle \xrightarrow{a} w$ is redundant.

(M'_3) For each $p \in Atom$,

$$V^{M'}(p) \triangleq (V^M(p) \cap S^{M'}) \cup \{\langle u, v\rangle \mid u\mathcal{Z}v \text{ and } u \in V^M(p)\}.$$

The model N' is constructed analogously.

(N'_1) $S^{N'} \triangleq (S^N - \pi_2(\mathcal{Z})) \cup \mathcal{Z}^{-1} \cup \bigcup_{a\in A_1} S^{N,\mathcal{Z}}_{a+}$.

(N'_2) For each $a \in A$, $R^{N'}_a \subseteq S^{N'} \times S^{N'}$ is given below

$$R^{N'}_a \triangleq \begin{cases} R^{\blacktriangle'}_a \cup \{\langle\langle v, u\rangle, w\rangle \mid u\mathcal{Z}v, w \in S^N \cap S^{N'} \text{ and } vR^N_a w\} & \text{if } a \in A_1 \\ R^{\blacktriangle'}_a & \text{otherwise} \end{cases}$$

(N_3') For each $p \in Atom$,

$$V^{N'}(p) \triangleq (V^N(p) \cap S^{N'}) \cup \{\langle v, u\rangle \mid u\mathcal{Z}v \text{ and } v \in V^N(p)\}.$$

Here

$$S_{a+}^{N,\mathcal{Z}} \triangleq \pi_2(\mathcal{Z}) \cap \pi_2(R_a^N \cap (\mathcal{Z}^{-1} \circ \overline{R_a^M \circ \mathcal{Z}}\,))$$

and

$$R_a^{\blacktriangle\prime} \triangleq (R_a^N \cap (S^{N'})^2) \cup \{\langle w, \langle v, u\rangle\rangle \mid u\mathcal{Z}v \text{ and } wR_a^N v\} \\ \cup \{\langle\langle v, u\rangle, \langle v', u'\rangle\rangle \mid uR_a^M u', vR_a^N v', u\mathcal{Z}v \text{ and } u'\mathcal{Z}v'\}.$$

Then it is not difficult to check that

$$\mathcal{Z}_1 : M_s \underline{\leftrightarrow} M'_{\langle s,t\rangle}, \;\; \mathcal{Z}_2 : N_t \underline{\leftrightarrow} N'_{\langle t,s\rangle} \;\text{ and }\; \mathcal{Z}_3 : M'_{\langle s,t\rangle} \succeq_{(A_1,A_2)} N'_{\langle t,s\rangle}$$

where

$\mathcal{Z}_1 \triangleq \{\langle u, u\rangle \mid u \in S^M \cap S^{M'}\} \cup \{\langle u, \langle u, v\rangle\rangle \mid u\mathcal{Z}v\}$
$\mathcal{Z}_2 \triangleq \{\langle v, v\rangle \mid v \in S^N \cap S^{N'}\} \cup \{\langle v, \langle v, u\rangle\rangle \mid u\mathcal{Z}v\}$
$\mathcal{Z}_3 \triangleq \{\langle\langle u, v\rangle, \langle v, u\rangle\rangle \mid u\mathcal{Z}v\}$.

Moreover, $\mathcal{Z}_3$ satisfies the conditions (2.1) and (2.2), as desired. □

Proposition 2.5. *Let $A_1, A_2 \subseteq A$ with $A_1 \cap A_2 = \emptyset$. Then, for each A_1', A_1'', A_2' and A_2'' such that $A_1' \cup A_1'' = A_1$ and $A_2' \cup A_2'' = A_2$, it holds that*

$$\succeq_{(A_1',A_2')} \circ \succeq_{(A_1'',A_2'')} = \succeq_{(A_1,A_2)} .$$

Proof. ($\subseteq$) Let $M_s \succeq_{(A_1',A_2')} \circ \succeq_{(A_1'',A_2'')} N_t$. Then we have that $\mathcal{Z}_1 : M_s \succeq_{(A_1',A_2')} N_v'$ and $\mathcal{Z}_2 : N_v' \succeq_{(A_1'',A_2'')} N_t$ for some N_v', $\mathcal{Z}_1$ and $\mathcal{Z}_2$. It is routine to check that $\mathcal{Z}_1 \circ \mathcal{Z}_2 : M_s \succeq_{(A_1,A_2)} N_t$.

($\supseteq$) Assume that M and N are disjoint and $\mathcal{Z} : M_s \succeq_{(A_1,A_2)} N_t$. By Proposition 2.4, we may suppose that $\mathcal{Z}$ is an injective partial function from S^M to S^N. To complete the proof, we intend to construct a pointed model N_v' such that $M_s \succeq_{(A_1',A_2')} N_v' \succeq_{(A_1'',A_2'')} N_t$. Put

$$N' \triangleq (M \uplus N) \mid R^{N'}.$$

Here for each $a \in A$,

$$R_a^{N'} \triangleq \begin{cases} R_a^N \cup (\mathcal{Z} \circ R_a^N) & \text{if } a \in A_2' - A_2'' \\ R_a^M \cup (\mathcal{Z}^{-1} \circ R_a^M) & \text{if } a \in A_1'' - A_1' \\ R_a^M \cup R_a^N \cup (\mathcal{Z}^{-1} \circ R_a^M) & \text{if } a \in A_2'' - A_2' \\ R_a^M \cup R_a^N \cup (\mathcal{Z} \circ R_a^N) & \text{if } a \in A_1' - A_1'' \\ R_a^M \cup R_a^N & \text{otherwise} \end{cases}$$

Next we show that the pointed model N'_t is the desired one. The proof will be divided into two steps.

Claim 1 $\mathbf{i}_{S^M,S^{N'}} \cup \mathcal{Z} : M_s \succeq_{(A'_1,A'_2)} N'_t$.

Let $\langle w, w' \rangle \in \mathbf{i}_{S^M,S^{N'}} \cup \mathcal{Z}$. Then it is obvious that w and w' satisfy the same proposition letters and hence the pair $\langle w, w' \rangle$ satisfies the condition (**atoms**). In the following, we deal with the other conditions (**forth**) and (**back**) by considering two cases.

Case 1 $\langle w, w' \rangle \in \mathbf{i}_{S^M,S^{N'}}$

Then $w = w'$. By the construction of N', it is easy to see that the pair $\langle w, w' \rangle$ satisfies (**forth**). Next we verify that such pair also satisfies (**back**). Assume that $w' R_a^{N'} v'$ with $a \in A - A'_1$. Since $S^M \cap S^N = \emptyset$ and $w \in S^M$, by the definition of $R^{N'}$, it is not difficult to see that either $w' R_a^M v'$ or $w'(\mathcal{Z} \circ R_a^N)v'$. For the former, (**back**) holds trivially. For the latter, we get $a \in A'_2 - A''_2$ and there exists $u' \in S^N$ such that $w' \mathcal{Z} u'$ and $u' R_a^N v'$. Hence, due to $a \in A'_2 \subseteq A_2$ and $\mathcal{Z} : M_{w'} \succeq_{(A_1,A_2)} N_{u'}$, we obtain $w' R_a^M v$ (i.e., $w R_a^M v$) and $v \mathcal{Z} v'$ for some $v \in S^M$, as desired.

Case 2 $\langle w, w' \rangle \in \mathcal{Z}$

Then $w' \in S^N$. We deal with (**forth**) and (**back**) in turn.

(**forth**) Let $a \in A - A'_2$ and $w R_a^M v$. Hence $w' \mathcal{Z}^{-1} \circ R_a^M v$. If $a \in (A''_1 - A'_1) \cup (A''_2 - A'_2)$ then, by the construction of $R_a^{N'}$, we get $w' R_a^{N'} v$, moreover, $\langle v, v \rangle \in \mathbf{i}_{S^M,S^{N'}} \cup \mathcal{Z}$ as desired. If $a \notin (A''_1 - A'_1) \cup (A''_2 - A'_2)$ then $a \notin A_2$ due to $a \notin A'_2$ and $A_2 = A'_2 \cup A''_2$. Further, it follows from $\mathcal{Z} : M_w \succeq_{(A_1,A_2)} N_{w'}$ and $w R_a^M v$ that $w' R_a^N v'$ and $v \mathcal{Z} v'$ for some $v' \in S^N$. Since $a \notin A'_2$ and $a \notin A''_1 - A'_1$, by the definition of $R_a^{N'}$, we have $w' R_a^{N'} v'$. Clearly, $\langle v, v' \rangle \in \mathbf{i}_{S^M,S^{N'}} \cup \mathcal{Z}$.

(**back**) Let $a \in A - A'_1$ and $w' R_a^{N'} v'$. Since $a \notin A'_1$ and $w' \in S^N$, by the definition of $R^{N'}$, we get either $w' R_a^N v'$ or $w'(\mathcal{Z}^{-1} \circ R_a^M)v'$. If $w' R_a^N v'$ then $a \notin A''_1$ and it immediately follows from $\mathcal{Z} : M_w \succeq_{(A_1,A_2)} N_{w'}$ and $a \notin A_1 (= A'_1 \cup A''_1)$ that $w R_a^M v$ and $v \mathcal{Z} v'$ for some $v \in S^M$ as desired. Next we consider another case where $w'(\mathcal{Z}^{-1} \circ R_a^M)v'$. In such case, $w_0 \mathcal{Z} w'$ and $w_0 R_a^M v'$ for some $w_0 \in S^M$. Since $\mathcal{Z}$ is an injective partial function from S^M to S^N, $w_0 = w$ immediately follows from $w_0 \mathcal{Z} w'$ and $w \mathcal{Z} w'$. Then, $w R_a^M v'$ and $\langle v', v' \rangle \in \mathbf{i}_{S^M,S^{N'}} \cup \mathcal{Z}$ as desired.

Claim 2 $\mathfrak{i}^{-1}_{S^N,S^{N'}} \cup \mathcal{Z} : N'_t \succeq_{(A''_1,A''_2)} N_t$.

Let $\langle w', w\rangle \in \mathfrak{i}^{-1}_{S^N,S^{N'}} \cup \mathcal{Z}$. Similarly, we verify that this pair satisfies the conditions (**forth**) and (**back**) by considering two cases.

Case 1 $\langle w', w\rangle \in \mathfrak{i}^{-1}_{S^N,S^{N'}}$

Hence $w' = w$. Clearly, for each $a \in A - A''_1$ and $v \in S^N$, $wR^N_a v$ implies $w'(= w)R^{N'}_a v$. Thus, the pair $\langle w', w\rangle$ satisfies the condition (**back**). Next we check that such pair also satisfies (**forth**). Let $a \in A - A''_2$ and $w'R^{N'}_a v'$. Then, by the definition of $R^{N'}$, we have either $w'R^N_a v'$ or $w'(\mathcal{Z}^{-1} \circ R^M_a)v'$. For the former, the verifying is straightforward. For the latter, we get $a \in A''_1 - A'_1$, $u'\mathcal{Z}w'$ and $u'R^M_a v'$ for some $u' \in S^M$. Thus, it follows from $a \in A''_1 \subseteq A_1$ and $\mathcal{Z} : M_{u'} \succeq_{(A_1,A_2)} N_{w'}$ that $w'R^N_a v$ (that is $wR^N_a v$) and $v'\mathcal{Z}v$ for some $v \in S^N$, as desired.

Case 2 $\langle w', w\rangle \in \mathcal{Z}$

(**forth**) Let $a \in A - A''_2$ and $w'R^{N'}_a v'$. Since $w' \in S^M$, by the definition of $R^{N'}$, it is easy to see either $w'R^M_a v'$ or $w'(\mathcal{Z} \circ R^N_a)v'$. If $w'R^M_a v'$ then $a \notin A'_2$ and it immediately follows from $\mathcal{Z} : M_{w'} \succeq_{(A_1,A_2)} N_w$ and $a \notin A_2$ $(= A'_2 \cup A''_2)$ that $wR^N_a v$ and $v'\mathcal{Z}v$ for some $v \in S^N$ as desired. If $w'(\mathcal{Z} \circ R^N_a)v'$ then $w'\mathcal{Z}w'_0$ and $w'_0 R^N_a v'$ for some $w'_0 \in S^N$. Then, applying the assumption that $\mathcal{Z}$ is an injective partial function again, we have $w'_0 = w$. Hence, $wR^N_a v'$ and $\langle v', v'\rangle \in \mathfrak{i}^{-1}_{S^N,S^{N'}} \cup \mathcal{Z}$.

(**back**) Let $a \in A - A''_1$ and $wR^N_a v$. Then it follows from $w'\mathcal{Z}w$ and $wR^N_a v$ that $w'\mathcal{Z} \circ R^N_a v$. If $a \in (A'_1 - A''_1) \cup (A'_2 - A''_2)$ then $w'R^{N'}v$ by the definition of $R^{N'}$, moreover, $\langle v, v\rangle \in \mathfrak{i}^{-1}_{S^N,S^{N'}} \cup \mathcal{Z}$ as desired. If $a \notin (A'_1 - A''_1) \cup (A'_2 - A''_2)$ then $a \notin A'_1$ and $a \notin A''_1$ due to $a \in A - A''_1$. Hence $a \notin A_1$. Then it follows from $\mathcal{Z} : M_{w'} \succeq_{(A_1,A_2)} N_w$ and $wR^N_a v$ that $w'R^M_a v'$ and $v'\mathcal{Z}v$ for some $v' \in S^M \subseteq S^{N'}$. Clearly, we also have $w'R^{N'}_a v'$ because of $w'R^M_a v'$ and $a \notin A'_2 - A''_2$. □

Corollary 2.6. *If $A_1 \neq \emptyset$ or $A_2 \neq \emptyset$, then*

$$\succeq_{\theta_1} \circ \succeq_{\theta_2} \circ \cdots \circ \succeq_{\theta_n} = \succeq_{(A_1,A_2)}$$

where, all the pairs in $A_1 \times A_2$ are arranged in a permutation $\{\theta_i\}_{1\leq i\leq n}$ with $n = |A_1 \times A_2|$ if $A_1 \neq \emptyset$ and $A_2 \neq \emptyset$, otherwise, $\{\theta_i :\ 1 \leq i \leq n\} = A_k$ with $n = |A_k|$ if $A_k \neq \emptyset$ $(k = 1$ or $2)$.

Laura Bozzelli et al. have obtained the same conclusion for B-refinement [8], which corresponds to the case where $A_1 = \emptyset$ and $A_2 = B$.

3 CC-refinement modal logic

This section presents CC-refinement modal logic (CCRML), which is obtained from the modal system $\boldsymbol{K}$ by adding CC-refinement quantifiers.

Definition 3.1 (Language $\mathcal{L}_{CC}$). *Let A be a finite set of actions and Atom a set of proposition letters. The language $\mathcal{L}_{CC}$ of CC-refinement modal logic is generated by the BNF grammar below, where $\emptyset \neq A_1,\ A_2 \subseteq A$ with $A_1 \cap A_2 = \emptyset$, $p \in Atom$ and $a \in A$:*

$$\varphi ::= p \mid \neg\varphi \mid (\varphi_1 \wedge \varphi_2) \mid \Box_a\varphi \mid \exists_{(A_1,A_2)}\varphi$$

The modal operator $\Diamond_a$ and propositional connectives $\rightarrow$, $\vee$, $\leftrightarrow$, $\top$ and $\bot$ are defined in the standard manner. Moreover, we also write $\forall_{(A_1,A_2)}\varphi$ for $\neg\exists_{(A_1,A_2)}\neg\varphi$.

For the sake of simplicity, this paper assumes that $A_1 \neq \emptyset$ and $A_2 \neq \emptyset$. Section 5 will discuss how to dispense with this assumption. If both A_1 and A_2 are singletons, say $A_1 = \{a_1\}$ and $A_2 = \{a_2\}$, we write $\exists_{(a_1,a_2)}\varphi$ (or $\forall_{(a_1,a_2)}\varphi$) instead of $\exists_{(A_1,A_2)}\varphi$ (resp., $\forall_{(A_1,A_2)}\varphi$). Moreover, we shall write $\bigwedge \Diamond_a\exists_{(a_1,a_2)}\Phi$ (or $\Box_a \bigvee \exists_{(a_1,a_2)}\Phi$) for $\bigwedge_{\varphi\in\Phi} \Diamond_a\exists_{(a_1,a_2)}\varphi$ (resp., $\Box_a \bigvee_{\varphi\in\Phi} \exists_{(a_1,a_2)}\varphi$) to ease the notation.

In this paper, the cover operator ∇_a is also adopted. As usual, for each $a \in A$, $\nabla_a\Phi$ is defined as $\Box_a \bigvee_{\varphi\in\Phi} \varphi \wedge \bigwedge_{\varphi\in\Phi} \Diamond_a\varphi$, where Φ is a finite set of formulas. It is well known that the operators $\Box_a$ and $\Diamond_a$ may be defined in terms of ∇_a. Formally, $\Box_a\varphi$ and $\Diamond_a\varphi$ are captured by $\nabla_a\emptyset \vee \nabla_a\{\varphi\}$ and $\nabla_a\{\varphi, \top\}$ respectively. More information about the cover operator may be found in [6, 11].

Given a model M, the notion of a formula $\varphi \in \mathcal{L}_{CC}$ being satisfied in M at a state s is defined inductively as follows:

$M_s \models p$	iff	$s \in V^M(p)$, where $p \in Atom$
$M_s \models \neg\varphi$	iff	$M_s \not\models \varphi$
$M_s \models \varphi_1 \wedge \varphi_2$	iff	$M_s \models \varphi_1$ and $M_s \models \varphi_2$
$M_s \models \Box_a\varphi$	iff	for all $t \in R_a^M(s), M_t \models \varphi$
$M_s \models \exists_{(A_1,A_2)}\varphi$	iff	$M_s \succeq_{(A_1,A_2)} N_t$ and $N_t \models \varphi$ for some N_t

As usual, for each $\varphi \in \mathcal{L}_{CC}$, φ is valid, in symbols $\models \varphi$, if $M_s \models \varphi$ for every pointed model M_s. It is easy to see that $\mathcal{L}_{CC}$-satisfiability is invariant under bisimulations, which is a known result:

Proposition 3.2. *If $M_s \underline{\leftrightarrow} N_t$ then*

$$M_s \models \psi \text{ iff } N_t \models \psi \quad \text{for all formulas } \psi \in \mathcal{L}_{CC}.$$

Proof. Proceeding by induction on ψ. $\square$

It is not difficult to see that

$$\begin{array}{lll} M_s \models \nabla_b \exists_{(a_1,a_2)}\Phi & \text{iff} & M_s \models \Box_b \bigvee \exists_{(a_1,a_2)}\Phi \text{ and } M_s \models \bigwedge \Diamond_b \exists_{(a_1,a_2)}\Phi \\ & \text{iff} & \forall t(sR_b^M t \Rightarrow \exists \varphi \in \Phi(M_t \models \exists_{(a_1,a_2)}\varphi)) \ \text{ and} \\ & & \forall \varphi(\varphi \in \Phi \Rightarrow \exists t(sR_b^M t \ \text{ and } \ M_t \models \exists_{(a_1,a_2)}\varphi)). \end{array}$$

Thus, to assert that a pointed model M_s satisfies a given formula $\nabla_b \exists_{(a_1,a_2)}\Phi$, we are required to provide a witness Ω which is a set of pointed models and satisfies the following conditions:

(1) for each t such that $sR_b^M t$, $N_u \models \varphi$ and $M_t \succeq_{(A_1,A_2)} N_u$ for some $\varphi \in \Phi$ and $N_u \in \Omega$;

(2) for each $\varphi \in \Phi$, $N_u \models \varphi$ and $M_t \succeq_{(A_1,A_2)} N_u$ for some $N_u \in \Omega$ and t with $sR_b^M t$.

In the following, two auxiliary notions are introduced, which describe a witness Ω for a given cover formula and hence capture its semantical characterization.

Definition 3.3 (b-support and b-ccrefine)**.** *Given $\Phi \subseteq_f \mathcal{L}_{CC}$ (i.e., Φ is a finite subset of $\mathcal{L}_{CC}$), a_1, a_2 and $b \in A$ and a pointed model M_s, for every set (of pointed models) Ω such that $\forall N_t \in \Omega\ \exists u \in R_b^M(s)\ (M_u \succeq_{(a_1,a_2)} N_t \models \bigvee \Phi)$, we say*
(**1**) *Ω b-**supports** Φ w.r.t. a_1, a_2 and M_s if $\forall \varphi \in \Phi\ \exists N_t \in \Omega\ (N_t \models \varphi)$,*
(**2**) *Ω b-**ccrefines** M_s w.r.t. a_1, a_2 and Φ if $\forall u \in R_b^M(s) \exists N_t \in \Omega(M_u \succeq_{(a_1,a_2)} N_t)$.*

Here are some elementary properties of these concepts, which are trivial but will be needed in the next section.

Lemma 3.4.

(1) *$M_s \models \Box_b \bigvee \exists_{(a_1,a_2)}\Phi$ iff there exists a set Ω which b-**ccrefines** M_s w.r.t. a_1, a_2 and Φ.*

(2) *$M_s \models \bigwedge \Diamond_b \exists_{(a_1,a_2)}\Phi$ iff there exists a set Ω which b-**supports** Φ w.r.t. a_1, a_2 and M_s.*

(3) *$M_s \models \nabla_b \exists_{(a_1,a_2)}\Phi$ iff there exists a set Ω which b-**ccrefines** M_s w.r.t. a_1, a_2 and Φ, and b-**supports** Φ w.r.t. a_1, a_2 and M_s.*

Proof. Straightforward. □

Proposition 3.5. *For all A_1', A_2', A_1'', $A_2'' \subseteq A$ such that $(A_1' \cup A_1'') \cap (A_2' \cup A_2'') = \emptyset$,*

$$\models \exists_{(A_1',A_2')}\exists_{(A_1'',A_2'')}\psi \leftrightarrow \exists_{(A_1'',A_2'')}\exists_{(A_1',A_2')}\psi.$$

Proof. Follow from Proposition 2.5, immediately. □

4 Axiom system

A sound and complete axiom system for the logic CCRML will be presented in this section. Similar to the axiom system **RML** [8], the uniform substitution rule is not sound in CCRML. For example, although $p \to \forall_{(a_1,a_2)} p$ is valid for every proposition letter p, the formula $\Diamond_{a_2}\top \to \forall_{(a_1,a_2)}\Diamond_{a_2}\top$ is not valid. Therefore CCRML is not a normal modal logic.

We use $\mathcal{L}_p$ to denote the set of all propositional formulas in $\mathcal{L}_{CC}$. Clearly, the fragment of $\mathcal{L}_{CC}$ consisting of all formulas containing no CC-refinement quantifier is indeed the multi-agent modal language $\mathcal{L}_K$, which may be axiomatized by the system $\boldsymbol{K}$ [7].

Axiom schemes

Here $a_1, a_2, a, b \in A$, $A_1, A_2, B \subseteq A$, $A_1 \cap A_2 = \emptyset$, $p \in Atom$, $\Gamma \subseteq_f \mathcal{L}_K$, and $\Phi, \Phi_b \subseteq_f \mathcal{L}_{CC}$.

Prop	*All propositional tautologies*	
K	$\Box_a(\varphi \to \psi) \to (\Box_a\varphi \to \Box_a\psi)$	
CCR	$\forall_{(a_1,a_2)}(\varphi \to \psi) \to (\forall_{(a_1,a_2)}\varphi \to \forall_{(a_1,a_2)}\psi)$	
CCRp1	$\forall_{(a_1,a_2)} p \leftrightarrow p$	
CCRp2	$\forall_{(a_1,a_2)} \neg p \leftrightarrow \neg p$	
CCRD	$\exists_{(A_1,A_2)}\varphi \leftrightarrow (\exists_{\theta_1}\cdots\exists_{\theta_{\lvert A_1\times A_2\rvert}})\varphi$	where all the pairs in $A_1 \times A_2$ are arranged in a permutation $\{\theta_i\}_{1\le i\le \lvert A_1\times A_2\rvert}$
CCRKco1	$\exists_{(a_1,a_2)}\nabla_{a_1}\Gamma \leftrightarrow \bot$	if $\vdash_{\boldsymbol{K}} \alpha \leftrightarrow \bot$ for some $\alpha \in \Gamma$
CCRKco2	$\exists_{(a_1,a_2)}\nabla_{a_1}\Gamma \leftrightarrow \Box_{a_1}\bigvee\exists_{(a_1,a_2)}\Gamma$	if $\nvdash_{\boldsymbol{K}} \alpha \leftrightarrow \bot$ for all $\alpha \in \Gamma$
CCRKcontra	$\exists_{(a_1,a_2)}\nabla_{a_2}\Phi \leftrightarrow \bigwedge\Diamond_{a_2}\exists_{(a_1,a_2)}\Phi$	
CCRKbis	$\exists_{(a_1,a_2)}\nabla_b\Phi \leftrightarrow \nabla_b\exists_{(a_1,a_2)}\Phi$	where $b \neq a_1, a_2$
CCRKconj	$\exists_{(a_1,a_2)}\bigwedge_{b\in B}\nabla_b\Phi_b \leftrightarrow \bigwedge_{b\in B}\exists_{(a_1,a_2)}\nabla_b\Phi_b$	

Rules

MP $\dfrac{\varphi \to \psi, \varphi}{\psi}$ **NK** $\dfrac{\varphi}{\Box_a\varphi}$ **NCCR** $\dfrac{\varphi}{\forall_{(a_1,a_2)}\varphi}$

Table 1: Axiom system of CCRML

The axiom schemes and rules for CCRML are given in Table 1. **CCR** is the $\forall_{(a_1,a_2)}$-over-implies distribution which allows us to transform $\forall_{(a_1,a_2)}(\varphi \to \psi)$ into

$\forall_{(a_1,a_2)}\varphi \to \forall_{(a_1,a_2)}\psi$ and enables further reasoning to take place. **CCRp1** and **CCRp2** capture the condition (**atoms**) in Definition 2.1 in terms of formulas, which is important because it provides a logical foundation to remove the quantifier $\forall_{(a_1,a_2)}$ in front of propositional formulas. **CCRD** gives a syntactic description of Corollary 2.6, which lets us transform $\exists_{(A_1,A_2)}$ into a stack of the quantifiers of the form $\exists_{(a_1,a_2)}$. **CCRKco1** reveals that the operator $\exists_{(a_1,a_2)}\nabla_{a_1}$ preserves the inconsistency of $\mathcal{L}_K$-formulas. **CCRKconj** is the $\exists_{(a_1,a_2)}$-over-$\wedge$ distribution. It seems reasonable that calling **CCRKco2**, **CCRKcontra** and **CCRKbis** $\exists_{(a_1,a_2)}$-∇_a crossing laws, which allow us to transform $\exists_{(a_1,a_2)}\nabla_a\Phi$ into a formula of the form $F_a(\exists_{(a_1,a_2)}\Phi)$, where the format F_a depends on the sort of the action a.

As usual, $\vdash \alpha$ ($\vdash_{\boldsymbol{K}} \alpha$) means that α is a theorem in CCRML (resp., ***K*** [7]).

4.1 $\oplus$-construction

This subsection will make some technical preparations for establishing the soundness of the axiom system. Since CCRML contains CC-refinement quantifiers, in order to show the validity of the axioms, we often need to construct a desired model based on the given ones. The $\oplus$-construction presented below will be used to construct these models in a uniform fashion.

Definition 4.1. *Given $B \subseteq A$, let Ω_b be a set of pointed models for each $b \in B$ and M_s a pointed model such that all the models in $\bigcup_{b\in B}\Omega_b \cup \{M_s\}$ are pairwise disjoint. The model N is obtained from $M \uplus (\biguplus_{b\in B, M'_v\in\Omega_b} M')$ by adding a new state s' and imposing the following clauses:*

(1) *for each $b \in B$, $s'R_b^N u$ iff $M'_u \in \Omega_b$ for some M',*

(2) *for each $b \notin B$, $R_b^N(s') = R_b^M(s)$, and*

(3) *for each $p \in Atom$, $s' \in V^N(p)$ iff $s \in V^M(p)$.*

The pointed model $N_{s'}$ is denoted by $(M_s \oplus \{\Omega_b\}_{b\in B})_{s'}$. An illustration for this construction is given in Figure 5.

Convention. In the remainder of the paper, we always suppose that all the models in $\bigcup_{b\in B\subseteq A}\Omega_b \cup \{M_s\}$ are pairwise disjoint whenever the model $M_s \oplus \{\Omega_b\}_{b\in B}$ is used.

The model $M_s \oplus \{\Omega_b\}_{b\in B}$ has some interesting properties whenever M_s and $\{\Omega_b\}_{b\in B}$ meet the requirement presented below.

Definition 4.2 (B-nice)**.** *Given $B \subseteq A$, for each $b \in B$, let $\Phi_b \subseteq_f \mathcal{L}_{CC}$ and Ω_b be a set of pointed models, and let M_s be a pointed model. We say these Φ_b, Ω_b ($b \in B$) and M_s are B-**nice** w.r.t. two given actions $a_1, a_2 \in A$ if the following conditions hold:*

(1) *if $a_1 \in B$ then $\Omega_{a_1} = \Omega_1 \cup \Omega_2$ for some Ω_1 and Ω_2 which satisfy that Ω_1*

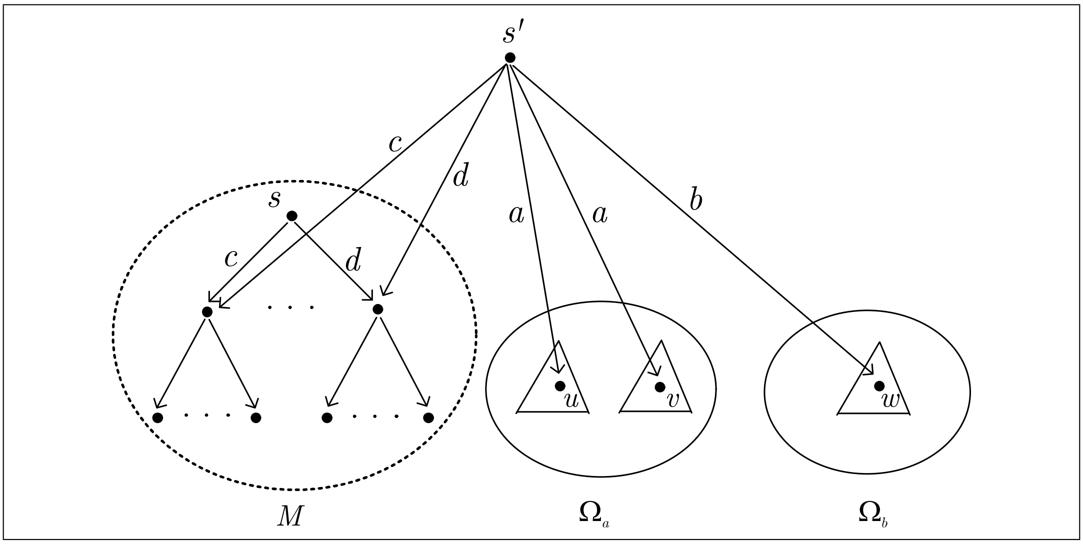


Figure 5: $(M_s \oplus \{\Omega_b\}_{b\in B})_{s'}$ with $A = \{a, b, c, d\}$ and $B = \{a, b\}$

*a_1-**ccrefines** M_s w.r.t. a_1, a_2 and Φ_{a_1} and $\forall\varphi \in \Phi_{a_1} \exists N_t \in \Omega_2$ $(N_t \models \varphi)$,*

(2) *if $a_2 \in B$ then the set Ω_{a_2} a_2-**supports** Φ_{a_2} w.r.t. a_1, a_2 and M_s, and*

(3) *for each b $(\neq a_1, a_2) \in B$, the set Ω_b b-**ccrefines** M_s w.r.t. a_1, a_2 and Φ_b, and b-**supports** Φ_b w.r.t. a_1, a_2 and M_s.*

In verifying Lemma 4.5, 4.6, 4.7 and 4.8 in the next section, we will get desired models by applying the $\oplus$-construction to suitable B-**nice** objects. These desired models enjoy the following property.

Lemma 4.3. *With the notations as in Definition 4.2, if Φ_b, Ω_b $(b \in B)$ and M_s are B-**nice** w.r.t. a_1 and a_2 then, for each new state s',*

$$M_s \succeq_{(a_1,a_2)} (M_s \oplus \{\Omega_b\}_{b\in B})_{s'} \models \bigwedge_{b\in B} \nabla_b \Phi_b.$$

Proof. Let $M'_{s'} \triangleq (M_s \oplus \{\Omega_b\}_{b\in B})_{s'}$. Since Φ_b, Ω_b $(b \in B)$ and M_s are B-**nice** w.r.t. a_1 and a_2, for each $b \in B - \{a_2\}$ and $u \in R_b^M(s)$, we may choose and fix a binary relation $\mathcal{Z}_{bu}$ such that $\mathcal{Z}_{bu} : M_u \succeq_{(a_1,a_2)} N_t$ for some $N_t \in \Omega_b$; moreover, for each $b \in B - \{a_1\}$ and $v \in R_b^{M'}(s')$, we may fix a binary relation $\mathcal{Z}_{bv}$ such that $\mathcal{Z}_{bv} : M_u \succeq_{(a_1,a_2)} N_v$ for some $N_v \in \Omega_b$ and $u \in R_b^M(s)$. Put

$$\mathcal{Z} \triangleq \mathfrak{i}_{S^M, S^{M'}} \cup \{\langle s, s'\rangle\} \cup \bigcup_{\substack{b\in B-\{a_1\} \\ v\in R_b^{M'}(s')}} \mathcal{Z}_{bv} \cup \bigcup_{\substack{b\in B-\{a_2\} \\ u\in R_b^M(s)}} \mathcal{Z}_{bu}.$$

Clearly, $\mathcal{Z} : M_s \succeq_{(a_1,a_2)} M'_{s'}$. Furthermore, by the construction of $M'_{s'}$, since Φ_b, Ω_b ($b \in B$) and M_s are B-***nice*** w.r.t. a_1 and a_2, it is not difficult to see that $M'_{s'} \models \bigwedge_{b\in B} \nabla_b \Phi_b$. $\square$

4.2 Soundness

This subsection devotes itself to establish the soundness of the axiom system CCRML. We begin with giving some validities.

Lemma 4.4. *If $\{\theta_i\}_{1\le i\le |A_1\times A_2|}$ is a permutation of all the pairs in $A_1 \times A_2$, then*

$$\models \exists_{(A_1,A_2)}\varphi \leftrightarrow (\exists_{\theta_1} \cdots \exists_{\theta_{|A_1\times A_2|}})\varphi.$$

Proof. It follows immediately from Corollary 2.6 and Proposition 3.5. $\square$

In fact, for each sequence $\{\theta_i\}_{1\le i\le n}$ of the pairs in $A_1 \times A_2$ such that each action in $A_1 \cup A_2$ occurs in θ_i for some $1 \le i \le n$, we always have

$$\models \exists_{(A_1,A_2)}\varphi \leftrightarrow (\exists_{\theta_1} \cdots \exists_{\theta_n})\varphi.$$

Lemma 4.5. *Let Φ be a finite set of $\mathcal{L}_{CC}$ formulas. Then*

(1) $\models \exists_{(a_1,a_2)}\nabla_{a_1}\Phi \to \Box_{a_1} \bigvee \exists_{(a_1,a_2)}\Phi$,

(2) $\models \Box_{a_1} \bigvee \exists_{(a_1,a_2)}\Phi \to \exists_{(a_1,a_2)}\nabla_{a_1}\Phi$ *whenever each φ in Φ is satisfiable.*

Proof. **(1)** Let $M_s \models \exists_{(a_1,a_2)}\nabla_{a_1}\Phi$. So there exists N_t such that

$$M_s \succeq_{(a_1,a_2)} N_t \models \nabla_{a_1}\Phi.$$

Then $N_t \models \Box_{a_1} \bigvee \Phi$, that is, for each $v \in R^N_{a_1}(t)$, $N_v \models \varphi$ for some $\varphi \in \Phi$. Moreover, since $M_s \succeq_{(a_1,a_2)} N_t$, for each $u \in R^M_{a_1}(s)$, there exists $v' \in R^N_{a_1}(t)$ such that $M_u \succeq_{(a_1,a_2)} N_{v'}$. Therefore, for each $u \in R^M_{a_1}(s)$, $M_u \models \exists_{(a_1,a_2)}\varphi$ for some $\varphi \in \Phi$, which implies $M_u \models \bigvee \exists_{(a_1,a_2)}\Phi$. Hence $M_s \models \Box_{a_1} \bigvee \exists_{(a_1,a_2)}\Phi$.

(2) Suppose that $M_s \models \Box_{a_1} \bigvee \exists_{(a_1,a_2)}\Phi$. Then, by Lemma 3.4 (1), there exists a set (say, E) of pointed models, which a_1-***ccrefines*** M_s w.r.t. a_1, a_2 and Φ. On the other hand, for each $\varphi \in \Phi$, since φ is satisfiable, we may choose arbitrarily and fix a pointed model $N^{\varphi}_{u^{\varphi}}$ such that $N^{\varphi}_{u^{\varphi}} \models \varphi$. Put

$$\Omega \triangleq E \cup \{N^{\varphi}_{u^{\varphi}} \mid \varphi \in \Phi\}.$$

Clearly, Ω, Φ and M_s are $\{a_1\}$-***nice*** w.r.t. a_1 and a_2. Thus, by Lemma 4.3, we get $M_s \succeq_{(a_1,a_2)} (M_s \oplus \Omega)_{s'} \models \nabla_{a_1}\Phi$ for some s'. Hence $M_s \models \exists_{(a_1,a_2)}\nabla_{a_1}\Phi$, as desired. $\square$

The above lemma implies the validity of the axiom scheme **CCRKco2**. Note that we can not adopt the following formula

$$\exists_{(a_1,a_2)}\nabla_{a_1}\Phi \leftrightarrow \Box_{a_1} \bigvee \exists_{(a_1,a_2)}\Phi \text{ where each } \varphi \text{ in } \Phi \text{ is satisfiable}$$

as an axiom scheme, because its side condition uses a semantical concept.

Fortunately, by relying on the completeness of $\boldsymbol{K}$, to prove the completeness of CCRML in Section 4.3, it is sufficient to require Φ to be a finite subset of $\mathcal{L}_K$ and to express the side condition in terms of $\boldsymbol{K}$-derivability (see, **CCRKco2** in Table 1).

Lemma 4.6. $\models \exists_{(a_1,a_2)}\nabla_{a_2}\Phi \leftrightarrow \bigwedge \Diamond_{a_2}\exists_{(a_1,a_2)}\Phi$ *for each* $\Phi \subseteq_f \mathcal{L}_{CC}$.

Proof. Let $M_s \models \exists_{(a_1,a_2)}\nabla_{a_2}\Phi$. So $M_s \succeq_{(a_1,a_2)} N_t \models \nabla_{a_2}\Phi$ for some N_t. We intend to show that $M_s \models \Diamond_{a_2}\exists_{(a_1,a_2)}\varphi$ for each $\varphi \in \Phi$. Let φ be any formula in Φ. Since $N_t \models \nabla_{a_2}\Phi$, we get $N_v \models \varphi$ for some $v \in R^N_{a_2}(t)$. Further, due to $M_s \succeq_{(a_1,a_2)} N_t$, $M_u \succeq_{(a_1,a_2)} N_v$ for some $u \in R^M_{a_2}(s)$. Then $M_u \models \exists_{(a_1,a_2)}\varphi$, and hence $M_s \models \Diamond_{a_2}\exists_{(a_1,a_2)}\varphi$ due to $sR^M_{a_2}u$.

Suppose that $M_s \models \bigwedge \Diamond_{a_2}\exists_{(a_1,a_2)}\Phi$. Then, by Lemma 3.4 (2), there exists a set (say, Ω) of pointed models, which a_2***-supports*** Φ w.r.t. a_1, a_2 and M_s. Clearly, Ω, Φ and M_s are $\{a_2\}$***-nice*** w.r.t. a_1 and a_2. Hence, by Lemma 4.3, we have $M_s \succeq_{(a_1,a_2)} (M_s \oplus \Omega)_{s'} \models \nabla_{a_2}\Phi$ for some s', which implies $M_s \models \exists_{(a_1,a_2)}\nabla_{a_2}\Phi$. □

Lemma 4.7. *Let* $b \in A$ *and* $\Phi \subseteq_f \mathcal{L}_{CC}$. *Then*

(1) $\models \exists_{(a_1,a_2)}\nabla_b\Phi \to \nabla_b\exists_{(a_1,a_2)}\Phi$ *whenever* $b \neq a_1, a_2$,

(2) $\models \nabla_b\exists_{(a_1,a_2)}\Phi \to \exists_{(a_1,a_2)}\nabla_b\Phi$.

Proof. **(1)** Suppose that $M_s \models \exists_{(a_1,a_2)}\nabla_b\Phi$. The analysis similar to that in the proof of Lemma 4.6 and 4.5 (1) shows

$$M_s \models \bigwedge \Diamond_b\exists_{(a_1,a_2)}\Phi \text{ and } M_s \models \Box_b \bigvee \exists_{(a_1,a_2)}\Phi.$$

Then it immediately follows that $M_s \models \nabla_b\exists_{(a_1,a_2)}\Phi$.

(2) Assume that $M_s \models \nabla_b\exists_{(a_1,a_2)}\Phi$. By Lemma 3.4 (3), there exists a set (say, Ω) of pointed models, which b***-ccrefines*** M_s w.r.t. a_1, a_2, and Φ and b***-supports*** Φ w.r.t. a_1, a_2 and M_s. It is clear that Ω, Φ and M_s are $\{b\}$***-nice*** w.r.t. a_1 and a_2. Then, by Lemma 4.3, $M_s \succeq_{(a_1,a_2)} (M_s \oplus \Omega)_{s'} \models \nabla_b\Phi$ for some s', and hence $M_s \models \exists_{(a_1,a_2)}\nabla_b\Phi$. □

It should be pointed out that Lemma 4.7 (1) is invalid without the assumption that $b \neq a_1, a_2$. Consider the models in Figure 6 (1). We have $\mathcal{Z}_1 : M_s \succeq_{(a_1,a_2)} N_t$ with $\mathcal{Z}_1 \triangleq \{\langle s,t\rangle, \langle s_1,t_1\rangle\}$ and $N_t \models \nabla_{a_1}\Phi$ with $\Phi = \{p_1,p_2\} \subseteq \mathit{Atom}$. Thus

$M_s \models \exists_{(a_1,a_2)}\nabla_{a_1}\Phi$, however, it is easy to see that $M_s \not\models \bigwedge \Diamond_{a_1}\exists_{(a_1,a_2)}\Phi$ and hence $M_s \not\models \nabla_{a_1}\exists_{(a_1,a_2)}\Phi$.

Similarly, for the models in Figure 6 (2), we have $\mathcal{Z}_2 : M_s \succeq_{(a_1,a_2)} N_t$ with $\mathcal{Z}_2 \triangleq \{\langle s,t\rangle, \langle s_1,t_1\rangle\}$ and $N_t \models \nabla_{a_2}\Phi$ with $\Phi = \{p_1\} \subseteq Atom$. Thus $M_s \models \exists_{(a_1,a_2)}\nabla_{a_2}\Phi$, but $M_s \not\models \Box_{a_2} \bigvee \exists_{(a_1,a_2)}\Phi$, so $M_s \not\models \nabla_{a_2}\exists_{(a_1,a_2)}\Phi$.

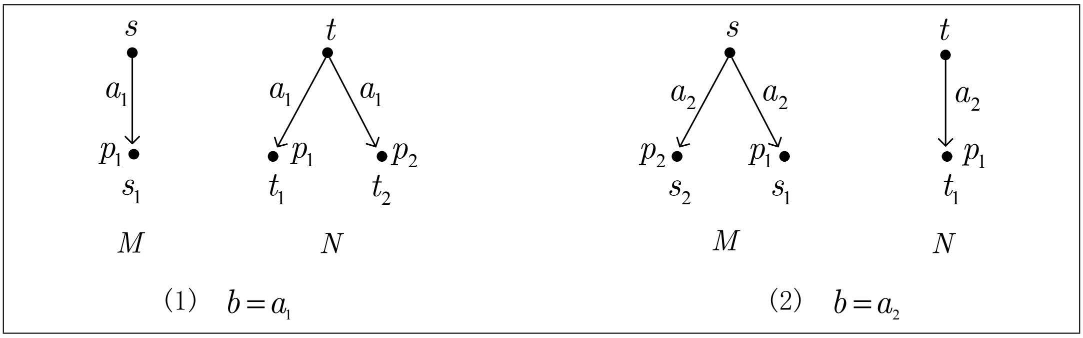


Figure 6: Counterexample used in Lemma 4.7 (1)

Lemma 4.8. *Let $\Phi_b \subseteq_f \mathcal{L}_{CC}$ for each $b \in B(\subseteq A)$. Then*

$$\models \exists_{(a_1,a_2)} \textstyle\bigwedge_{b\in B} \nabla_b\Phi_b \leftrightarrow \bigwedge_{b\in B} \exists_{(a_1,a_2)}\nabla_b\Phi_b.$$

Proof. If α is unsatisfiable for some $\alpha \in \bigcup_{b\in B}\Phi_b$, then

$$\models \bigwedge_{b\in B} \nabla_b\Phi_b \leftrightarrow \bot.$$

Hence

$$\models \exists_{(a_1,a_2)} \bigwedge_{b\in B} \nabla_b\Phi_b \leftrightarrow \bot \quad \text{and} \quad \models \bigwedge_{b\in B} \exists_{(a_1,a_2)}\nabla_b\Phi_b \leftrightarrow \bot.$$

Then it holds trivially that

$$\models \exists_{(a_1,a_2)} \bigwedge_{b\in B} \nabla_b\Phi_b \leftrightarrow \bigwedge_{b\in B} \exists_{(a_1,a_2)}\nabla_b\Phi_b.$$

In the following we deal with the nontrivial case where each α in $\bigcup_{b\in B}\Phi_b$ is satisfiable. It is obvious that

$$\models \exists_{(a_1,a_2)} \bigwedge_{b\in B} \nabla_b\Phi_b \rightarrow \bigwedge_{b\in B} \exists_{(a_1,a_2)}\nabla_b\Phi_b.$$

Next we consider the converse implication. Let $M_s \models \bigwedge_{b\in B} \exists_{(a_1,a_2)}\nabla_b\Phi_b$ and $b \in B$.

If $b \neq a_1, a_2$, by Lemma 4.7 (1), it follows from $M_s \models \exists_{(a_1,a_2)}\nabla_b\Phi_b$ that $M_s \models \nabla_b\exists_{(a_1,a_2)}\Phi_b$. Then, by Lemma 3.4 (3), there exists a set (say, Ω_b) of pointed models, which b-***ccrefines*** M_s w.r.t. a_1, a_2 and Φ_b, and b-***supports*** Φ_b w.r.t. a_1, a_2 and M_s.

If $b = a_2$, by Lemma 4.6, we get $M_s \models \bigwedge \Diamond_{a_2}\exists_{(a_1,a_2)}\Phi_{a_2}$. Then, by Lemma 3.4 (2), there exists a set (say, Ω_{a_2}) of pointed models, which a_2-***supports*** Φ_{a_2} w.r.t. a_1, a_2 and M_s.

If $b = a_1$, by Lemma 4.5 (1), we obtain $M_s \models \Box_{a_1} \bigvee \exists_{(a_1,a_2)}\Phi_{a_1}$. Further, by Lemma 3.4 (1), there exists a set (say, K) of pointed models, which a_1-***ccrefines*** M_s w.r.t. a_1, a_2 and Φ_{a_1}. Moreover, for each $\varphi \in \Phi_{a_1}$, since φ is satisfiable, we may choose arbitrarily and fix a pointed model $N^{\varphi}_{u^{\varphi}}$ such that $N^{\varphi}_{u^{\varphi}} \models \varphi$. Then put

$$\Omega_{a_1} \triangleq K \cup \{N^{\varphi}_{u^{\varphi}} \mid \varphi \in \Phi_{a_1}\}.$$

Clearly, M_s and these sets Φ_b and Ω_b with $b \in B$ are B-***nice*** w.r.t. a_1 and a_2. Then, by Lemma 4.3, $M_s \succeq_{(a_1,a_2)} (M_s \oplus \{\Omega_b\}_{b\in B})_{s'} \models \bigwedge\limits_{b\in B} \nabla_b\Phi_b$ for some s'. Hence it holds that $M_s \models \exists_{(a_1,a_2)} \bigwedge_{b\in B} \nabla_b\Phi_b$, as desired. □

Now we arrive at the soundness of the axiom system.

Theorem 4.9 (Soundness)**.** *For each $\psi \in \mathcal{L}_{CC}$, $\vdash \psi$ implies $\models \psi$.*

Proof. As usual, it is enough to show that all the axiom schemes are valid, and the rules **MP**, **NK** and **NCCR** are sound. It is trivial to check that the axiom schemes **Prop**, **K**, **CCR**, **CCRp1**, **CCRp2** and **CCRKco1** are valid, and the rules **MP**, **NK** and **NCCR** are sound. From Lemma 4.4, 4.5, 4.6, 4.7 and 4.8, it follows that the axiom schemes **CCRD**, **CCRKco2**, **CCRKcontra**, **CCRKbis** and **CCRKconj** are valid. □

4.3 Completeness

This subsection intends to establish the completeness of the axiom system CCRML. This follows by the same method as in [8]. We will show that each $\mathcal{L}_{CC}$-formula is provably equivalent to a ***K***-formula, which brings the completeness of CCRML based on the completeness of ***K***.

We firstly give some general statements as the preparations for the reduction argument.

Proposition 4.10. *Let φ_1, φ_2, $\psi \in \mathcal{L}_{CC}$ and $p \in Atom$. Then*

$$\vdash \varphi_1 \leftrightarrow \varphi_2 \textit{ implies } \vdash \psi[\varphi_1\backslash p] \leftrightarrow \psi[\varphi_2\backslash p]$$

Proof. By induction on the structure of ψ. $\square$

Proposition 4.11.

(1) $\vdash \forall_{(a_1,a_2)}(\varphi \wedge \psi) \leftrightarrow \forall_{(a_1,a_2)}\varphi \wedge \forall_{(a_1,a_2)}\psi$.
(2) $\vdash \exists_{(a_1,a_2)}(\varphi \vee \psi) \leftrightarrow \exists_{(a_1,a_2)}\varphi \vee \exists_{(a_1,a_2)}\psi$.
(3) $\vdash \forall_{(a_1,a_2)}\varphi \vee \forall_{(a_1,a_2)}\psi \to \forall_{(a_1,a_2)}(\varphi \vee \psi)$.
(4) $\vdash \exists_{(a_1,a_2)}(\varphi \wedge \psi) \to \exists_{(a_1,a_2)}\varphi \wedge \exists_{(a_1,a_2)}\psi$.

Proof. Trivially. $\square$

Proposition 4.12. *For each propositional formula β, we have*

(1) $\vdash \forall_{(a_1,a_2)}\beta \leftrightarrow \beta$
(2) $\vdash \exists_{(a_1,a_2)}\beta \leftrightarrow \beta$

Proof. Clearly, Item 2 is implied by Item 1. We intend to show $\vdash \beta \to \forall_{(a_1,a_2)}\beta$ and $\vdash \forall_{(a_1,a_2)}\beta \to \beta$ in turn.

For the former, since each propositional formula is provably equivalent to a formula in disjunctive normal form (**DNF**, for short) [10] and the axiom system CCRML contains all propositional tautologies, we may assume $\vdash \beta \leftrightarrow \bigvee_{\theta\in\Theta}\bigwedge_{r\in\theta} r$, where $\bigvee_{\theta\in\Theta}\bigwedge_{r\in\theta} r$ is a **DNF** formula. Clearly, each θ ($\in \Theta$) is a finite set of literals (proposition letter or the negation of a proposition letter). Then $\vdash \beta \to \forall_{(a_1,a_2)}\beta$ may be inferred as follows.

1.	$\vdash \beta \longrightarrow \bigvee_{\theta\in\Theta}\bigwedge_{r\in\theta} r$	**Prop**
2.	$\vdash \beta \longrightarrow \bigvee_{\theta\in\Theta}\bigwedge_{r\in\theta} \forall_{(a_1,a_2)} r$	**CCRp1**, **p2**, Proposition 4.10
3.	$\vdash \beta \longrightarrow \bigvee_{\theta\in\Theta}\forall_{(a_1,a_2)}\bigwedge_{r\in\theta} r$	Proposition 4.11(1), 4.10
4.	$\vdash \beta \longrightarrow \forall_{(a_1,a_2)}\bigvee_{\theta\in\Theta}\bigwedge_{r\in\theta} r$	Proposition 4.11(3), **Prop**, **MP**
5.	$\vdash \beta \longrightarrow \forall_{(a_1,a_2)}\beta$	Proposition 4.10, **Prop**

Similarly, for the latter, we assume $\vdash \beta \leftrightarrow \bigwedge_{\theta\in\Theta}\bigvee_{r\in\theta} r$, where $\bigwedge_{\theta\in\Theta}\bigvee_{r\in\theta} r$ is a **CNF** formula [10]. Then we have

6.	$\vdash \beta \longrightarrow \bigwedge_{\theta\in\Theta}\bigvee_{r\in\theta} r$	**Prop**
7.	$\vdash \forall_{(a_1,a_2)}\beta \longrightarrow \forall_{(a_1,a_2)}\bigwedge_{\theta\in\Theta}\bigvee_{r\in\theta} r$	**NCCR**, **CCR**, **MP**
8.	$\vdash \forall_{(a_1,a_2)}\beta \longrightarrow \bigwedge_{\theta\in\Theta}\forall_{(a_1,a_2)}\bigvee_{r\in\theta} r$	Proposition 4.11(1), 4.10
9.	$\vdash \forall_{(a_1,a_2)}\beta \longrightarrow \bigwedge_{\theta\in\Theta}\bigvee_{r\in\theta} r$	★
10.	$\vdash \forall_{(a_1,a_2)}\beta \longrightarrow \beta$	

(★) It is easy to see that, for each $\theta \in \Theta$,

$$\vdash \bigvee_{r\in\theta} r \longleftrightarrow (\neg r_1 \to (\neg r_2 \to \cdots (\neg r_{n-1} \to r_n)\cdots)),$$

where $\theta = \{r_1, \cdots, r_n\}$. Then, using **Prop**, **MP**, **CCRp1**, **CCRp2**, **NCCR** and Proposition 4.10 repeatedly, we have

$$\vdash \forall_{(a_1,a_2)}\bigvee_{r\in\theta} r \longrightarrow \bigvee_{r\in\theta} r.$$

This, together with 8., implies 9. as desired. □

The above proposition generalizes the axioms **CCRp1** and **CCRp2**, which guarantees that the CC-refinement quantifier over any propositional formula may be eliminated using proof-theoretical methods. Unfortunately, for a $\mathcal{L}_K$-formula φ, $\exists_{(a_1,a_2)}\varphi$ is not always logically equivalent to φ. Thus, $\nvdash \exists_{(a_1,a_2)}\varphi \leftrightarrow \varphi$ due to the soundness. However, it holds that $\exists_{(a_1,a_2)}\varphi$ is provably equivalent to a $\mathcal{L}_K$-formula in CCRML. To show this, some auxiliary results and notions are needed.

Proposition 4.13. *Let $\beta \in \mathcal{L}_p$ and $\psi \in \mathcal{L}_{CC}$. Then*

$$\vdash \exists_{(a_1,a_2)}(\beta \wedge \psi) \leftrightarrow (\beta \wedge \exists_{(a_1,a_2)}\psi).$$

Proof. By Proposition 4.11 (4), 4.12 and 4.10, it is easy to get

$$\vdash \exists_{(a_1,a_2)}(\beta \wedge \psi) \longrightarrow (\beta \wedge \exists_{(a_1,a_2)}\psi).$$

To complete the proof, we shall show $\vdash (\beta \wedge \exists_{(a_1,a_2)}\psi) \longrightarrow \exists_{(a_1,a_2)}(\beta \wedge \psi)$. Clearly,

$$\vdash \neg(\beta \wedge \psi) \longrightarrow (\beta \to \neg\psi).$$

Then, by **NCCR**, **CCR**, **Prop** and **MP**, we obtain

$$\vdash \forall_{(a_1,a_2)}\neg(\beta \wedge \psi) \longrightarrow (\forall_{(a_1,a_2)}\beta \to \forall_{(a_1,a_2)}\neg\psi).$$

Thus, by **Prop**, it follows that

$$\vdash (\forall_{(a_1,a_2)}\beta \wedge \exists_{(a_1,a_2)}\psi) \longrightarrow \exists_{(a_1,a_2)}(\beta \wedge \psi).$$

Finally, by Proposition 4.12 and 4.10, $\vdash (\beta \wedge \exists_{(a_1,a_2)}\psi) \to \exists_{(a_1,a_2)}(\beta \wedge \psi)$ holds. □

Next we recall the notion of ***disjunctive formula in cover logic*** (***df***, for short) [23]. The ***df*** formulas are generated by the BNF grammar below, where $\emptyset \neq B \subseteq A$ and $\beta_0 \in \mathcal{L}_p$.

$$\beta ::= (\beta \vee \beta) \mid \beta_0 \mid (\beta_0 \wedge \bigwedge_{b \in B} \nabla_b\{\beta, \cdots, \beta\})$$

Proposition 4.14 ([23])**.** *For each $\varphi \in \mathcal{L}_K$, there is a **df** formula β such that $\vdash_{\boldsymbol{K}} \varphi \leftrightarrow \beta$.*

Now we can show that any formula of the form $\exists_{(a_1,a_2)}\beta$ with $\beta \in \mathcal{L}_K$ can be provably reduced to a $\mathcal{L}_K$-formula.

Proposition 4.15. *For each $\beta \in \mathcal{L}_K$,*

$$\vdash \exists_{(a_1,a_2)}\beta \leftrightarrow \xi \quad \text{for some } \xi \in \mathcal{L}_K.$$

Proof. Since the axiom system presented in this paper contains the system $\boldsymbol{K}$, by Proposition 4.14, for each $\varphi \in \mathcal{L}_K$, we have $\vdash \varphi \leftrightarrow \alpha$ for some ***df*** formula α. Further, by Proposition 4.10, we may assume that β is a ***df*** formula. In the following, we proceed by the induction on β.

For $\beta \in \mathcal{L}_p$, it holds by Proposition 4.12.

For $\beta \equiv \beta_1 \vee \beta_2$, by Proposition 4.11 (2), we get

$$\vdash \exists_{(a_1,a_2)}\beta \longleftrightarrow \exists_{(a_1,a_2)}\beta_1 \vee \exists_{(a_1,a_2)}\beta_2.$$

Then, by the induction hypothesis and Proposition 4.10, it follows that

$$\vdash \exists_{(a_1,a_2)}\beta \longleftrightarrow \xi_1 \vee \xi_2 \ \text{ for some } \xi_1,\ \xi_2 \in \mathcal{L}_K.$$

For $\beta \equiv \beta_0 \wedge \bigwedge_{b\in B} \nabla_b \Phi_b$ with $\beta_0 \in \mathcal{L}_p$ and $\Phi_b \subseteq$ ***df*** for each $b \in B$, by Proposition 4.13, we have

$$\vdash \exists_{(a_1,a_2)}\beta \longleftrightarrow \beta_0 \wedge \exists_{(a_1,a_2)} \bigwedge_{b\in B} \nabla_b \Phi_b.$$

Further, by **CCRKconj** and Proposition 4.10, it follows that

$$\vdash \exists_{(a_1,a_2)}\beta \longleftrightarrow \beta_0 \wedge \bigwedge_{b\in B} \exists_{(a_1,a_2)} \nabla_b \Phi_b.$$

Now the proof is completed by showing that, for each $b \in B$,

$$\vdash \exists_{(a_1,a_2)}\nabla_b\Phi_b \longleftrightarrow \xi_b \ \text{ for some } \xi_b \in \mathcal{L}_K. \tag{$**$}$$

Since $\Phi_b \subseteq$ ***df*** $\subseteq \mathcal{L}_K$, applying the axiom schemes **CCRKco1**, **CCRKco2**, **CCRKcontra** and **CCRKbis**, it follows that $\vdash \exists_{(a_1,a_2)}\nabla_b\Phi_b \longleftrightarrow \theta$ for some θ in which the quantifiers $\exists_{(a_1,a_2)}$ are over only the formulas in Φ_b ($\subseteq$ ***df***). Finally, by the induction hypothesis and Proposition 4.10, the claim $(**)$ follows, as desired. □

At this point, we can show that all $\mathcal{L}_{CC}$-formulas can be provably reduced to $\mathcal{L}_K$-formulas, which is the crucial step in establishing the completeness of CCRML.

Proposition 4.16. *For each $\psi \in \mathcal{L}_{CC}$, there exists a formula $\varphi \in \mathcal{L}_K$ such that $\vdash \psi \leftrightarrow \varphi$.*

Proof. By the axiom scheme **CCRD** and Proposition 4.10, we only need to deal with the formulas in which all CC-refinement quantifiers are of the form $\exists_{(a,b)}$ with $a, b \in A$. We proceed by the induction on the number $n(\psi)$ of the occurrences of the CC-refinement quantifiers in ψ.

For $n(\psi) = 0$, trivially. If $n(\psi) > 0$, we can always find a subformula of ψ, which is of the form $\exists_{(a_1,a_2)}\theta$ with $\theta \in \mathcal{L}_K$. Then, by Proposition 4.15 and 4.10, it follows easily that $\vdash \psi \leftrightarrow \psi'$ for some $\psi' \in \mathcal{L}_{CC}$ with $n(\psi') < n(\psi)$. This enables the induction proof to work well. □

Proposition 4.17. *Let $\psi \in \mathcal{L}_{CC}$ and $\varphi \in \mathcal{L}_K$ such that $\vdash \psi \leftrightarrow \varphi$. If φ is a theorem in $\boldsymbol{K}$, then so is ψ in CCRML.*

Proof. Since the axiom system $\boldsymbol{K}$ is contained in CCRML, we have $\vdash \varphi$ due to $\vdash_{\boldsymbol{K}} \varphi$. Further, $\vdash \psi$ follows immediately from $\vdash \psi \leftrightarrow \varphi$. □

Theorem 4.18 (Completeness)**.** *For each $\psi \in \mathcal{L}_{CC}$, $\models \psi$ implies $\vdash \psi$.*

Proof. By Proposition 4.16, $\vdash \psi \leftrightarrow \xi$ for some $\xi \in \mathcal{L}_K$. Then, by Theorem 4.9, we get $\models \psi \leftrightarrow \xi$, which implies $\models \xi$ because of $\models \psi$. Hence $\vdash_{\boldsymbol{K}} \xi$ due to the completeness of $\boldsymbol{K}$. Thus, by Proposition 4.17, we get $\vdash \psi$, as desired. □

To establish the completeness of CCRML, we have showed that all $\mathcal{L}_{CC}$-formulas can be provably reduced to $\mathcal{L}_K$-formulas. It follows easily from our proof that there is an algorithm for transforming each $\mathcal{L}_{CC}$-formula ψ into a $\mathcal{L}_K$-formula φ such that $\vdash \psi$ iff $\vdash_{\boldsymbol{K}} \varphi$. Thus, due to the decidability of the system $\boldsymbol{K}$, we have

Theorem 4.19 (Decidability)**.** *CCRML is decidable.* □

In the proof of Proposition 4.15, our argument is given based on the assumption that β is a ***df*** formula. Such assumption does not obstruct getting the above conclusion because each $\mathcal{L}_K$-formula can be effectively transformed into an equivalent ***df*** formula.

5 Discussion

This paper provides a sound and complete axiom system for the CC-refinement quantifiers $\exists_{(A_1,A_2)}$ under the assumption that neither A_1 nor A_2 is empty. This assumption is not particularly restrictive. Similar to $\exists_{(A_1,A_2)}$, the quantifiers $\exists_{(A_1,\emptyset)}$ and $\exists_{(\emptyset,A_2)}$ can be reduced to the quantifiers $\exists_{(a,\emptyset)}$ and $\exists_{(\emptyset,a)}$ respectively due to Corollary 2.6. Moreover, the proofs and constructions with minor modification still work if we drop such assumption. We leave it to the reader to check this. Here we only explain the necessary modification of the axiom system.

The axiom system for $\exists_{(A_1,\emptyset)}$ is obtained from the axiom system in Table 1 by replacing the quantifier $\forall_{(a_1,a_2)}$ (or, $\exists_{(a_1,a_2)}$) with $\forall_{a_1}$ ($\exists_{a_1}$, resp.) in all the axioms and rules, deleting the axiom **CCRKcontra**, rewriting **CCRD** according to Corollary 2.6 and replacing the side condition $b \neq a_1, a_2$ of **CCRKbis** with $b \neq a_1$.

To obtain the axiom system for $\exists_{(\emptyset,A_2)}$, in addition to replacing the quantifier $\forall_{(a_1,a_2)}$ (or, $\exists_{(a_1,a_2)}$) with $\forall_{a_2}$ ($\exists_{a_2}$, resp.) in all the axioms and rules and rephrasing the axiom **CCRD** based on Corollary 2.6, we erase the axioms **CCRKco1** and **CCRKco2**, and replace the side condition $b \neq a_1, a_2$ of **CCRKbis** with $b \neq a_2$. The obtained axiom system is indeed the one given in [8].

Based on the explanation given above, it is straightforward to integrate all these axiom systems into one. This paper prefers to adopt the present framework for the sake of simplicity.

References

[1] F. Aarts, F. Heidarian, and F. Vaandrager. A theory of history dependent abstractions for learning interface automata. In *M. Koutny, I. Ulidowski (Eds.), International Conference on Concurrency Theory, CONCUR 2012, LNCS*, volume 7454, pages 240–255. Springer, 2012.

[2] F. Aarts and F. Vaandrager. Learning I/O automata. In *P. Gastin, F. Laroussinie (Eds.), International Conference on Concurrency Theory, CONCUR 2010, LNCS*, volume 6269, pages 71–85. Springer, 2010.

[3] L. Aceto, I. Fábregas, D. de F. Escrig, A. Ingólfsdóttir, and M. Palomino. Relating modal refinement, covariant-contravariant simulations and partial bisimulations. In *F. Arbab and M. Sirjani (Eds.), Fundamentals of Software Engineering, FSEN 2011, LNCS*, volume 7141, pages 268–283. Springer, 2011.

[4] J. Baeten, D. van Beek, B. Luttik, J. Markovski, and J. Rooda. *Partial bisimulation.* SE Report 2010-04, Systems Engineering Group, Department of Mechanical Engineering, Eindhoven University of Technology, 2010.

[5] P. Balbiani, A. Baltag, H. van Ditmarsch, A. Herzig, T. Hoshi, and T. De Lima. 'knowable' as 'known after an announcement'. *Review of symbolic Logic*, 1(3):305–334, 2008.

[6] M. Bílková, A. Palmigiano, and Y. Venema. Proof systems for the coalgebraic cover modality. In *C. Areces, R. Goldblatt (Eds.), Advances in Modal Logic*, volume 7, pages 1–21. College Publication, 2008.

[7] P. Blackburn, M. de Rijke, and Y. Venema. *Modal logic.* Cambridge University Press, 2002.

[8] L. Bozzelli, H. van Ditmarsch, T. French, J.Hales, and S. Pinchinat. Refinement modal logic. *Information and Computation*, 239:303–339, 2014.

[9] L. Bozzelli, H. van Ditmarsch, and S. Pinchinat. The complexity of one-agent refinement modal logic. *Theoretical Computer Science*, 7519:303–339, 2015.

[10] I. Chiswell and W. Hodges. *Mathematical logic, Oxford texts in logic 3.* Oxford university press, 2007.

[11] G. d'Agostino and G. Lenzi. An axiomatization of bisimulation quantifiers via the μ-calculus. *Theoretical Computer Science*, 338(1-3):64–95, 2005.

[12] G. d'Agostino and G. Lenzi. A note on bisimulation quantifiers and fixed points over transitive frames. *Journal of Logic and Computation*, 18(4):601–614, 2008.

[13] I. Fábregas, D. de F. Escrig, and M. Palomino. Non-strongly stable orders also define interesting simulation relations. In *A. Kurz, M. Lenisa and A. Tarlecki (Eds.), Algebra and Coalgebra in Computer Science, CALCO 2009, LNCS*, volume 5728, pages 221–235. Springer, 2009.

[14] I. Fábregas, D. de F. Escrig, and M. Palomino. Logics for contravariant simulations. In *J. Hatcliff and E. Zucca (Eds.), Formal Techniques for Distributed Systems, Joint 12th IFIP WG 6.1 International Conference, FMOODS 2010 and 30th IFIP WG 6.1 International Conference, FORTE 2010, LNCS*, volume 6117, pages 224–231. Springer, 2010.

[15] I. Fábregas, D. De F. Escrig, and M. Palomino. Equational characterization of covariant-contravariant simulation and conformance simulation semantics. In *L. Aceto and P. Sobociński (Eds.), Seventh Workshop on Structural Operational Semantics 2010, ar Xiv:1008.2108, Electronic Proceedings in Theoretical Computer Science, EPTCS 2010*, volume 32, pages 1–14, 2010.

[16] T. French. *Bisimulation quantifiers for modal logic.* PhD thesis, University of Western Australia, 2006.

[17] M. Hennessy and R. Milner. Algebraic laws for indeterminism and concurrency. *Journal of ACM*, 32:137–162, 1985.

[18] K.G. Larsen. Modal specifications. In *J. Sifakis (Eds.), Automatic Verification Methods for Finite State Systems, International Workshop, Grenoble, France, LNCS*, volume 407, pages 232–246. Springer, 1989.

[19] K.G. Larsen and B. Thomsen. A modal process logic. In *Third Annual Symposium on Logic in Computer Science*, pages 203–210. IEEE Computer Society Press, 1988.

[20] H. van Ditmarsch and T. French. Simulation and information. In *J.J.C. Meyer and J.M. Broersen (Eds.), Knowledge Representation for Agents and Multi-agent Systems, KRAMAS 2008, LNAI*, volume 5605, pages 51–65. Springer, 2009.

[21] H. van Ditmarsch, T. French, and S. Pinchinat. Future event logic-axioms and complexity. In *L. Beklemishev, V. Goranko and V. Shehtman (Eds.), Advances in Modal logic, Moscow*, volume 8, pages 77–99. College Publications, 2010.

[22] R.J. van Glabbeek. The linear time-branching time spectrum I. In *J.A. Bergstra, A. Ponse and S.A. Smolka (Eds.), Handbook of Process Algebra*, pages 5–97. Elsevier, 2001.

[23] Y. Venema. *Lecture notes on the modal μ-calculus.* Draft, 2012.

Received 1 December 2016

Did the Neo-Babylonians Construct a Symbolic Logic for Legal Proceedings?

Andrew Schumann
University of Information Technology and Management in Rzeszow, Rzeszow, Poland.

Abstract

In this paper, I show that the first proto-axiomatic system of symbolic logic was created within the Sumerian-Akkadian legal culture (i.e. much earlier than Aristotle and the Stoics lived) to make justice effective and transparent. The point is that there were excavated many well-preserved Neo-Babylonian trial records and we can reconstruct logically legal proceedings reported in them. There are direct evidences that their trial decisions were drawn automatically just by applying some inference rules. A very similar structure and the same perfect logicality are observed in legal documents in Aramaic and Greek from Elephantine. Hence, symbolic logic existed at least since the Neo-Babylonians and was preserved by Arameans and continued by Greeks.

1 Introduction

All the life of Akkadians was structured by two systems: (i) justice (*dīnu*) and (ii) omens (*alaktu*) [10], [14]. So, there were detailed lists of signs (*ittātu* / GISKIM.MEŠ) of two types: (i) articles of the law code; (ii) omens (forecastings) about everyting with correlated events for divination. Each article of the code and each omen for divination has the same logical structure "If (Akkadian *šumma*) S, then P." The code was used for legal proceedings (to serve *dīnu*) and each court looked for an appropriate article of the code for a claim S to infer the trial decision P from the article "If S, then P" and the proved claim S just by *modus ponens*. The omens (as a part of *alaktu*) were used for prognoses and each scholar looked for an appropriate omen for an event S to infer the forecasting P from the omen "If S, then P" and the appearance of S just by *modus ponens*, as well [18]. Hence, *modus ponens* was the most significant inference rule in the Sumerian-Akkadian culture. Also, we can find many examples of *modus tollens* in legal proceedings and forecastings.

In Greece, Chrysippus (279 – 206 B.C.) defined *modus ponens* and *modus tollens* for the first time correctly. His samples:

- *modus ponens*: "If it is day, it is light; but in fact it is day; therefore it is light" (Sextus Empiricus, *Against the Logicians* II, 224);
- *modus tollens*: "If it is day, it is light; but it is not light; therefore it is not day" (Sextus Empiricus, *Against the Logicians* II, 225).

We do not know whether Chrysippus applied his propositional logic in a legal hermeneutics. But his prominent Roman follower, Marcus Tullius Cicero (106 – 43 B.C.) did it well in his *Topica* [8]. The Babylonians did it well too, but many and many centuries earlier, and not only in legal proceedings, but also in divination.

In this paper, I propose a kind of formalization for *dīnu* and I show that legal proceedings reported in Neo-Babylonian trial records were based on an axiomatic system, where trial decisions were drawn automatically just by applying some inference rules to articles of a code or to previous trial decisions. It is a really significant fact that Neo-Babylonian trial records and business contracts, including inheritance records, marriage contracts, and other documents concerning money, movable and immovable properties, consist of complex syllogisms without logical fallacies at all. These syllogisms are well formalizable within the contemporary propositional logic. Omens (forecastings), i.e. *alaktu*, can be formalized logically also, but within a non-classical logic.

The Talmud is a continuation of Aramaic legal tradition that became a continuation of Neo-Babylonian (Akkadian) law in turn. The Talmudic hermeneutics studied logically in [1] – [2], [20] – [25] has the following two parts: (i) a *classical propositional logic* (used before the Talmud by the Assyrians and Neo-Babylonians for inferring court decisions on the basis of laws, Hebrew: *dīn*, Aramaic: *dīn'a*, a derivative word from the Akkadian *dīnu*); (ii) some *non-classical logics* for deducing new laws from the text of the Torah (Hebrew: *halaḵah*, a derivative word from the Akkadian *alaktu*). Both parts have their roots in the Babylonian tradition: (i) at first, a kind of classical propositional logic was established by the Sumerians to interpret the law codes; in these codes the casuistic law formulation was used: "if/when (Akkadian: *šumma*) this or that occurs, this or that must be done" in the same way how it is formulated in the Bible, e.g. "If [*w 'im*] he has not been redeemed in any of those ways, he and his children with him shall go free in the jubilee year" (*Leviticus* 25:54); so, a trial decision looked like an inference by *modus pones* or by other logical rules (such as *modus tollens*, substitution rule, and many other rules) from an appropriate article in the law code, see [25]; (ii) some first non-classical logics were invented within a forecasting tradition, where since the Old Babylonian period (ca. 1800 B.C.) each omen (forecasting) rule was formulated in the form of implication: "if (*šumma*) P, [then] Q"; so, a concrete forecasting looked like an inference by *modus pones* or *modus tollens*, too [18]; but to induce new omen rules their authors appealed to some non-classical logics.

Conventionally, Aristotle (384 – 322 B.C.) is considered a father of symbolic logic. Nevertheless, we can assume that the Greek logic (rather the Stoic one) was based on a Sumerian-Akkadian legal hermeneutics. On the basis of trial records [11], [12] (notice that some Hebrew trial records in Aramaic made within a Babylonian tradition are found at Elephantine in Egypt [6]), it is possible to show that judges knew a propositional logic well that was applied in deducing court decisions [25]. Hence, the origin of classical symbolic logic should have been connected to establishing a logical tradition of the Sumerian-Akkadian jurisprudence at first. The legal tradition of the Talmud is a direct continuation of the Babylonian tradition.

The structure of my paper is based on the following conclusion drawn by applying the Mill's joint method of agreement and difference:

> The oldest fragments over the world consisting of complex syllogisms without logical fallacies (x) are presented by Mesopotamian legal documents (trial records and business contracts) (A);
>
> The Neo-Babylonian legal documents are made in Akkadian (B) within the Mesopotamian legal tradition (A);
>
> The Neo-Babylonian legal documents (B) are well formalizable as an axiomatic system extending the classical propositional logic to draw legal conclusions (y);
>
> The Aramaic legal documents written in Aramaic (C) and excavated in Elephantine belong to the Jews (D);
>
> They (C) are made within the Aramaic legal tradition of the Achaemenid dynasty continuing the Akkadian legal tradition of Neo-Babylonians (B);
>
> The Aramaic legal documents of Elephantine (C) are well formalizable within the classical propositional logic of drawing legal conclusions and this logic is the same as of the Neo-Babylonian trial records (y);
>
> The Greek (then Byzantine) legal documents of Elephantine (E) are

a direct continuation of the Aramaic legal tradition of Elephantine (C);

These documents (E) are well formalizable within the classical propositional logic of drawing legal conclusions, too (y).

Therefore, most probably, the first symbolic propositional logic (x y) was founded within the Mesopotamian legal tradition to draw legal conclusions (A B)

Formally:

A occurs together with x

A B occur together with x y

A B C D E occur together with x y

Therefore A is the cause, or the effect, or a part of the cause of x and B is the cause, or the effect, or a part of the cause of y

Hence, my reasoning follows the Mill's joint method of agreement and difference to claim that the Sumerian-Akkadian jurisprudence (A) was a possible cause of using propositional logic correctly (x) and the Neo-Babylonian ways of making trial decisions (B) was a possible cause of using the axiomatic system extending the classical propositional logic to draw legal conclusions (y).

For the first time this idea was put forward in [25]. In that paper, I showed a strong connection of symbolic logic to the Ancient legality. Now I am proposing a logical formalization of legal reasoning belonging to New-Babylonians, Achaemenid Arameans, and Greeks of Hellenistic kingdoms to show that at the time of first Greek logicians there was a well-developed tradition of symbolic logic in legality rooted in the New-Babylonian and Achaemenid jurisprudence.

2 Logical Context

Main presuppositions of symbolic logic are formulated as follows:

1. each sentence (proposition or statement) is either true or false;
2. a true sentence can be always verified by a reference to an appropriate fact described in this sentence and taking place in reality indeed;

3. there are logical inference rules which give only true conclusions from true premises;

4. *modus ponens* and *modus tollens* are two basic inference rules. Let A and B be some propositions which can be either true or false. Then the rule of *modus ponens* is defined thus: Assume that 'If A, then B' and A are true, then we can conclude that B is true, too. The rule of *modus tollens*: 'If A, then B' is true, but B is false, then A is false, also.

For the first time, a symbolic logic with these presuppositions was explicitly introduced by Chrysippus (ca. 279 – 206 B.C.), the Stoic philosopher. There are no preserved texts authored by him, only some quotations. Thereby among all these quotations, the best illustration of his reasoning is given in the *Topica* written by Cicero (106 B.C. – 43 B.C.). All the examples of Chrysippian conclusions in this treatise are taken from the legal hermeneutics, e.g.:

> If a slave has not been declared free either by the censor, or by the praetor's rod, or by the will of his master, he is not free: but none of those things is the case: therefore he is not free [7].

> If S is not A ('a slave declared free by the censor') or S is not B ('a slave declared free by the praetor's rod') or S is not C ('a slave declared free by the will of his master'), then S is not D ('free'). S is not A. S is not B. S is not C.
>
> ---
>
> Then S is not D

In this reasoning there are applied two inference rules: (i) *introducing disjunction* ('S is not A; S is not B; S is not C; then S is not A or S is not B or S is not C); (ii) *modus ponens* ('If S is not A or S is not B or S is not C, then S is not D; S is not A or S is not B or S is not C; then S is not D'). Hence, we deal with the conclusions with two logical steps here. It is not so simple. But nevertheless, according to the *Topica*, even complicated logical conclusions of the Chrysippian symbolic logic were actively being applied in jurisprudence and, most probably, in trial decisions at the time of Cicero.

It is a great surprise that the Chrysippian symbolic logic was being applied in trial records, business contracts, and testaments much much earlier than Chrysippus lived, namely since some (not all) first legal documents of the third dynasty of Ur (ca. 2047 – 2030 B.C.).

Now that is the question whether a symbolic logic can be used unintentionally (i.e. without studying a logic as a system previously). Let us consider the following Lewis Carroll's logical puzzle. We have the four premises and we should find out a conclusion from them:

> (a) None of the unnoticed things, met with at sea, are mermaids.
>
> (b) Things entered in the log, as met with at sea, are sure to be worth remembering.
>
> (c) I have never met with anything worth remembering, when on a voyage.
>
> (d) Things met with at sea, that are noticed, are sure to be recorded in the log.

Everybody who did not study a symbolic logic deeply cannot solve this task in any way. But the answer is very simple and made only by the four logical steps: 'I have never met with a mermaid at sea'. Another similar puzzle proposed by Lewis Carroll:

> (a) All babies are illogical.
>
> (b) Nobody is despised who can manage a crocodile.
>
> (c) Illogical persons are dispised.

It can be solved now by three steps in inferring. The true answer is as follows: 'Anyone who can manage a crocodile is not a baby'. But it is impossible to be solved by anyone who did not study logic in advance, too.

Now let us take a task assuming only one step in inferring:

> (a) My gardener is well worth listening to on military subjects;
>
> (b) Nobody is really worth listening to on military subjects, unless he can remember the battle of Waterloo.

How many people are able to conclude correctly from these two premises? Only a few respondents. But in Babylonian trial records, business contracts, and testaments there are correct conclusions with ten and even much more logical steps. And we face no logical fallacies at all. It cannot be occasionally. The point is that symbolic logic is not an innate knowledge and it is not an innate ability. The Babylonian legal tradition was the only tradition in the human history before Aristotle and Chrysippus that had had a perfect logical competence: very large and complicated logical conclusions without fallacies. It is a greatest mystery in the history of science how it was possible. There is neither logical textbook nor logical treatise in Sumerian or Akkadian excavated still (but we have some late logical parts of Talmudic books in Aramaic). But we have no excavated grammar of Akkadian or Sumerian, too, although there are known first dictionaries – Sumerian and Akkadian lexical lists ordered by topic, such as ur_5-ra = *ḫu-bul-*lu_4, see [15].

3 Historical Context

There are the following cuneiform codices [13], [29]: (i) the *Code of Ur-Nammu* (ca. 2047 – 2030 B.C.) belonging to the third dynasty of Ur and written in Sumerian; (ii) the *Code of Lipit-Ištar* (ca. 1870 B.C. – c. 1860 B.C.) written also in Sumerian and belonging to the fifth king of the first dynasty of Isin; (iii) the *Laws of Ešnunna* (ca. 1790 B.C.) written in Akkadian and belonging to a king of Ešnunna, the city located in northern Mesopotamia on the Tigris river and becoming politically important after the fall of the third dynasty of Ur; (iv) the *Code of Hammurabi* (ca. 1728 – 1686 B.C.) written in Akkadian and belonging to Hammurabi, the sixth king of the first Babylonian dynasty; (v) the *Middle Assyrian Laws* (since the twelfth century B.C.) written in the Assyrian dialect of Akkadian and much connected to Tiglath-Pileser I, the king of the Assyrian Empire; (vi) the *Hittite Laws* (ca. 1650 – 1100 B.C.) written in Hittite and containing the corpus of laws of the Hittite Empire; (vii) the *Neo-Babylonian Laws* (626 B.C. – 539 BC) written in the Neo-Babylonian dialect of Akkadian and containing the corpus of laws of the Neo-Babylonian Empire. Aramaic was the everyday language at the time of Neo-Babylonians, but this language became the language of administration and culture only since the Achaemenid Empire (ca. 550 B.C. – 330 B.C.), although cuneiform texts in Akkadian remained popular, also. Hence, (viii) the *Achaemenid Laws* were written in the Neo-Babylonian dialect of Akkadian as well as in Aramaic in the Aramaic script. Since establishing the Hellenistic Empire by Alexander the Great (334 B.C. – 323 B.C.), (ix) the *Hellenistic Laws* were written mainly in Greek, but also there was used Aramaic for the same purpose.

In Assyriology there is a quite popular claim that these codes and laws did not function as normative and binding legislation. For example, Raymond Westbrook puts forward a hypothesis that the codes were legal treatises, i.e. a secondary legal authority just to show possible legal lists with individual items formulated in casuistic style [29]. There is even a more radical opinion that the codes were only a kind of ideology like the Bible was for Christians and it is until now. For instance, in Christianity different Biblical commandments such as ‘thou shalt not steal’ or ‘thou shalt not covet’ never played a basis of legislation. In Christianity the Bible was never used in criminal and civil laws. It is explained by that since Constantine the Great (ca. 272 A.D. – 337 A.D.) Christianity was accepted in the Roman Empire only as ideology – the Roman laws are maintained to be applied as criminal and civil laws further. The idea that the Babylonian codes and laws never were a form of legislation is a unconscious “Christianization” of Babylonians.

Among different arguments supporting this “Christianization” of Babylonians the strongest is that we never find explicit references to these codes as sources of law. The problem is that many cuneiform tablets are not published yet. Nevertheless, in unpublished texts there are these references in fact. For example, in some fragments of trial and business documents of Old-Assyrian period (ca. 2025 B.C. – 1378 B.C.) there is an explicit expression “the words of the stele” (*awāt naruāim*), see for the details in [28], with the meaning to be such a reference. Also, we know that these references were not necessary. They were just assumed implicitly. So, there was excavated a very important Hebrew ostracon at Meẓad Hašavyahu, dated to from 630 B.C. to 609 B.C., the time of Josiah, the king of Judah [5, p. 568]. The text of this ostracon is as follows:

ישמע אדני השר
את דבר עבדה עבדך
קצר היה עבדך בח
צר אסם ויקצר עבדך
ויכל ואסם כימם לפני שב
ת כאשר כל עבדך את קצר וא
סם כימם ויבא הושעיהו בן שב
י ויקח את בגד עבדך כאשר כלת
את קצרי זה ימם לקח את בגד עבדך
וכל אחי יענו לי הקצרם אתי בחם
השמש אחי יענו לי אמן נקתי מא
לתך ויקח בגדי ואמלא לשר להש
ב את בגד עבדך ותתן אלו רח
מם והשבת את בגד עבדך ולא תדהם ן

Its translation:

> Let my lord, the governor, hear the word of his servant! Your servant is a reaper. Your servant was in Haẓar 'Asam, and your servant reaped, and he finished, and he was storing up (the grain) during these days before the Sabbath. When your servant had finished the harvest, and was storing (the grain) during these days, Hošeyahu came, the son of Šobi, and he seized the garment of your servant, when I had finished my harvest. It (is already now some) days (since) he took the garment of your servant. And all my companions can bear witness for me – they who reaped with me in the heat of the harvest – yes, my companions can bear witness for me. Amen! I am innocent from guilt. And he stole my garment! It is for the governor to give back the garment of his servant. So grant him mercy in that you return the garment of your servant and do not be displeased [26, p. 96].

As we see, the author of the text makes a statement in court. He affirms that he has at least two witnesses who can support his claim – it is a sufficient amount of witnessing at a Jewish (Babylonian) court [17]. In his claim he is appealing to the unjust act of the fortress's governor who confiscated the garment on the eve of the Sabbath. Whether the fortress's governor could take this garment? According to the Torah, he cannot:

> 26 If thou at all take thy neighbour's raiment to pledge, thou shalt deliver it unto him by that the sun goeth down:
> 27 For that is his covering only, it is his raiment for his skin: wherein shall he sleep? and it shall come to pass, when he crieth unto me, that I will hear; for I am gracious (KJV, *Exodus* 22).

> 12 And if the man be poor, thou shalt not sleep with his pledge:
> 13 In any case thou shalt deliver him the pledge again when the sun goeth down, that he may sleep in his own raiment, and bless thee: and it shall be righteousness unto thee before the Lord thy God (KJV, *Deuteronomy* 24).

Thus, the author refers to the Torah only implicitly. But it is enough, because only the Torah was accepted as a primary source of legislation at the time of Josiah, the king of Judah (ca. 649 – 609 B.C.). The same situation can hold for Neo-Babylonian trial decisions. We do not need explicit references there too.

The Hebrew ostracon is interesting also as it assumes that an appropriate trial decision should be drawn only by two logical steps through applying *modus ponens* and *substitution rule*:

> 'If thou at all take thy neighbour's raiment to pledge, thou shalt deliver it unto him by that the sun goeth down' is true (because it is an axiom from the Torah);
>
> 'the fortress's governor took the raiment of the claimant' is true (because it is verified by witnessing).
>
> ---
>
> The fortress's governor should deliver the raiment unto the claimant back.

In the Neo-Babylonian trial records much more logical steps occur in drawing trial decisions and always it is supposed that there are axioms taken from a code, although without a direct reference. For instance, for stealing one item belonging to the god, the thief should give the same item thirtyfold. According to the Neo-Babylonian trial records, it is an axiom. But we know that this statement is an article of the *Hammurabi Code* [19, p. 82]. Thus, we can suggest that this statement was an axiom for courts since the time of Hammurabi.

In this paper, I am proposing a very strong structuralist argument confirming the hypothesis of Babylonian legislation: (i) the symbolic-logical reasoning of trial records are very complicated; so, they assume a perfect logical competence of their authors, although this competence is typical only for logicians today; hence, we can assume that this extraordinary competence was a part of legal hermeneutics existed indeed; (ii) in this symbolic-logical reasoning there are assumed some axioms; many of them occur in some Babylonian codes; thus, we can suggest that all these axioms in trial conclusions were taken from appropriate articles in Babylonian codes[1].

[1] It is worth noting that my argumentation on legislation in Babylonia was criticized by the referee. According to that opinion, my argumentation supports the thesis indeed that legislation would exist in Mesopotamia. Nevertheless, first, there are a few of fragments which can be examined as Neo-Babylonian laws and they do not allow us to reconstruct a Neo-Babylonian code as a whole (too few fragments), therefore we can reconstruct some implicit references only to laws from different periods in history, but "it is not at all clear that a Neo-Babylonian court would have known the text of an earlier law 'code,' such as the Laws of Ešnunna". Second, "there are other examples of Neo-Babylonian trial records for which we do not have parallel 'legislation' at all". Hence, by the referee, it seems simpler to suggest that "the courts are operating on the basis of a sort of customary law, without reference to writing". In my paper, I have not addressed the existence of Akkadian legal decisions from earlier periods. About them please see [13]. I agree with the referee

4 Axiomatic System for Legal Proceedings

The Sumerian law codes were first over the world, e.g. *Ur-Nammu* (ca. 2047 – 2030 B.C.) and *Lipit-Ištar* (ca. 1870 B.C. – ca. 1860 B.C.). Later the Akkadians, the successors of Sumerians, continued this Sumerian legal tradition and established some own codes, such as *Hammurabi* (1728 – 1686 B.C.), see Figure 1.

There were excavated many trial records of the Neo-Babylonian period, i.e. of the time beginning with the rise of the Babylonian king Nabopolassar in 626 B.C. and lasting until the end of the reign of Nabonidus because of the Achaemenid conquest in 539 B.C. There are also many trial records thereafter, but they were written within the same unchanged cuneiform textual tradition of Neo-Babylonia [11], [12]. Therefore they can be called Neo-Babylonian, too. All these trial records are especially interesting from the point of view of symbolic logic. The matter is that these texts demonstrate a high logical competence of their authors. Furthermore, we can claim that these documents obviously show an axiomatization of legal proceedings. The thing is that Neo-Babylonian trial records are well preserved and many of them are large enough to be reconstructed logically.

In the Neo-Babylonian trial records, suitable articles from law codes are assumed to be axioms. Facts established by the court give another set of axioms. And according to some logical inference rules, just automatically the court makes an appropriate decision from axioms, i.e. from facts and articles of law codes, in a logically correct way. We can assume that the same situation was before the Neo-Babylonian period, too: (i) the law codes of different periods have had many unchanged articles and the *Hammurabi Code* (see Fig.1) can be examined as a significant early sample of all these legal codifications; (ii) the court procedure has remained almost the same since the very beginning; (iii) there are many evidences that the law codes were considered as a set of axioms for logical inferring, e.g. in some fragments of trial and business documents of Old-Assyrian period (that took place from 2025 B.C. to 1378 B.C.) we find out an expression "the words of the stele" (*awāt naruāim*), see for the details in [28], with the meaning to be a reference to an axiom for inferring from it.

Each article from codes is formulated as follows: "*if* (*šumma*) you do such-and-such, *then* a trial decision will be as follows". So, on the one hand, in the antecedent of implication we deal with some combinations of illegal and legal actions conducted

that we need "a broader inquiry into the relationship between written legislation and legal decision making". On my part, I can add that it is better to organize a joint project uniting logicians and Assyriologists to study this (proto)legislation in Mesopotamia. For instance, I am sure that a (proto)logical textbook can exist in fragments among commentary texts. But to see this fact we need a cooperation of logicians and Assyriologists.

Figure 1: The stele of the *Law Code of Hammurabi*, Louvre Museum; *by courtesy of Vladimir Sazonov.*

before the court: “steal”, “possess”, “strike”, “causes the boat to sink”, etc. and, on the other hand, in the succedent of implication we deal with some legal consequences established by the court: “compensate”, “kill”, etc. There are the following main classes of people in the *Law Code of Hammurabi* [19]: (i) "man" or "free person" (*awīlu*), to denote men, women, and minors; (ii) "commoner" (*muškēnu*); (iii) "male slave" (*wardu*) and "female slave" (*amtu*). Some classes to denote familial relationships: (i) "son" (*māru*); (ii) "wife" (*aššatu*); (iii) "first-ranking wife" (*ḫīrtu*); (iv) "widow" (*almattu*), etc. Also, there are some abstract entities such as "god" (i.e. a temple as legal person) or "palace" (i.e. a king as legal person) to denote a form of possession. And some professional groups: "boatman"; "merchant", etc.

Let us formalize the articles of the law in the following way. Suppose, small letters a, b, c, …, x, y, z, … are variables to denote things. Capital letters are to denote classes of natural and legal persons: A (*awīlu*-class); M (*muškēnu*-class); T (temple or palace as legal persons), B (boatmen), etc. The expression Ax means that y belongs to the *awīlu*-class, My means that y belongs to the *muškēnu*-class, and Tz means that z belongs to the category of legal persons.

First-order formulas are defined in the following standard manner: (i) each a, b, c, …, x, y, z, … is a first-order formula; (ii) each Ax, My, Tz, … is a first-order formula; (iii) each Boolean combination of formulas defined in (i) and (ii) is a first-order formula (Boolean connectives: $\neg$ negation, & conjunction, $\vee$ disjunction, $\Rightarrow$ implication, $\otimes$ strong disjunction).

Let an action "verb" in the antecedent or succedent of articles of the law be denoted by [[verb]]. Each action is an n-place relation among people of different classes and things. For example: [[steal (Ax, z)]] means that x of the *awīlu*-class steals z; [[strike (Ax, Ay)]] means that x of the *awīlu*-class strikes y of the same class; [[boat-to-sink (Bx, y)]] means that a boatman x causes the boat y to sink; [[possess (Ty, z)]] means that y is a legal person who possesses z; [[compensate (Ax, Ty, z)]] means that x of the *awīlu*-class must compensate the legal person y by paying z; [[killed (Ax)]] means that x of the *awīlu*-class is sentenced to be killed, etc.

Second-order formulas are defined thus: (i) each [[steal (Ax, z)]], [[strike (Ax, Ay)]], [[boat-to-sink (Bx, y)]], [[possess (Ty, z)]], [[compensate (Ax, Ty, z)]], [[killed (Ax)]], … is a second-order formula; (ii) each Boolean combination of formulas defined in (i) is a second-order formula (Boolean connectives: negation, & conjunction, $\vee$ disjunction, $\Rightarrow$ implication, $\otimes$ strong disjunction).

Now, let us consider some articles from codes as axioms for legal proceedings. Some examples:

A1

> If a man x of the *awīlu*-class steals a thing z and it belongs to a temple or king y, then x must replace z thirtyfold. If z belongs to a commoner y, he must replace it tenfold. If the thief does not have anything to give,

he shall be killed:

$(([[\text{steal }(Ax,\ z)]]\ \&\ [[\text{possess }(Ty,\ z)]]) \Rightarrow [[\text{compensate }(Ax,\ Ty,\ z \cdot 30)]])\ \&\ (\ [[\text{compensate }(Ax,\ Ty,\ z \cdot 30)]] \Rightarrow [[\text{killed }(Ax)]])$

&

$(([[\text{steal }(Ax,\ z)]]\ \&\ [[\text{possess }(Cy,\ z)]]) \Rightarrow [[\text{compensate }(Ax,\ Cy,\ z \cdot 10)]])\ \&\ (\ [[\text{compensate }(Ax,\ Cy,\ z \cdot 10)]] \Rightarrow [[\text{killed }(Ax)]])$

In the *Hammurabi Code*:

(vi 57-69) *šumma awīlum lu alpam lu immeram lu imēram lu šaḫâm ulu elippam išriq šumma ša ilim šumma ša ekallim adi 30-šu inaddin šumma ša muškēnim adi 10-šu iriab šumma šarrâqânum sa nadânim la išu iddâk*

§8 If a man steals an ox, a sheep, a donkey, a pig, or a boat – if it belongs either to the god or to the palace, he shall give thirtyfold; if it belongs to a commoner, he shall replace it tenfold; if the thief does not have anything to give, he shall be killed [19, p. 82].

A2

If a man x of the *awīlu*-class strikes another man, he must pay 60 shekels of silver:

$[[\text{strike }(Ax,\ Ay)]] \Rightarrow [[\text{compensate }(Ax,\ Ay,\ 60\text{ shekels})]]$

In the *Hammurabi Code*:

(xl 82-87) *šumma mār awīlim lēt mār awīlim ša kīma šuāti imtaḫaṣ 1 mana kaspam išaqqal*

§203 If a member of the *awīlu*-class should strike the cheek of another member of the *awīlu*-class who is his equal, he shall weigh and deliver 60 shekels of silver [19, p. 121]

A3

If a boatman x causes the boat y to sink, he must compensate all things z contained there:

([[boat-to-sink (Bx, y)]] & [[contain (y, z)]] & [[possess (Ay, z)]]) $\Rightarrow$ [[compensate (Bx, Ay, z)]]

In the *Laws of Ešnunna* (ca. 1790 B.C.):

(A i 25-26) *šumma malaḫḫum īgīma elippam uṭṭebbe mala uṭebbû umalla*

§5 If the boatman is negligent and causes the boat to sink, he shall restore as much as he caused to sink [19, p. 60]

Hence, all the articles are formulated as a set of implications connected by conjunctions: $(A_1 \Rightarrow B_1)\&(A_2 \Rightarrow B_2)\&\ldots\&(A_n \Rightarrow B_n)$. In the meanwhile, the antecedents $A_1, A_2, \ldots, A_n$ of these implications are understood as particulars and their succedents $B_1, B_2, \ldots, B_n$ as appropriate generals.

One of the main features of law codes explicitly stated first by Cicero in his *Topica* [8] (the work devoted to legal hermeneutics in the Stoic/Sumerian-Akkadian way) is a full enumeration of particulars $A_1, A_2, \ldots, A_n$, related to one general B, i.e. a full list of implications $A_1 \Rightarrow B, A_2 \Rightarrow B, \ldots, A_n \Rightarrow B$ with the same B, as it was supposed in any law code since *Hammurabi*. These $A_1, A_2, \ldots, A_n$ should be exclusive thereby. It means that they should be connected by strong disjunctions "either ... or ... " (Akkadian: "*ūl* ... *ūl*...", symbolically: "$\ldots \otimes$ ..."). In this case there is the following equivalence: $(A_1 \otimes A_2 \otimes \cdots \otimes A_n) \Leftrightarrow B$. From this we can draw the following conclusion:

$$\frac{\dfrac{A_1 \Rightarrow B; A_2 \Rightarrow B; \ldots; A_n \Rightarrow B; C \Rightarrow \neg A_1; C \Rightarrow \neg A_2; \ldots; C \Rightarrow \neg A_n}{(A_1 \otimes A_2 \otimes \cdots \otimes A_n) \Rightarrow B; \quad C \Rightarrow \neg(A_1 \otimes A_2 \otimes \cdots \otimes A_n)}}{C \Rightarrow \neg B.}$$

Cicero formulates this rule thus:

Next, the enumeration of the parts (sc. of the whole), which is handled in the following way: If someone has not been freed by either having his name entered in the census-roll or by being touched with the rod or by a provision in a will, then he is not free. None of these applies to the individual in question. Therefore he is not free [8, p. 121].

This logical rule implemented in any code may be named a *completeness of legal information*. This completeness means that if we take any factual verified case C of an indictment, then for any general B from the code, each court can announce either a verdict $C \Rightarrow B$ or a verdict $C \Rightarrow \neg B$ inferred from the code just logically.

Each court has considered only cases supported by documents or couples of witnesses. Let us exemplify this feature. In one trial record, $^{\mathrm{m}}$Rēmanni-Bēl son of $^{\mathrm{m}}$Tērik-šarrūssu has claimed that his sister $^{\mathrm{f}}$Bābunu and her children are a part of household of $^{\mathrm{m}}$Nabû-mukīn-apli son of $^{\mathrm{m}}$Amurru-šuma-iddinam wrongly. Nevertheless, $^{\mathrm{m}}$Rēmanni-Bēl cannot have supported his claim by any document:

1. *di-i-ni ša*$_2$ $^{\mathrm{m}}$*re-man-ni-*$^{\mathrm{d}}$EN A-*šu*$_2$ *ša*$_2$ $^{\mathrm{m}}$*te-rik*-LUGAL-*ut-su*
2. *a-na muḫ-ḫi* $^{\mathrm{f}}$*ba-bu-nu u*$_3$ DUMU. MEŠ-*šu*$_2$ UN.MEŠ E$_2$
3. *ša*$_2$ $^{\mathrm{md}}$NA$_3$-<*mu*>-*ki-in*-IBILA DUMU-*šu*$_2$ *ša*$_2$ $^{\mathrm{md}}$KUR.GALMU-*id-di-nam*
4. *it-ti* $^{\mathrm{md}}$NA$_3$-DU-IBILA *a-na ma-ḫar* $^{\mathrm{lu2}}$DI.KU$_5$.MEŠ
5. *ša*$_2$ $^{\mathrm{md}}$NA$_3$-*na-'-id* LUGAL TIN. TIR^{ki} *id-bu-bu um-ma* $^{\mathrm{f}}$*ba-bu-nu*
6. *ša*$_2$ *i-na* E$_2$-*ku-nu*^{2} *a-ḫa-ta-a ši-i* $^{\mathrm{lu2}}$DI.KU$_5$.MEŠ
7. $^{\mathrm{m}}$*re-man-ni-*$^{\mathrm{d}}$EN *iš-ta-'-a-lu um-ma* $^{\mathrm{f}}$*ba-bu-nu*
8. NIN-*ka ul-tu im-ma-ti ki-i* E$_2$$^{\mathrm{md}}$KUR.GAL-MU-MU
9. AD *ša*$_2$ $^{\mathrm{md}}$NA$_3$- -DU-IBILA *ši-i* $^{\mathrm{m}}$*re-man-ni-*$^{\mathrm{d}}$EN *iq-bi*
10. *um-ma* 40 MU.AN.NA.MEŠ *an-na-a-ti* $^{\mathrm{f}}$*ba-bu-nu*
11. NIN-*a* $^{\mathrm{md}}$KUR.GAL-MU-MU *ta-pal-laḫ di-i-ni a-na muḫ-ḫi-šu*$_2$
12. *it-ti* $^{\mathrm{md}}$KUR.GAL-MU-MU AD *ša*$_2$ $^{\mathrm{md}}$NA$_3$-DU-IBILA *ad-di-bu-ub*
13. *u*$_3$ *a-di i-na-an-na iš-tu* E$_2$-*šu*$_2$ *la u*$_2$-*še-ṣi-iš*
14. $^{\mathrm{m}}$*re-ma-an-ni-*$^{\mathrm{d}}$EN *mim-ma i-da-tu ša*$_2$ *di-i-ni a-na* UGU
15. $^{\mathrm{f}}$*ba-bu-nu it-ti* $^{\mathrm{md}}$KUR.GALMU-MU AD *ša*$_2$ $^{\mathrm{md}}$NA$_3$-DU-A
16. *id-bu-bu a-na* $^{\mathrm{lu2}}$DI.KU$_5$.MEŠ *la u*$_2$-*kal-li-im*
17. $^{\mathrm{lu2}}$DI.KU$_5$.MEŠ *dib-bi-šu*$_2$-*nu-ti iš-mu-ma mim-ma i-da-tu*$_4$
18. *ša*$_2$ *di-i-ni la i-mu-ru-u iš-ta-lumu* 40 MU.AN.NA.MEŠ
19. *an-na-a-ti* $^{\mathrm{f}}$*ba-bu-nu* $^{\mathrm{md}}$KUR. GAL-MU-MU AD *ša*$_2$ $^{\mathrm{md}}$NA$_3$DU-IBILA
20. *tu*$_3$-*pal-laḫ man-ma di-i-ni u*$_3$ *paqa-ri*
21. *ina muḫ-ḫi-šu*$_2$ *la ir-ši*
22. $^{\mathrm{f}}$*ba-bu-nu u*$_3$ DUMU.MEŠ-*šu*$_2$ UN.MEŠ E$_2$ *ša*$_2$ $^{\mathrm{md}}$KUR.GALMU-MU
23. *a-na* $^{\mathrm{md}}$NA$_3$-DU-IBILA *id-di-nu*

(1–5) The case which $^{\mathrm{m}}$Rēmanni-Bēl son of $^{\mathrm{m}}$Tērik-šarrūssu argued against $^{\mathrm{m}}$Nabû-mukīn-apli, regarding $^{\mathrm{f}}$Bābunu and her children, members of the household of $^{\mathrm{m}}$Nabû-mukīn-apli son of $^{\mathrm{m}}$Amurru-šuma-iddinam, before the judges of Nabonidus, king of Babylon, thus:

(5–6) "$^{\mathrm{f}}$Bābunu, who is in your household, is my sister!"

(6–7) The judges questioned $^{\mathrm{m}}$Rēmanni-Bēl thus:

(7–9) "Since when has $^{\mathrm{f}}$Bābunu, your sister, been part of the household of $^{\mathrm{m}}$Amurru-šuma-iddinam, father of $^{\mathrm{m}}$Nabû-mukīn-apli?"

(9–10) $^{\mathrm{m}}$Rēmanni-Bēl said thus:

(10–13) "For these past 40 years, $^{\mathrm{f}}$Bābunu, my sister has served $^{\mathrm{m}}$Amurru-šuma-iddinam. I argued a case regarding her against $^{\mathrm{m}}$Amurru-šuma-iddinam, father of $^{\mathrm{m}}$Nabû-mukīn-apli, but he has not let her go from his household until now!"

(14–16) $^{\mathrm{m}}$Rēmanni-Bēl did not show the judges any evidence of the case regarding $^{\mathrm{f}}$Bābunu which he argued against $^{\mathrm{m}}$Amurru-šuma-iddinam, father of $^{\mathrm{m}}$Nabû-mukīn-apli.

(17–18) The judges heard their arguments. They did not see any evidence of the case. They conferred.

(18–21) For these 40 years, $^{\mathrm{f}}$Bābunu served $^{\mathrm{m}}$Amurru-šuma-iddinam, father of $^{\mathrm{m}}$Bēl-mukīn-apli. He did not have any case or claimant against him.

(22–23) They assigned $^{\mathrm{f}}$Bābunu and her children, the members of the household of $^{\mathrm{m}}$Amurru-šuma-iddinam to $^{\mathrm{m}}$Nabû-mukīn-apli [12, p. 23-24].

Symbolically, we apply *modus tollens*:

1. If a man brings a claim to present a case, he must support his claim about the case by documents or witnesses [the axiom supposed in any code];

2. There are no documents and witnesses for this case [a fact proved by the court];

Then, the case is not valid to be approved.

The case of the claim, C, is beyond all the n cases A_1, A_2, $\ldots$, A_n given in a code (killing, stealing, etc.), which are usually considered by the court, i.e. we have: $C \Rightarrow \neg(A_1 \otimes A_2 \otimes \cdots \otimes A_n)$, where each case A_i is supported by appropriate documents and couples of witnesses. Then this case is out of any legal proceeding, B, i.e.: $C \Rightarrow \neg B$. In other words, if someone is C, he is guilty either in A_1 (killing) or in A_2 (stealing) or ... in A_n. Formally: $C \Rightarrow (A_1 \otimes A_2 \otimes \cdots \otimes A_n)$. The court has considered all documents and witnesses in respect to the defendant and the case is not A_1 (killing) and it is not A_2 (stealing) and ... it is not A_n: $(\neg A_1 \wedge \neg A_2 \wedge \cdots \wedge \neg A_n)$. This expression is the same as $\neg(A_1 \vee A_2 \vee \cdots \vee A_n)$. In turn, from the latter formula it follows that $\neg(A_1 \otimes A_2 \otimes \cdots \otimes A_n)$, because there is a tautology: $\neg(A_1 \vee A_2 \vee \cdots \vee A_n) \Rightarrow \neg(A_1 \otimes A_2 \otimes \cdots \otimes A_n)$. Then, by *modus tollens* we deduce that he is not C (he is not guilty, i.e. not defendant):

$$C \Rightarrow (A_1 \otimes A_2 \otimes \cdots \otimes A_n);$$

$$\neg(A_1 \otimes A_2 \otimes \cdots \otimes A_n);$$

$$\overline{\neg C}$$

Hence, not only articles of the law, such as **A1**, **A2**, and **A3**, were examined as axioms for legal proceedings, but also facts established by the court by appealing to signed documents or presented witnesses. Let us show how a document has supported a view of defendant partly. There is a signed document that $^{\mathrm{m}}$Iddin-Marduk possesses 480 *gur* of dates and the boatman $^{\mathrm{m}}$Amurru-natan son of $^{\mathrm{m}}$Ammaya is responsible to deliver them safely. Nevertheless, $^{\mathrm{m}}$Nergal-rēṣūa the slave of $^{\mathrm{m}}$Iddin-Marduk testimonies that 47 *gur* 1 *pi* are missing. But there is a contract that the boatman $^{\mathrm{m}}$Amurru-natan gave back 7 *gur* 1 *pi* of dates from missing items. The trial decision is that $^{\mathrm{m}}$Amurru-natan must pay 40 *gur* of dates, the missing amount of those dates, and assigned them to $^{\mathrm{m}}$Iddin-Marduk:

1. $^{\mathrm{md}}$U.GUR-*re-ṣu-u*$_2$*-a* $^{\mathrm{lu2}}$*qal-la ša*$_2$ $^{\mathrm{md}}$MU-$^{\mathrm{d}}$AMAR.UTU
2. *a-na* $^{\mathrm{lu2}}$DI.KU$_5$.MEŠ *ša*$_2$ $^{\mathrm{md}}$NA$_3$IM.TUK LUGAL TIN.TIR$^{\mathrm{ki}}$
3. *iq-bi um-ma*
3. $^{\mathrm{m}}$MU-$^{\mathrm{d}}$AMAR.UTU EN-*a*
4. 4 ME 80 GUR ZU$_2$.LUM.MA *e-pi-ru-tu*
5. *ul-tu* EDIN *a-na* $^{\mathrm{giš}}$MA$_2$.MEŠ *ša*$_2$ $^{\mathrm{md}}$KUR.GAL-*na-tan*
6. $^{\mathrm{lu2}}$MA$_2$.LAḪ5 A-*šu*$_2$ *ša*$_2$ $^{\mathrm{m}}$*am-ma-a u*$_2$*-še-li-ma*
7. *pu-ut* EN.NUN-*tim ša*$_2$ ZU$_2$. LUM.MA *u*$_2$*-ša*$_2$*-aš*$_2$*-ši-iš*
8. $^{\mathrm{giš}}$MA$_2$.MEŠ *a-na* TIN.TIR$^{\mathrm{ki}}$ *u*$_2$*-še-la-am-ma*
9. *ši-pir-tu*$_4$ *ša*$_2$ $^{\mathrm{m}}$MU-$^{\mathrm{d}}$AMAR. UTU *id-di-nam-ma*
10. 4 ME 80 GUR ZU$_2$.LUM. MA *ina lib*$_3$-[*bi-šu*$_2$] *ša*$_2$*-ṭi-ir*
11. *re-eš* ZU$_2$.LUM.MA *aš*$_2$*-ši-ma* 47 GUR 1 PI

12. *ina lib*$_3$*-bi ma-ṭu-*⌈*u*$_2$⌉ *a-na* UGU
13. *mi-ṭi-tu*$_4$ *ša*$_2$ ZU$_2$.LUM.MA *it-ti* ^{md}KUR.GAL-*na-tan*
14. *ar-gum*$_2$*-ma u*$_2$-ŠAR-X-RI *um-ma* ZU$_2$.⌈LUM.MA⌉-*ka*
15. *ul aš*$_2$*-ši ar*$_2$?*-ki ba-ti-qu* X X X ...
16. 4! GUR 1 PI ⌈ZU$_2$⌉.[LUM. MA]
17. *u*$_3$ *ku-tal-la ša*$_2$ ⌈^{giš}MA$_2$-*ni*⌉ X-X *u*$_2$
18. ZU$_2$.LUM.MA *šu*$_2$*-nu-tu*$_2$ *i-na* X-*šu*$_2$-[
19. *rik-su it-ti-šu*$_2$ *ni-iš-ku-us*
20. *um-ma* 7 GUR 1 PI ZU$_2$.LUM. MA
21. ^{md}KUR.GAL-*na-tan ina sar-tu iš-šu-u*$_2$
22. *ar*$_2$*-ki ri-ik-su šu-a-tu*$_2$ ^{md}KUR. GAL-[*na-tan*]
23. *šuṭ-ur-ma a-di u*$_4$*-mu an-ni-i* X
24. *i-na-an-na i-na maẖ-ri-ku-nu ub-la-aš*$_2$
25. EŠ.BAR-*a-ni šuk-na*
25. ^{lu2}DI.KU$_5$.MEŠ *dib-bi-šu*$_2$*-nu*
26. *iš-mu-u*$_2$ *rik-su šu-a-tu*$_2$ *u ši-pir-tu*$_4$
27. *ša*$_2$ ^{m}MU-^{d}AMAR.UTU *ša*$_2$ 4 ME 80 GUR ZU$_2$.LUM.MA
28. *ina lib*$_3$*-bi šaṭ-ru ša*$_2$ ^{md}U.GUR*re-ṣu-u*$_2$*-a ub-la*
29. *ma-ẖar-šu-nu iš-tas-su-u*$_2$ ^{md}KUR.GAL-*na-tan*
30. *i-ša*$_2$*-lu-ma na-šu-u*$_2$ *ša*$_2$ ZU$_2$. LUM.MA *ša*$_2$ *ina sar-tu*$_4$
31. *na-šu-u*$_2$ *e-li ra-ma-ni-šu*$_2$ *u*$_2$*-kin-ma*
32. 40 GUR ZU$_2$.LUM.MA *mi-ṭi-tu*$_4$ *ša*$_2$ ZU$_2$.LUM.MA *šu*$_2$*-nu-šu*$_2$
33. *e-li* [^{m}]^{d}KUR.GAL-*na-tan ip-rusu-ma*
34. *a-na* ^{md}U.GUR-*re-ṣu-u*$_2$*-a* ^{lu2}[*qal-la ša*$_2$] ^{m}MU-^{d}AMAR.UTU
35. *id-di-nu ina* EŠ.BAR ⌈*di-i-ni*⌉ *šu-a-tim*
36. ^{md}U.GUR-[GI ^{lu2}DI.KU$_5$] DUMU *ši-gu-u*$_2$*-a*
37. ^{md}NA$_3$-ŠEŠ.MEŠ-MU ^{lu2}DI. KU$_5$ [DUMU]⌈*e-gi-bi*⌉
38. ^{md}NA$_3$-[MU-GI].NA ^{lu2}DI.KU$_5$ DUMU *ir-a-*[*ni*]
39. ^{md}EN-[ŠEŠ.MEŠ]-⌈MU⌉ ^{lu2}DI.KU$_5$ DUMU ^{md}ZALAG^{d}30
40. ^{md}EN-[KAR]-⌈*ir*⌉ ^{lu2}DI.KU$_5$ DUMU ^{md}30-*tab-ni*
41. ^{md}NA$_3$-MU-GAR-*un* DUB. SAR DUMU ^{lu2}GAL-DU$_3$
42. TIN.TIR^{ki} ITI ŠE U$_4$ 4-*kam*$_2$
43. MU 10-*kam*$_2$ ^{md}NA$_3$-IM.TUK LUGAL TIN.TIR^{ki}

^{na4}KIŠIB ^{md}U.GUR-GI ^{lu2}DI.KU$_5$
^{na4}KIŠIB ^{md}NA$_3$-ŠEŠ.MEŠ-MU [^{lu2}DI].KU$_5$
^{na4}KIŠIB ^{md}NA$_3$-MU-GI.NA ^{lu2}DI. KU$_5$
^{na4}KIŠIB ^{md}EN-ŠEŠ.MEŠ-MU ^{lu2}DI.KU$_5$
^{na4}[KIŠIB] ^{md}EN-KAR-[*ir*] ^{lu2}DI.KU$_5$

(1–3) ^{m}Nergal-rēṣūa the slave of ^{m}Iddin-Marduk said thus to the judges of Nabonidus, king of Babylon:

(3–6) "^{m}Iddin-Marduk, my master, loaded a shipment of 480 *kur* of dates for transport (?) from the hinterland on the boats belonging to ^{m}Amurru-natan, the boatman, son of ^{m}Ammaya."

(7) "He had him bear the responsibility for keeping the dates."

(8–10) "He brought the boats to Babylon and he gave me ^{m}Iddin-Marduk's message. 480 Gur of dates was written i[n it]."

(11–12) "I took account of the dates, and 47 *gur* 1 *pi* were missing."

(12–14) I raised a claim against ^{m}Amurru-natan concerning the missing amount of the dates and . . . thus:

(14–15) " 'I did not take your dates."'

(15) "Afterwards, an informer . . .

(16) " '4 Gur 1 Pi of dates . . .

(17) " 'and behind my boat . . .

(18) " 'those dates in . . .

(19–20) "We contracted a contract stating thus: '^{m}Amurru-natan illegally took 7 *gur* 1 *pi* of dates.' "

(22–23) "After ^{m}Amurru-[natan] wrote this contract until today . . .

(24) Now, I have brought him before you."

(25) "Establish our decision!"

(25–26) The judges heard their arguments.

(26–29) They read before them that contract and ^{m}Iddin-Marduk's message in which 480 Gur of dates was written which ^{m}Nergalrēṣūa brought.

(29–30) They questioned ^{m}Amurru-natan.

(30–31) (Regarding) the taking of the dates, he established about himself that they were taken illegally.

(32–35) They decided that ^{m}Amurru-natan must pay 40 *gur* of dates, the missing amount of those dates, and assigned them to ^{m}Iddin-Marduk, [slave] of ^{m}Nergal-rēṣua.

(35) At the decision of this case:

(36) ^{m}Nergal-[ušallim, the judge,] descendant of Šigûa;

(37) ^{m}Nabû-aẖẖē-iddin, the judge, [descendant of] Egibi;

(38) ^{m}Nabû-[šuma-ukī]n, the judge, descendant of Ir'an[ni];

(39) ^{m}Bēl-[aē]-iddin, the judge, descendant of Nūr-Sîn;

(40) ^{m}Bēl-ēṭir, the judge, descendant of Sîn-tabni;

[Scribe:]

(41) ^{m}Nabû-šuma-iškun, the scribe, descendant of Rāb-bānê.

[Date:]
(42–43) Babylon. 4 Addaru, year 10 of Nabonidus, king of Babylon.
[Seals of authorities]
[Left edge:]
Seal of $^{\mathrm{m}}$Nergal-ušallim, the judge;
Seal of $^{\mathrm{m}}$Nabû-aẖẖē-iddin [the jud]ge;
Seal of $^{\mathrm{m}}$Nabû-šuma-ukīn, the judge;
[Right edge:]
Seal of $^{\mathrm{m}}$Bēl-aẖẖē-iddin, the judge;
[Seal] of $^{\mathrm{m}}$Bēl-ēṭ[ir], the judge [12, p. 28-32]

1. If a boatman has a contract to deliver some items safely, he is responsible for them and must repay them if they are missing [the axiom from the code, compare to **A3**];

2. The boatman $^{\mathrm{m}}$Amurru-natan son of $^{\mathrm{m}}$Ammaya has a contract to deliver 480 *gur* of dates belonging to $^{\mathrm{m}}$Iddin-Marduk [the fact established by the trial];

3. $^{\mathrm{m}}$Nergal-rēṣūa the slave of $^{\mathrm{m}}$Iddin-Marduk testimonies that 47 *gur* 1 *pi* are missing [the claim];

4. There is a contract that the boatman $^{\mathrm{m}}$Amurru-natan gave back 7 *gur* 1 *pi* of dates from missing items [the fact established by the trial];

Then, $^{\mathrm{m}}$Amurru-natan son of $^{\mathrm{m}}$Ammaya must repay 40 *gur* of dates to $^{\mathrm{m}}$Iddin-Marduk on 4 Addaru, year 10 of Nabonidus, king of Babylon.

This conclusion is perfect logically and we can even formalize it within an axiomatic system for the Neo-Babylonian legal proceedings (where articles from a code and facts confirmed by a trial are axioms and propositional logic is to deduce from axioms some new conclusions). Assume that we have axioms of the following two types: (i) the articles of the law formulated *explicitly*, as well as the directly derivated propositions from the articles by using logical inference rules (the articles of the law formulated *implicitly*); (ii) the facts as signed documents or sworn testimonies checked by the court.

In this axiomatic system we apply some inference rules. The two basic rules for inferring from axioms A, B, $A \Rightarrow B$ are as follows:

Modus ponens (**MP**):

$$\frac{A; \qquad A \Rightarrow B}{B}$$

Modus tollens (**MT**):

$$\frac{\neg B; \qquad A \Rightarrow B}{\neg A}$$

Some other inference rules:

Introducing conjunction (**I&**):

$$\frac{A; \qquad B}{A \& B}$$

Deleting conjunction (**D&**):

$$\frac{A \& B}{A; \qquad B}$$

Introducing disjunction (**I**$\vee$):

$$\frac{A; \qquad B}{A \vee B}$$

Let S[B] mean that B is a variable a, b, c, $\ldots$, x, y, z, $\ldots$ or Ax, My, Tz, $\ldots$ belonging to a formula S. Let A be a constant. An additional inference rule:

Substitution rule (**SR**):

$$\frac{\mathrm{S}[B]; \qquad A \Rightarrow B}{\mathrm{S}[A]}$$

In other words, if $A \Rightarrow B$ holds, then we can substitute A for any occurrence of B in a formula S. For example, in expressions such as ([[steal (Ax, z)]] & [[possess (Ty, z)]]) $\Rightarrow$ [[compensate (Ax, Ty, $z \cdot 30$)]] from a code, we can replace variables by constants: Ax by '$^{\mathrm{m}}$Aḫa-iddin', Ty by 'the Lady-of-Uruk and Nanaya', and z by 'ducks', etc. if there are implications: '$^{\mathrm{m}}$Aḫa-iddin' $\Rightarrow Ax$, 'the Lady-of-Uruk and Nanaya' $\Rightarrow Ty$, and 'ducks' $\Rightarrow z$, validated by a court.

Inference rules **MP**, **MT**, **I&**, **D&**, **I**$\vee$, **SR** are enough for inferring all the trial decisions from the code and facts checked by the court, but in order to reduce a length of deduction we can introduce some new inference rules of the form:

$$\frac{A; \qquad B}{C}$$

if and only if the formula $(A \& B) \Rightarrow C$ is a tautology of propositional logic.

5 Automatic Logical Conclusions Drawn in Neo-Babylonian Trial Records

Each excavated Neo-Babylonian trial record can be examined as a conclusion drawn from axioms automatically. Let us consider some examples. One of them is denoted by YBC 3771, it was found in Uruk, and dated to 12.XII.3 of Cambyses (22 March, 526 B.C.), see [11, p. 178-181]. In this trial record, two judges determine that Bēl-iqīša, who led away 5 sheep belonging to 'Ištar of Uruk and Nanaya' (a temple), must repay 155 sheep to the property of this temple, because 150 sheep is the **thirtyfold** penalty for five branded sheep and the five unbranded lambs are supposed to be born after stealing:

1. [1-*en* UDU *pu-ḫal* 4 UDU U_8.MEŠ] NIGIN 5 *ṣe-e-nu šá* MUL-tu_4 *še-en-du*
2. ⌈ù⌉ [5 *par*]-*rat*.ME *ta-mi-ma-a-ta* NIGIN 10 *ṣe-e-nu*
3. NIG_2.GA $^{\mathrm{d}}$INNIN UNUG$^{\mathrm{ki}}$ *u dna-na-a šá qa-pu-ut-*tu_4
4. *šá mda-nu*-LUGAL-URI_3 DUMU-*šú šá* $^{\mathrm{md}}$LUGAL-DU *šá ina* ITI APIN MU 2-*kám*
5. $^{\mathrm{m}}$*ka-am-bu-zi-ia* LUGAL TIN. TIR$^{\mathrm{ki}}$ LUGAL KUR.KUR $^{\mathrm{md}}$ ENBA-*šá*
6. DUMU-*šú šá* $^{\mathrm{m}}$*ṣil-la-a ina* ŠU.2 $^{\mathrm{md}}$a-nu-LUGAL-URI_3 A-*šú šá* $^{\mathrm{m}}$LUGAL-DU *i-bu-ku-ma*
7. *ina* ITI ŠE MU 3-*kám mri-mut u mdba-ú*-APIN-*eš*

8. $^{\mathrm{lu2}}$DI.KU$_5$.ME 150 *ṣe-e-nu ku-um ṣe-e-nu šá* $^{\mathrm{d}}$15

9. *šen-de-e-ti* 1-*en a-di* 30 *ù* 5 *parrat ta-mi-ma-a-ta*

10. NIGIN 155 *ṣe-e-nu a-na e-ṭe*$_3$*-ru šá* $^{\mathrm{d}}$INNIN UNUG$^{\mathrm{ki}}$

11. *i-na ṭup-pi iš-ṭu-ru-ma e-li* $^{\mathrm{md}}$ENBA-*šá ú-kin-nu*

12. U$_4$ 25-*kám šá* ITI ŠE MU 3-*kám ṣe-e-nu a'* 155 $^{\mathrm{md}}$EN-BA-*šá*

13. DUMU-*šú šá* $^{\mathrm{m}}$*ṣil-la-a ibba-kám-ma ina* E$_2$.AN.NA *i-šim-mi-it-ma*

14. *a-na* NIG$_2$.GA E$_2$.AN.NA *i-namdin* $^{\mathrm{m}}$IR$_3$-$^{\mathrm{d}}$U.GUR DUMU-*šú šá* $^{\mathrm{m}}$DU-A

15. DUMU $^{\mathrm{m}}$*e-gi-bi pu-ut e-ṭe*$_3$*-ru šá ṣe-e-nu-a'*

16. 155 *na-ši-i i-na ú-šu-uz-zu šá* $^{\mathrm{md}}$NA$_3$-DU-⌈IBILA⌉

17. $^{\mathrm{lu2}}$ŠA$_3$.TAM E$_2$.AN.NA DUMU-*šú šá* $^{\mathrm{m}}$*na-di-nu* DUMU $^{\mathrm{m}}$*da-bi-bi*

18. $^{\mathrm{md}}$NA$_3$-ŠEŠ-MU $^{\mathrm{lu2}}$SAG-LUGAL $^{\mathrm{lu2}}$EN *pi-qit-ti* E$_2$.AN.NA

19. $^{\mathrm{lu2}}$*mu-kin-nu* $^{\mathrm{m}}$IR$_3$-$^{\mathrm{d}}$AMAR.UTU DUMU-*šú šá* $^{\mathrm{m}}$NUMUN-*ia* DUMU $^{\mathrm{m}}$*e-gi-bi*

20. $^{\mathrm{md}}$30-APIN-*eš* DUMU-*šú šá* $^{\mathrm{md}}$NA$_3$-MU-SI.SA$_2$ DUMU $^{\mathrm{m}}$DU$_3$-DINGIR

21. $^{\mathrm{md}}$EN-SUM-IBILA DUMU-*šú šá* $^{\mathrm{md}}$AMAR.UTU-MU-MU DUMU $^{\mathrm{md}}$EN-IBILA-URI$_3$

22. $^{\mathrm{m}}$*na-di-nu* DUB.SAR DUMU $^{\mathrm{m}}$*e-gi-bi*

23. $^{\mathrm{m}}$IR$_3$-$^{\mathrm{d}}$[AMAR.UTU] DUB.SAR DUMU $^{\mathrm{md}}$EN-IBILA-URI$_3$

24. UNUG$^{\mathrm{ki}}$ ITI ŠE U$_4$ 12-*kám* MU 3-*kám* $^{\mathrm{m}}$*kám-bu-zi-ia*

25. LUGAL TIN.TIR$^{\mathrm{ki}}$ LUGAL KUR.KUR

(1–6) [1 ram 4 ewes] total 5 sheep branded with a star and 5 unblemished lambs, a total of 10 sheep, property of Ištar of Uruk and Nanaya, from the pen of Anu-šarra-uṣur son of Šarrukīn, which in Araḫšamna, year 2 of Cambyses, king of Babylon, king of the lands, Bēl-iqīša son of Ṣillaya led away (in payment) from Anu-šarra-uṣur son of Šarrukīn.

(7–11) In Addaru, year 3, Rīmūt and Bau-ēreš, the judges, wrote in a tablet and determined for Bēl-iqīša to pay 150 sheep, **thirtyfold** for the sheep branded for Ištar and 5 unbranded lambs, a total of 155 sheep, for repayment to Ištar of Uruk.

(12–14) On 25 Addaru, year 3, Bēliqīša son of Ṣillaya shall bring these 155 sheep, brand them in the Eanna and give them to the property of the Eanna.

(14–16) Arad-Nergal son of Mukīnapli descendant of Egibi assumes responsibility for the repayment of these 155 sheep.

(16–17) In the presence of Nabûmukīn-apli, the šatammu of the Eanna, son of Nādinu descendant of Dābibī;

(18) Nabû-aḫa-iddin, the royal official in charge of the Eanna.

(19) Witnesses: Arad-Marduk, son of Zēriya descendant of Egibi;

(20) Sīn-ēreš son of Nabû-Šumu-līšir descendant of Ibni-ili;

(21) Bēl-nādin-apli son of Mardukšuma-iddin descendant of Bēl-apla-uṣur;

(22) Nādinu, the scribe, descendant of Egibi;

(23) Arad-Marduk, the scribe, descendant of Bēl-apla-uṣur.

(24–25) Uruk. 12 Addaru, year 3 of Cambyses, king of Babylon, king of the lands [11, p. 179-181].

This trial record is symbolically represented as an inference by *modus ponens* (**MP**) as follows:

1. If a man steals X sheep and it belongs to the god (to a temple), then he must replace it **thirtyfold** (i.e. the amount of $X \cdot 30$) [the axiom from the code, see **A1**];

2. Ištar of Uruk and Nanaya is a temple [it is a fact, because 'a temple' is a generalization for the case of 'Ištar of Uruk and Nanaya'];

3. Bēl-iqīša son of Ṣillaya led away 5 sheep belonging to Ištar of Uruk and Nanaya [the fact established by the trial];

4. But 5 unblemished lambs are a property of Ištar of Uruk and Nanaya too [the fact established by the trial];

5. Arad-Nergal son of Mukīnapli descendant of Egibi is a guarantor for the repayment of Bēl-iqīša son of Ṣillaya [the fact established by the document]

Then, Bēl-iqīša son of Ṣillaya must repay 155 sheep to Ištar of Uruk and Nanaya on 25 Addaru, year 3 of Cambyses, king of Babylon.

More formally step by step:

Axiom **A1**:

(1) ([[steal (Ax, z)]] & [[possess (Ty, z)]]) $\Rightarrow$ [[compensate (Ax, Ty, $z \cdot 30$)]]

Facts proved by the trial:

(2) $a \Rightarrow Ax$, where $a =$ 'Bēl-iqīša son of Ṣillaya'

(3) $b \Rightarrow Ty$, where $b =$ 'Ištar of Uruk and Nanaya'

(4) $c \Rightarrow z$, where $c =$ '5 sheep'

(5) [[steal (a, c)]]

(6) [[possess (b, c)]]

Thus, (1) – (6) are axioms for the court, and basing on them the court should draw a logical conclusion. According to the substitution rule, we can substitute a for Ax, b for Ty, and c for z in **A1**. Then we obtain by **SR**:

(7) ([[steal (a, c)]] & [[possess (b, c)]]) $\Rightarrow$ [[compensate (a, b, $c \cdot 30$)]]

From (5) and (6) we deduce by **I&**:

(8) [[steal (a, c)]] & [[possess (b, c)]]

Then, finally, by **MP** (*modus ponens*) we conclude from (7) and (8) that

(9) [[compensate (a, b, $c \cdot 30$)]]

This (9) means that Bēl-iqīša son of Ṣillaya must compensate Ištar of Uruk and Nanaya by paying 150 sheep. Why must he repay additional 5 unblemished lambs? These 5 were born after stealing. Hence, on the one hand, they belong to Ištar of Uruk, too, but, on the other hand, they were not stolen. Therefore they must be given back without any additional compensation. In order to deduce this statement, we should appeal to the property of the completeness of legal information. The 5 unblemished lambs were born to the 5 sheep after stealing. Hence, they belong to Ištar of Uruk, but they were not stolen. Let A_1, A_2, $\ldots$, A_n be all the n cases related to the obligation B to compensate additionally. The case of the 5 lumbs, C, is beyond all the n cases A_1, A_2, $\ldots$, A_n, i.e. we have: $C \Rightarrow \neg(A_1 \otimes A_2 \otimes \ldots, \otimes A_n)$. Then $C \Rightarrow \neg B$. So, they must be given back without compensations.

The next example about additional compensations:

1. 2 UZ.TUR$^{\text{mušen}}$ *ša*$_2$ ⌈$^{\text{d}}$GAŠAN⌉ UNUG$^{\text{ki}}$ *u*$_3$ $^{\text{d}}$*na-na-*[*a ša*$_2$ *qa-puut-tu*$_4$]

2. *ša*$_2$ $^{\text{m}}$*ni-din-tu*$_4$ *u*$_3$ $^{\text{m}}$*gu-za-nu* DUMU.MEŠ *ša*$_2$ $^{\text{md}}$*na-na-a* [MU ... U$_4$ 11-*kam*$_2$ *ša*$_2$ ITI AB]

3. MU 2-*kam*$_2$ $^{\text{m}}$*kam-bu-zi-ia* LUGAL TIN.TIR$^{\text{ki}}$ LUGAL [KUR. KUR ...

4. *ša*$_2$ KA$_2$.GAL $^{\text{d}}$15 *di-i-ku-ma i-na ṭi-ṭu*$_3$ [*qit-bu-ru* ...]

5. $^{\text{m}}$ŠEŠ-SUM.NA *u*$_3$ $^{\text{md}}$*a-nu*-ŠEŠ. MEŠ-TIN-⌈iṭ⌉ [DUMU.MEŠ *ša*$_2$ $^{\text{md}}$NA$_3$-KAD$_2$ $^{\text{md}}$*na-na-a*-ŠEŠ-MU]

6. DUMU-*šu*$_2$ *ša*$_2$ $^{\text{md}}$*na-na-a*-KAM$_2$ *u*$_3$ $^{\text{m}}$ŠEŠ-SUM.NA [DUMU-*šu*$_2$ *ša*$_2$ $^{\text{m}}$*ki-na-a* ...] ID$_2$

7. *ša*$_2$ MUŠEN.MEŠ *i-na meš-ḫišu*$_2$*-nu di-i-ku-ma i-*[*na ṭi-ṭu*$_3$ *iqte-bi-ru*]

8. *a-na ma-ḫar* $^{\text{md}}$NA$_3$-DUIBILA $^{\text{lu2}}$ŠA$_3$.TAM E$_2$.AN. NA DUMU-*šu*$_2$ *ša*$_2$ $^{\text{m}}$[*na-di-nu* DUMU *da-bi-bi*]

9. *u*$_3$ $^{\text{md}}$NA$_3$-ŠEŠ-MU $^{\text{lu2}}$SAG. LUGAL $^{\text{lu2}}$EN *pi-qit-tu*$_4$ E$_2$.AN. NA [...]

10. *ina* UKKIN *iq-bu-u*$_2$ *um-ma* U$_4$ 11-*kam*$_2$ *ša*$_2$ ITI AB MU 2-*kam*$_2$ *ni-i-ni u*$_3$ $^{\text{md}}$*na-na-a*-MU

11. DUMU-*šu*$_2$ *ša*$_2$ $^{\text{md}}$*in-nin*-NUMUN-DU$_3$ *it-ti a-ḫa-meš ina ku-tal* BAD$_3$ ID$_2$ *ni-ḫi-*⌈*ir*⌉*-ru*

12. 2 UZ.TUR$^{\text{mušen}}$.ME NIG$_2$.GA $^{\text{d}}$GAŠAN UNUG$^{\text{ki}}$ *ša*$_2$ *qa-pu-uttu*$_4$ *ša*$_2$ $^{\text{m}}$*ni-din-tu*$_4$ *u*$_3$ *mgu-za-nu*

13. DUMU.MEŠ *ša*$_2$ $^{\text{md}}$*na-na-a*-MU *ki-i ni-du-ku i-na ṭi-ṭu*$_3$ *ni-iqte-bir*

14. ⌈*pag*⌉*-ra-nu ša*$_2$ UZ.TUR$^{\text{mušen}}$-*a'* 2 *ša*$_2$ $^{\text{m}}$ŠEŠ-MU *u*$_3$ $^{\text{md}}$DIŠ-ŠEŠ. MEŠ-TIN-*iṭ* DUMU.MEŠ

15. *ša*$_2$ $^{\text{md}}$NA$_3$-KEŠDA-*ir* $^{\text{md}}$*na-naa*-ŠEŠ-MU DUMU-*šu*$_2$ *ša*$_2$ $^{\text{md}}$*na-na-a*-APIN-*eš* 16. $^{\text{m}}$ŠEŠ-SUM.NA DUMU-*šu*$_2$ *ša*$_2$ $^{\text{m}}$*ki-na-a u* $^{\text{md}}$*na-na-a*-MU DUMU-*šu*$_2$ *ša*$_2$ $^{\text{md}}$*in-nin*-NUMUN-DU$_3$

17. *i-du-ku-ma ina ṭi-ṭu*$_3$ *iq-bi-ri i-na* UKKIN $^{\text{lu2}}$*qi-pa-a-nu u* $^{\text{lu2}}$DUMU DU$_3$-*i*.[MEŠ]

18. *in-nam-ru-ma ki-i pi-i* $^{\text{lu2}}$*mu-kin-nu-tu ša*$_2$ $^{\text{m}}$ŠEŠ-SUM.NA

19. $^{\text{md}}$*a-nu*-ŠEŠ.MEŠ-TIN-*iṭ* $^{\text{md}}$*na-na-a*-ŠEŠ-MU *u*$_3$ $^{\text{m}}$ŠEŠMU *i-na* UKKIN *qi*$_2$*-pa-a-nu*

20. *u*$_3$ $^{\text{lu2}}$DUMU DU$_3$-*i*.ME *e-li ram-ni-šu*$_2$*-nu* ⌈*u*$_3$⌉ [*ma-ḫar*] $^{\text{md}}$NA$_3$-DU-IBILA

21. $^{\text{lu2}}$ŠA$_3$.TAM E$_2$.AN.NA $^{\text{md}}$NA$_3$-ŠEŠ-MU $^{\text{lu2}}$SAGLUGAL $^{\text{lu2}}$EN *pi-qit-tu*$_4$ E$_2$.AN.NA UKKIN

22. ^{lu2}TIN.TIR^{ki}.ME *u*$_3$ ^{lu2}UNUG^{ki}-*a-a ki* UZ.TUR^{mušen} 1-*en a-di* 30 *ku-um* UZ.TUR^{mušen}.ME-*a'*

23. 2 *e-li* ^{m}ŠEŠ-MU *u*$_3$ ^{md}DIŠ-PAP. ME-TIN-*iṭ* DUMU.MEŠ *ša*$_2$ ^{md}NA$_3$-KAD$_2$ ^{md}*na-na-a*-ŠEŠ-MU

24. DUMU-*šu*$_2$ *ša*$_2$ ^{md}*na-na-a*-KAM ^{m}ŠEŠ-MU DUMU-*šu*$_2$ *ša*$_2$ ^{m}*ki-na-a u* ^{md}*na-na-a*-MU DUMU-*šu*$_2$ *ša*$_2$ ^{md}INNIN.NANUMUN-DU$_3$

25. ^{lu2}EN *ar'*-<*ni*> *šu*$_2$*-nu ša*$_2$ *la in-nam-ru a-na e-ṭe-ru a-na* NIG$_2$. GA E$_2$.AN.NA *šul-lu-un-du*

26. *e-li-šu*$_2$*-nu ip-ru-su* UZ.TUR. MEŠ-*a* 60-*šu ib-ba-ku-nim-ma a-na* NIG$_2$.GA E$_2$.AN.NA

27. *i-nam-di-nu ina u*$_2$*-šu-uz-zu ša*$_2$ ^{md}NA$_3$DU-IBILA ^{lu2}ŠA$_3$.TAM E$_2$.AN.NA DUMU ^{m}*da-bi-bi*

28. ^{md}NA$_3$-ŠEŠ-MU ^{lu2}SAGLUGAL ^{lu2}EN SIG$_5$ E$_2$.AN.NA

28. ^{lu2}*mu-kin-nu* ^{md}30-KAM$_2$ DUMU-*šu*$_2$ *ša*$_2$ ^{md}NA$_3$-MUSI.SA

29. DUMU ^{m}*ib-ni*-DINGIR ^{md}UTU-DU-IBILA DUMU*šu*$_2$ *ša*$_2$ ^{md}DI.KU$_5$-PAP.ME-MU DUMU ^{m}*ši-gu-u*$_2$*-a*

30. ^{m}*la-a-ba-ši*-^{d}AMAR.UTU DUMU-*šu*$_2$ *ša*$_2$ ^{m}IR$_3$-^{d}EN DUMU *me-gi-bi* ^{md}AMAR. UTU-MU-ŠEŠ DUMU-*šu*$_2$ *ša*$_2$ ^{md}EN-TIN-*iṭ*

31. DUMU ^{m}*bu-u*$_2$*-ṣu* $^{m\langle d\rangle}$ENKAR-^{d}NA$_3$ ^{lu2}SAG ^{md}*a-nu*-MUDU$_3$ DUMU-*šu*$_2$ *ša*$_2$ ^{md}NA$_3$SUR DUMU ^{md}[PN]

32. ^{md}INNIN-ŠEŠ-MU DUMU*šu*$_2$ *ša*$_2$ ^{md}NA$_3$-DU$_3$-ŠEŠ DUMU ^{m}KUR-i ^{m}*lu-ṣa-ana*-ZALAG$_2$-^{d}UTU DUMU-*šu*$_2$ *ša*$_2$ ^{m}*šu-la-a*

33. DUMU ^{lu2}E$_2$.MAŠ-^{d}MAŠ ^{md}DIŠ-ŠEŠ-MU DUMU-*šu*$_2$ *ša*$_2$ ^{m}ŠU DUMU ^{m}KUR-*i*

33–34. ^{m}*na-di-nu* DUB.SAR DUMU-*šu*$_2$ *ša*$_2$ ^{md}EN-ŠEŠ. MEŠ-BA-*ša*$_2$ DUMU ^{m}*e-gi-bi* ^{m}IR$_3$-^{d}AMAR. UTU DUB.SAR DUMU-*šu*$_2$ *ša*$_2$ ^{m}[^{d}AMAR. UTU-MU-MU DUMU ^{md}EN-A-URI$_3$]

35. UNUG^{ki} ITI AB U$_4$ 12-*kam*$_2$ MU 2-kam$_2$ ^{m}*kam-bu-zi*-[*ia* LUGAL TIN.TIR^{ki} LUGAL KUR.KUR]

(1–4) 2 ducks, property of the Lady-of-Uruk and Nanaya [from the pen of] ^{m}Nidintu and ^{m}Guzānu, sons of ^{m}Nanaya-iddin [... on 11 Tebētu] year 2 of Cambyses king of Babylon king [of the lands] at the Ištar Gate, killed and [buried] in mud.

(5–10) ^{m}Aẖa-iddin and ^{m}Anuaẖẖē-bulliṭ, [sons of ^{m}Nabû-kāṣir, ^{m}Nanaya-aẖa-iddin] son of ^{m}Nanaya-ēreš and ^{m}Aẖa-iddin [son of Kīnaya ...] in whose working-area the birds were killed [and buried in mud], said thus before ^{m}Nabû-mukīn-apli, *šatammu* of the Eanna, son of ^{m}[Nādinu descendant of Dābibī], and ^{m}Nabû-aẖa-iddin the *ša rēš šarri*, the administrator of the Eanna [...], in the assembly:

(10–11) "On 11 Ṭebētu, year 2, we were digging below the canal wall, together with ^{m}Nanaya-iddin son of mInnin-zēra-ibni."

(12–13) "When we killed 2 ducks, property of the Lady-of-Uruk, from the pen of ^{m}Nidintu and ^{m}Guzānu, sons of ^{m}Nanaya-iddin, we buried them in mud."

(14–18) The corpses of these 2 birds that ^{m}Aẖa-iddin and ^{m}Anu-aẖẖē-bulliṭ sons of ^{m}Nabûkāṣir, ^{m}Nanaya-aẖa-iddin son of ^{m}Nanaya-ēreš, ^{m}Aẖa-iddina son of ^{m}Kīnaya, and ^{m}Nanaya-iddin son of ^{m}Innin-zēra-ibni killed and buried in mud were inspected in the assembly of the *qīpu* officials and the *mār banī*.

(18–26) In accordance with the testimony of ^{m}Aẖa-iddin, ^{m}Anu-aẖẖē-bulliṭ, ^{m}Nanaya-aẖa-iddin and ^{m}Aẖa-iddin against themselves in the assembly of the *qīpu* officials and the *mār banī*, and [before] ^{m}Nabû-mukīn-apli, the *šatammu* of the Eanna, ^{m}Nabû-aẖa-iddin, the *ša rēš šarri*, administrator of the Eanna, the assembly of Babylonians and Urukians – they decided that ^{m}Aẖa-iddin and ^{m}Anu-aẖẖē-bulliṭ sons of ^{m}Nabû-kāṣir, ^{m}Nanaya-aẖa-iddin son of ^{m}Nanaya-ēreš, ^{m}Aẖa-iddin son of ^{m}Kīnaya, and ^{m}Nanaya-iddin son of ^{m}Inninzēra-ibni, their accomplice in crime who was not seen, must pay a **thirtyfold** restitution for the 2 ducks to the property of the Eanna.

(26–27) They shall bring and pay these 60 ducks to the property of the Eanna.

(27) In the presence of ^{m}Nabûmukīn-apli, the *šatammu* of the Eanna, descendant of Dābibī.

(28) ^{m}Nabû-aẖa-iddin, the *ša rēš šarri*, administrator of the Eanna.

(28) Witnesses: ^{m}Sîn-ēreš son of ^{m}Nabû-šumu-līšir descendant of Ibni-ilī;

(29) ^{m}Šamaš-mukīn-apli son of ^{m}Madānu-aẖẖē-iddin descendant of Šigûa;

(30–31) ^{m}Lâbāši-Marduk son of ^{m}Arad-Bēl descendant of Egibi; ^{m}Marduk-šuma-uṣur son of ^{m}Bēluballiṭ descendant of ^{m}Būṣu;

(31) ^{m}Bēl-eṭēri-Nabû, *ša rēši*; ^{m}Anušuma-ibni son of ^{m}Nabû-ušēzib descendant of [PN];

(32–33) ^{m}Innin-aẖa-iddin son of ^{m}Nabû-bāni-aẖi descendant of Kurī; ^{m}Lûṣa-ana-nūr-Šamaš son of ^{m}Šulaya descendant of Šangû-Ninurta;

(33) ^{m}Anu-aẖa-iddin son of ^{m}Gimillu descendant of Kurī;

(33–34) $^{\mathrm{m}}$Nādinu, the scribe, son of $^{\mathrm{m}}$Bēl-aẖẖē-iqīša descendant of Egibi; $^{\mathrm{m}}$Arad-Marduk, the scribe, son of [$^{\mathrm{m}}$Marduk-šuma-iddin descendant of Bēl-apla-uṣur]

(35) Uruk. 12 Ṭebētu, year 2 of Cambyses, king of Babylon, king of the lands [12, p. 47-52]

Formally:

1. If a man steals X ducks and they belong to the god (to a temple), then he must replace them **thirtyfold** (i.e. the amount of $X \cdot 30$ [the axiom from the code, see **A1**);

2. The Lady-of-Uruk and Nanaya is a temple [it is a fact, because 'a temple' is a generalization for the case of 'the Lady-of-Uruk and Nanaya'];

3. $^{\mathrm{m}}$Aẖa-iddin and $^{\mathrm{m}}$Anu-aẖẖē-bulliṭ sons of $^{\mathrm{m}}$Nabû-kāṣir, $^{\mathrm{m}}$Nanaya-aẖa-iddin son of $^{\mathrm{m}}$Nanaya-ēreš, $^{\mathrm{m}}$Aẖa-iddin son of $^{\mathrm{m}}$Kīnaya, and $^{\mathrm{m}}$Nanaya-iddin son of $^{\mathrm{m}}$Inninzēra-ibni killed 2 ducks and buried them in mud [the fact established by the trial];

4. The 2 ducks are a property of the Lady-of-Uruk and Nanaya [the fact established by the trial];

Then, $^{\mathrm{m}}$Aẖa-iddin and $^{\mathrm{m}}$Anu-aẖẖē-bulliṭ sons of $^{\mathrm{m}}$Nabû-kāṣir, $^{\mathrm{m}}$Nanaya-aẖa-iddin son of $^{\mathrm{m}}$Nanaya-ēreš, $^{\mathrm{m}}$Aẖa-iddin son of $^{\mathrm{m}}$Kīnaya, and $^{\mathrm{m}}$Nanaya-iddin son of $^{\mathrm{m}}$Inninzēra-ibni must repay 60 ducks to the Lady-of-Uruk and Nanaya on 12 Ṭebētu, year 2 of Cambyses, king of Babylon.

Axiom **A1**:

(1) ([[steal (Ax, z)]] & [[possess (Ty, z)]]) $\Rightarrow$ [[compensate $(Ax, Ty, z \cdot 30)$]]

Facts proved by the trial:

(2) $(a_1 \& a_2 \& a_3 \& a_4 \& a_5) \Rightarrow Ax$, where $(a_1 \& a_2) =$ '$^{\mathrm{m}}$Aẖa-iddin and $^{\mathrm{m}}$Anu-aẖẖē-bulliṭ sons of $^{\mathrm{m}}$Nabû-kāṣir', $a_3 =$ '$^{\mathrm{m}}$Nanaya-aẖa-iddin son of $^{\mathrm{m}}$Nanaya-ēreš', $a_4 =$ '$^{\mathrm{m}}$Aẖa-iddin son of $^{\mathrm{m}}$Kīnaya', and $a_5 =$ '$^{\mathrm{m}}$Nanaya-iddin son of $^{\mathrm{m}}$Inninzēra-ibni'

(3) $b \Rightarrow Ty$, where b = 'the Lady-of-Uruk and Nanaya'

(4) $c \Rightarrow z$, where c = '2 ducks'

(5) [[steal (a_1 & a_2 & a_3 & a_4 & a_5, c)]]

(6) [[possess (b, c)]]

Expressions (1) – (6) are axioms for the court. According to **SR**, we can substitute (a_1 & a_2 & a_3 & a_4 & a_5) for Ax, b for Ty, and c for z in **A1**. Then we obtain:

(7) ([[steal (a_1 & a_2 & a_3 & a_4 & a_5, c)]] & [[possess (b, c)]]) $\Rightarrow$ [[compensate (a_1 & a_2 & a_3 & a_4 & a_5, b, $c \cdot 30$)]]

From (5) and (6) by **I&**:

(8) [[steal (a_1 & a_2 & a_3 & a_4 & a_5, c)]] & [[possess (b, c)]]

From (7) and (8) by **MP**:

(9) [[compensate (a, b, $c \cdot 30$)]]

In Judaism there is the same sufficient condition to define a case as theft. This condition is called an acquisition (קניין) in the form of lifting up or pulling of thing belonging to another man. So, according to Judaism, killing 2 ducks and buring them in mud is a case of theft, too.

Another example of trial record is denoted BM 46660 (see [11, p. 43-44]) and tells us that Marduk-šarannu has accused Kīnaya of striking his son and, as a result, two siblings, a brother and a sister, guarantee that Kīnaya will appear at the court. If Kīnaya escapes, then the two must pay compensation to Marduk-šarannu:

1'. [*u* $^{\mathrm{m}}$*ki-na-a* DUMU-*šú šá* $^{\mathrm{m}}$BA]-*šá a-na*
2'. ⌈$^{\mathrm{lu2}}$DUMU⌉ [DU$_3$ x x x] *it-ti a-ḫameš*
3'. *il-la-ku-ú-ma di-i-nu* [*šá*]
4'. $^{\mathrm{md}}$AMAR.UTU-LUGAL-*a-nu a-na* $^{\mathrm{m}}$*ki-na-*[*a*]
5'. *iq-bu-ú um-mu* DUMU-*u-*⌈*a*⌉
6'. *ta-an-da-ḫa-aṣ ina* IGI $^{\mathrm{lu2}}$[...]
7'. *i-dab-bu-ub* $^{\mathrm{md}}$NA$_3$-[NUMUNMU]
8'. A-*šú šá* $^{\mathrm{m}}$ŠEŠ.MEŠ-*šá-iá u* [$^{\mathrm{f}}$*iṣṣur*-X]
9'. NIN-*šú pu-ut* $^{\mathrm{m}}$[*ki-na-a*]

10’. A-*šú šá* $^{\mathrm{m}}$BA-*šá-a na-*[*šu-u ki-i*]
11’. $^{\mathrm{m}}$*ki-na-a iḫ-te-*[*li-qu*]
12’. ZI.MEŠ *šá* DUMU-*šú sa*$_2$ $^{\mathrm{md}}$[AMAR.UTU-LUGAL-*a-nu*]
13’. $^{\mathrm{md}}$NA$_3$-NUMUN-MU *u* $^{\mathrm{f}}$*iṣ-*⌈*ṣur*⌉-[
14’. *ú-šal-lim-mu* $^{\mathrm{lu2}}$*mu-kin-nu* $^{\mathrm{m}}$[PN

(1’–3’) [. . . and Kīnaya son of Iq]īšaya will go to the *mār* [*banī*] together

(3’–7’) They (!) /He will argue the case [in which] Marduk-šarannu said thus to Kīnaya “You **struck** my son!” before the . . .

(7’–10’) Nabû-zēra-iddin son of Aḫḫūšaya and [Iṣṣur-X], his sister, assume responsibility for [Kīnaya] son of Iqīšaya.

(10’–14’) If Kīnaya escapes, Nabûzēra-iddin and Iṣṣur-[X] will pay compensation for the life of the son of Marduk-šarannu.

(14’–15’) Witnesses: PN

[11, p. 43-44].

Symbolically:

1. If a man **strikes** somebody, then he must pay compensation [the axiom from the code, see **A2**];

2. If a man cannot pay, his guarantors must pay [the axiom from a code we can reconstruct];

3. Kīnaya struck Marduk-šarannu’s son [the proven fact];

4. Nabû-zēra-iddin son of Aḫḫūšaya and his sister Iṣṣur-X are guarantors for Kīnaya;

5. If Kīnaya appears at the trial, he must pay compensation to Marduk-šarannu [the first conditional verdict];

6. If Kīnaya escapes, his two guarantors (Nabû-zēra-iddin son of Aḫḫūšaya and his sister Iṣṣur-X) must pay compensation to Marduk-šarannu [the second conditional verdict];

Then, either Kīnaya or his two guarantors (Nabû-zēra-iddin son of Aḫḫūšaya and his sister Iṣṣur-X) must pay compensation to Marduk-šarannu.

Axiom **A2**:

(1) [[strike (Ax, Ay)]] $\Rightarrow$ [[compensate (Ax, Ay, 60 shekels)]]

An axiom from a code which can be reconstructed:

(2) ($\neg$ [[pay (Ax, Ay, n)]] & [[guarantor (Az, Ax)]]) $\Rightarrow$ [[pay (Az, Ay, n)]]

which means that if a man x of the *awīlu*-class cannot pay n for y and there is a guarantor z of the same class, this guarantor must pay.

Paying is a more general case than compensating (the direct conclusion just semantically):

(3) [[compensate (Ax, Ay, n)]] $\Rightarrow$ [[pay (Ax, Ay, n)]]

Facts proved by the court:

(4) [[strike (a, b)]]

(5) $a \Rightarrow Ax$, where a = 'Kīnaya son of Iqīšaya'

(6) $b \Rightarrow Ay$, where b = 'the son of Marduk-šarannu'

(7) $c\&d \Rightarrow Az$, where c & d = 'Nabû-zēra-iddin son of Aḫḫūšaya and his sister Iṣṣur-X'

Expressions (1) – (7) are axioms for the court. According to **SR**, we can substitute a for Ax, b for Ay in **A2**. Then we obtain:

(8) [[strike (a, b)]] $\Rightarrow$ [[compensate (a, b, 60 shekels)]]

From (8) and (4) we have by **MP**:

(9) [[compensate (a, b, 60 shekels)]]

Now we can substitute [[compensate (Ax, Ay, n)]] for [[pay (Ax, Ay, n)]] in (2). Then we substitute a for Ax, b for Ay, c & d for Az, and '60 shekels' for n in (2). So, we have from (2) by **SR**:

(10) (¬ [[compensate (a, b, 60 shekels)]] & [[guarantor (c & d, a)]]) ⇒ [[compensate (c & d, b, 60 shekels)]]

Let us introduce a new inference rule:

$$\frac{A; \qquad (\neg A \wedge B) \Rightarrow C}{A \vee C},$$

because $(A\&((\neg A \wedge B) \Rightarrow C)) \Rightarrow (A \vee C)$ is a tautology of propositional logic. Then from (9) and (10) by this rule we infer:

(11) [[compensate (a, b, 60 shekels)]] ∨ [[compensate (c & d, b, 60 shekels)]]

Thus, we have deduced the trial decision denoted by (11) just automatically.

The next instance of conditional verdicts is taken from the text denoted by BM 31162, found in Opis, and dated to 23.VIII.40 Nebuchadnezzar, king of Babylon (5 November, 565 B.C.), see [11, p. 45-47]. In this trial record, Gudaya, the guarantor of a grain loan to Katimu', testifies that he presented Katimu' to Bau-ērēš (the creditor) to repay the debt. Bau-ērēš has pressed the charges that he has not been repaid by Katimu'. Gudaya must present two witnesses now. If Gudaya finds these witnesses for his claim, then he is clear. If Gudaya does not support his statement by witnessing, then Gudaya must repay the barley and the interest to Bau-ērēš:

1. *a-di* U$_4$ 1-*kám šá* ITI GAN ^{m}*guda-a*
2. A-*šú šá* ^{m}*ḫi-in-ni*-DINGIR.MEŠ 2 ^{lu2}DUMU-DU$_3$.MEŠ
3. ^{lu2}*mu-kin-ne-e-šú a-na* ^{uru}*ú-pi-ia ib-ba-kám-ma*
4. *a-na* ^{md}KA$_2$-KAM$_2$ A-*šú šá* ^{md}NA$_3$-DU$_3$-ŠEŠ
5. *ú-kan-ni šá* ^{m}*ka-ti-mu-'* A-*šú šá*
6. ^{m}*ḫa-gu-ru šá pu-ut še-pi-šú ina* ŠU.2
7. ^{md}KA$_2$-KAM$_2$ *iš-šu-ú ina a-danni-šú*
8. ^{m}*g[u-d]a-a i-bu-ka-šim-⌈ma⌉*
9. ⌈*a-na* ^{m}⌉^{d}KA$_2$-KAM$_2$ *id-di-nu*
10. *ki-i uk-tin-nu-uš za-ki*
11. *ki-i la uk-tin-nu-uš a-ki-i ú-ìl-tim*
12. ŠE.BAR *u* HAR.RA-*šú a-na* ^{md}KA$_2$-KAM$_2$ *it-ta-din*
13. ^{lu2}*mu-kin-nu* ^{m}*si-lim-*^{d}EN A-*šú šá*
14. ^{m}*ba-la-ṭu* ^{m}MU-^{d}AMAR.UTU A-*šú šá*
15. ^{md}NA$_3$-KI-*ia u* ^{lu2}UMBISAG ^{md}NA$_3$-ŠEŠ.MEŠ-MU

16. A-*šú šá* $^{\mathrm{m}}$*šu-la-a* A $^{\mathrm{m}}$*e-gi-bi* $^{\mathrm{uru}}$*úpi-ia*
17. ITI APIN U_4 23-*kám* MU 40-*kám*
18. $^{\mathrm{d}}$NA$_3$-NIG$_2$.DU-URI$_3$ LUGAL TIN.TIR$^{\mathrm{ki}}$

(1–9) By 1 Kislīmu, Gudaya son of Ḫinni-ilī shall bring two *mār banī* (as) his witnesses to Opis and establish, against Bau-ēreš son of Nabû-bāniaḫi, that, at the time (of the termination of the loan), Gudaya brought Katimu' son of Ḫagūru – for whose presence he (Gudaya) assumed guarantee to Bau-ēreš – to him (Bau-ēreš) and handed (Katimu') over to Bau-ēreš.

(10) If he (Gudaya) establishes (the case) against him (Bau-ēreš), he (Gudaya) is clear.

(11–12) If he (Gudaya) does not establish (the case) against him (Bauēreš), then he (Gudaya) shall pay Bauēreš barley and its interest according to the debt-note.

(13–14) Witnesses: Silim-Bēl son of Balāṭu;

(14–15) Iddin-Marduk son of Nabûittiya;

(15–16) and the scribe: Nabû-aḫḫēiddin son of Šulaya descendant of Egibi.

(16–18) Opis. 23 Araḫšamna, year 40 of Nebuchadnezzar, king of Babylon [12, p. 46].

Formally:

1. If a man takes a loan, he must repay the debt according to the debt-note in the presence of a guarantor [the axiom from a code we can reconstruct];

2. If a man cannot pay, his guarantors must pay [the axiom from a code we can reconstruct];

3. Gudaya son of Ḫinni-ilī was a guarantor that Katimu' took a loan from Bau-ēreš [the documented fact];

4. If Gudaya has two witnesses that he presented Katimu' to Bau-ēreš to repay the debt, Gudaya is free [the first conditional verdict];

5. If Gudaya has no witnesses that he presented Katimu' to Bau-ēreš, Gudaya must pay Bauēreš barley and its interest according to the debt-note [the second conditional verdict];

Then, either Gudaya is free or he must pay on 23 Araḫšamna, year 40 of Nebuchadnezzar, king of Babylon.

Axioms from a code:

(1) ([[loan (Ax, Ay, n)]] & [[guarantor (Az, Ax)]]) $\Rightarrow$ [[repay-debt (Ax, Ay, n)]]

Formula (1) means that if a man x of the *awīlu*-class takes a loan from y of the same class and x has a guarantor z of the same class, x must repay the debt according to the debt-note in the presence of a guarantor z.

(2) ($\neg$[[pay (Ax, Ay, n)]] & [[guarantor (Az, Ax)]]) $\Rightarrow$ [[pay (Az, Ay, n)]]

(3) [[repay-debt (Ax, Ay, n)]] $\Rightarrow$ [[pay (Ax, Ay, n)]]

Indeed, [[pay (Ax, Ay, n)]] is a generalization for [[repay-debt (Ax, Ay, n)]]. It is proved just semantically.

(4) ([[pay (Ax, Ay, n)]] & [[witness (Ap, Ax)]] & [[witness (Aq, Ax)]]) $\Rightarrow$ [[free (Ax, n)]]

Formula (4) means that a man x of the *awīlu*-class pays n for y of the same class in the presence of two witnesses p and q of the *awīlu*-class, then x is free for paying n.

(5) ($\neg$ ([[pay (Ax, Ay, n)]] & [[witness (Ap, Ax)]] & [[witness (Aq, Ax)]]) & [[guarantor (Az, Ax)]]) $\Rightarrow$ [[pay (Az, Ay, n)]]

Formula (5) means that if a man x of the *awīlu*-class does not pay n for y of the same class in the presence of two witnesses p and q of the *awīlu*-class and there is a guarantor z for x, then z must pay.

Facts proved by the court:

(6) $a \Rightarrow Ax$, where $a =$ 'Katimu' son of Ḫagūru'

(7) $b \Rightarrow Ay$, where $b =$ 'Bau-ēreš son of Nabû-bāniaḫi'

(8) $c \Rightarrow Az$, where c = 'Gudaya son of Ḫinni-ilī'

(9) [[loan (a, b, n)]]

(10) [[guarantor (c, a)]]

Expressions (1) – (10) are axioms for the court. According to **SR**, we can substitute a for Ax, b for Ay, c for Az in (1) to obtain the following expression:

(11) ([[loan (a, b, n)]] & [[guarantor (c, a)]]) $\Rightarrow$ [[repay-debt (a, b, n)]]

Now, according to **SR**, we can substitute [[repay-debt (Ax, Ay, n)]] for [[pay (Ax, Ay, n)]] and then a for Ax, b for Ay, c for Az in (2):

(12) ($\neg$ [[repay-debt (a, b, n)]] & [[guarantor (c, a)]]) $\Rightarrow$ [[repay-debt (c, b, n)]]

By substituting [[repay-debt (Ax, Ay, n)]] for [[pay (Ax, Ay, n)]] and then c for Ax, b for Ay in (4) we have by **SR**:

(13) ([[repay-debt (c, b, n)]] & [[witness (Ap, c)]] & [[witness (Aq, c)]]) $\Rightarrow$ [[free (c, n)]]

By substituting [[repay-debt (Ax, Ay, n)]] for [[pay (Ax, Ay, n)]] and then a for Ax, b for Ay, c for Az in (5) we have by **SR**:

(14) ($\neg$ ([[repay-debt (a, b, n)]] & [[witness (Ap, a)]] & [[witness (Aq, a)]]) & [[guarantor (c, a)]]) $\Rightarrow$ [[repay-debt (c, b, n)]]

Let us introduce a new inference rule:

$$\frac{A; \qquad \neg(A \wedge B) \Rightarrow C}{\neg B \Rightarrow C},$$

because $(A\&(\neg(A \wedge B) \Rightarrow C)) \Rightarrow (\neg B \Rightarrow C)$ is a tautology of propositional logic. Then from (10) and (14) we deduce:

(15) $\neg$ ([[repay-debt (a, b, n)]] & [[witness (Ap, a)]] & [[witness (Aq, a)]]) $\Rightarrow$ [[repay-debt (c, b, n)]]

From (9), (10) by **I&**:

(16) [[loan (a, b, n)]] & [[guarantor (c, a)]]

Then from (16) and (11) by **MP**:

(17) [[repay-debt (a, b, n)]]

Then we apply for (15) and (17) the following inference rule again:

$$\frac{A;\ (A \wedge B) \Rightarrow C}{B \Rightarrow C},$$

As a result, we obtain:

(18) ¬ ([[witness (Ap, a)]] & [[witness (Aq, a)]]) ⇒ [[repay-debt (c, b, n)]]

From (18) and (13) by inference rule **I**∨:

(19) (¬([[witness (Ap, a)]] & [[witness (Aq, a)]]) ⇒ [[repay-debt (c, b, n)]]) ∨ ((([[repay-debt (c, b, n)]] & ([[witness (Ap, c)]] & [[witness (Aq, c)]])) ⇒ [[free (c, n)]])

This (19) is just a verdict of the trial. So, it is obtained automatically, also.

Usually, any relationship between creditors and debtors was regulated by a legal proceeding that may be formalized as follows:

1. The creditor (C) has pressed the charges that the debtor (D) has not given back the X shekels taken from him.

2. This D is testifying at the trial: "The X shekels of C which I owed, I have paid to him in the presence of two witnesses: W_1 and W_2." In accordance with the words of the stele, it means that D is free.

3. If his witnesses W_1 and W_2 are confirming: "D has repaid the X shekels to C," then D must swear together with his witnesses and D is free and C forfeits his claims.

4. And if D's witnesses do not confirm D's statement, C must swear together with his witnesses W_3 and W_4 that D has taken the X shekels from C in the presence of W_3 and W_4 and D must pay C's money back."

This legal proceeding has the following logical structure:

1. If a man takes a loan, he must do it in the presence of two witnesses W_3 and W_4 [the axiom from a code];

2. If a man took a loan in the presence of two witnesses W_3 and W_4, he must repay the debt [the axiom from a code];

3. If a man repays the debt, he must do it in the presence of two witnesses W_1 and W_2 [the axiom from a code];

4. If a man repays the debt in the presence of two witnesses W_1 and W_2, he is free [the axiom from a code];

5. There are two witnesses W_3 and W_4 that a debtor took a loan from a creditor [a documented fact];

Then, either the debtor must repay the debt or if he repaid it in the presence of two witnesses W_1 and W_2, then he is free.

To sum up, the Neo-Babylonian trial records reconstructed in [11], [12] assume an axiomatization of justice to infer verdicts automatically. In some cases we see even quite long deductions. All these deductions are perfect logically and they are evidences that the Neo-Babylonian justice was formalized logically in fact. Thus, the first logical axiomatic system in all the world was proposed by the Neo-Babylonians at least or their predecessors. But it is very probable that the same system existed much more before and it was established by the Sumerians from the very beginning of Sumerian-Akkadian legal culture, see [25].

6 Aramaic and Greek Legal Documents of Elephantine

In Elephantine (island on the Nile in southern Egypt) there were excavated many papyri containing legal texts and compliled in Coptic, Aramaic, Greek, and Latin. The legal documents of Elephantine in Aramaic have many similarities between the Aramaic and Old Babylonian legal formulae [9]. For instance, the structure of documents and many legal phrases were directly taken from the (Neo-)Babylonian jurisprudence or were based on the Mesapotamian legal system as such [9]. This indicates that the Aramaic legal system exposed in the Achaemenid Empire is a continuation of the (Neo-)Babylonian tradition in fact.

Let us consider a Jewish business contract in Aramaic dated to 12 September, 471 B.C. This document was drawn up between Konaiah son of Zadak and Mahseiah son of Jedaniah. According to the contract, Konaiah has an access to Mahseiah's gateway to build there a wall to continue all along the common wall between their two properties (lines 3-4): [[build ('Konaiah', 'wall')]]. The wall is regarded as the property of Mahseiah (lines 4-5): [[possess ('Mahseiah', 'wall')]]. Konaiah and his heirs agree that Mahseiah and his heirs can build on that wall and will have a free access through the gateway and if Konaiah and his heirs deny these rights, then they will incur a penalty of five karsh (lines 6-14): [[claim ('Konaiah' & 'his heirs', ¬ [[possess ('Mahseiah' & 'his heirs', 'wall')]])]] ⇒ [[compensate ('Konaiah' & 'his heirs', 'Mahseiah' & 'his heirs', '5 karsh of pure silver')]]. The text of this contract is as follows:

> [*Date*]
>
> [1]On the 18th of Elul, that is day 28 of Pachons, year 15 of Xerxes the king,
>
> [*Parties*]
>
> said [2]Konaiah son of Zadak, an Aramean of Syene of the detachment of Varyazata, [2]to Mahseiah son of Jenadiah, an Aramean of Syene [3]of the detachment of Varyazata, saying:
>
> [*Building Rights*]
>
> I came to you and you gave me the gateway of the house of yours to build [4]a wall there.
>
> [*Investiture*]
>
> That wall is yours — (the wall) which adjoins the house of mine at its corner which is above. [5]That wall shall adjoin the side of my house from the ground upwards, from the corner of my house which is above to the house of Zechariah.
>
> [*Restraint Waiver I*]
>
> [6]Tomorrow or the next day, I shall not be able to restrain you from building upon that wall of yours.
>
> [*Penalty I*]
>
> [7]If I restrain you, I shall give you silver, 5 karsh by the stone(-weight)s of the king, pure silver,
>
> [*Reaffirmation I*]
>
> and that wall [8]is likewise (yours).
>
> [*Restraint Waiver II*]

And if Konaiah die tomorrow or the next day, son or daughter, brother or sister, [9]near or far, member of a detachment or town [8]shall not be able [9]to restrain Mahsah or a son of his from building upon [10]that wall of his.

[*Penalty II*]

Whoever shall restrain (one) of them shall give him the silver which is written above

[*Reaffirmation II*]

and the wall [11]is yours likewise and you have right to build upon it upwards.

[*Restraint Waiver III*]

And I, Konaiah, shall not be able [12]to say to Mahsah, saying: "(ERASURE: Not) That gateway is not yours and you shall not go out into

the street which is [13]between us and between the house of Peftuauneit the boatman."

[*Penalty III*]

If I restrain you, I shall give you the silver which is written above

[*Reaffirmation III*]

[14]and you have right to open that gateway and to go out into the street which is between us (and Pefṭuauneit).

[*Scribe*]

[15]Wrote Pelatiah son of Ahio this document at the instruction of Konaiah.

[*Witnesses*]

The witnesses herein: [16](2nd hand) witness Mahsah son of Isaiah; (3rd hand) witness Shatibarzana son of *'trly*; (3rd hand) witness Shatibarzana son of *'trly*; [17](4tn hand) witness Shemaiah son of Hosea; (5tn hand) witness Phrathanjana son of Artakarana; (6tn hand) [18]witness Bagadata son of Nabukudurri; (7th hand) *Ynbwly* son of Darga; (8tn hand) [19]witness Baniteresh son of Wahpre; (9tn hand) witness Shillem son of Hoshaiah.

[*Endorsement*]

[20]Document (sealing) of the wall which is built which Konaiah wrote for Mahsah [16, p. 152-157].

As we see, Konaiah and Mahseiah, although they are Jews, are called Arameans in the contract. This reflects the double-identity of Jews in Elephantine as well as in other parts of Achaemenid Empire: both Aramean and Jew. Among the eight witnesses, only three are Jews and the others display a mixed onomasticon: Persian, Caspian, Babylonian, and Egyptian (lines 16-19). All this attests to good neighbour relationships among different peoples. The contract is fomalizable in the following way:

(1) [[build (a, b)]] ⇒ [[possess (c, b)]]

(2) [[claim (a & d, ¬ [[possess (c & d, b)]])]] ⇒ [[compensate (a & d, c & d, n)]]

(3) 'Konaiah son of Zadak' ⇒ a

(4) 'wall' ⇒ b

(5) 'Mahseiah son of Jedaniah' ⇒ c

(6) 'his heirs' ⇒ d

(7) '5 karsh of pure silver' ⇒ n

Axioms (1) – (7) given in the contract allow a possible court to draw the following two conclusions just by *modus ponens* (**MP**) and substitution rule (**SR**):

1. Konaiah son of Zadak built up the wall. Then from (1), (3), (4), (5) it follows that it is a property of Mahseiah son of Jedaniah.

2. Konaiah and his heirs denied that Mahseiah and his heirs can build on that wall and have a free access through the gateway. Then from (2), (3), (4), (5), (6), (7) it is concluded that Konaiah and his heirs should incur a penalty of five karsh.

The next document compiled in Aramaic and dated to 2 January, 464 B.C. contains a judicial settlement of disputes between Dargamana son of Khvarshaina and Mahseiah son of Jedaniah. Dargamana complained that Mahseiah took his land, but neither party could produce a document of title who is landowner in fact. The court, headed by the Persian Damidata, settled the dispute as follows. Mahseiah with his wife and son swore by YHW, perhaps in the Elephantine Jewish Temple, that the land does not belong to Dargamana (lines 4-7): ¬ [[possess ('Dargamana', 'land')]]. As a consequence, Dargamana was satisfied by this oath (lines 11-12) and drew up the withdrawal that he or any child or sibling in his name will incur a penalty of twenty karsh, if they dispute this court decision (lines 12-16): [[claim ('Dargamana', [[possess ('Dargamana', 'land')]])]] ⇒ [[compensate ('Dargamana' & 'his heirs', 'Mahseiah' & 'his heirs', '20 karsh of pure silver')]]. The document:

RECTO

[*Date*]

[1]On the 18th of Kislev, that is d[ay 13+]4 (= 17) of Thoth, year 21 (of Xerxes the king), the beginning of the reign when [2]Artaxerxes the king sat on his throne,

[*Parties*]

said Dargamana son of Khvarshaina, a Khwarezmian whose place [3]is made in Elephantine the fortress of the detachment of Artabanu, [3]to Mahseiah son of Jedaniah, a Jew who is in the fortress of Elephantine [4]of the detachment of Varyazata, saying:

[*Complaint*]

You swore to me by YHW the God in Elephantine the fortress, you and your wife [5]and your son, all (told) 3, about the land of mine on account of which I complained against you before [6]Damidata and his colleagues the judges,

[*Oath I*]

and they imposed upon you for me the oath to swear by YHW on account of [7]that [6]land, [7]that it was not land of Dargamana, mine, behold I.

[*Boundaries*]

Moreover, behold the boundaries of that land [8]which you swore to me on account of it: my house, Dargamana is to the east of it; and the house of Konaiah son of Zadak, [9]a Jew of the detachment of Atropharna, is to the west of it; and the house of [Jeza]niah son of Uriah, [10]a Jew of the detachment of Varyazata, is below it; and the house of Espemet son of Peftuauneit, [11]a boatman of the rough waters, is above it.

[*Oath II*]

You swore to me by YHW

[*Satisfaction*]

and satisfied [12]my heart about that land.

[*Waiver of Suit*]

I shall not be able to institute against you suit or process — I, or son of mine or daughter [1]3of mine, brother or sister of mine near or far — about that land (against) you, or son of yours or daughter of yours, brother or sister of yours, near or far.

[*Penalty*]

[14]Whoever shall institute against you (suit) in my name about that land shall give you silver, 20, that is twenty, karsh by the stone(-weighf)s of [15]the king, silver 2 q(uarters) to the ten,

[*Affirmation of Investiture*]

and that land is likewise yours and you are withdrawn from [16]any suit (in) which they shall complain against you on account of that land.

[*Scribe and Place*]

Wrote Itu son of Abah [17]this [16]document [17]in Syene the fortress at the instruction of Dargamana.

[*Witnesses*]

(2nd hand) Witness Hosea son of Peṭekhnum; (3rd hand) witness [18]Gaddul son of Igdal; (4tn hand) witness Gemariah son of Ahio; (5tn hand) Meshullam son of Hosea; (6th hand) [19]Sinkishir son of Nabusumiskun; (7tn hand) witness Hadadnuri the Babylonian; (8th hand) [20]witness Gedaliah son of Ananiah; (9th hand) [21]witness Aryaicha son of Arvastahmara.

VERSO

[*Endorsement*]

[22]Document (sealing) of withdrawal which [Dargama]na son of Khvarshaina wrote for Mahseiah [16, p. 158-162].

Formally:

(1) $\neg$ [[possess (a, b)]].

(2) [[claim (a, [[possess (a, b)]])]] $\Rightarrow$ [[compensate (a & c, b, d & c, n)]]

(3) ‘Dargamana son of Khvarshaina’ $\Rightarrow a$

(4) ‘land’ $\Rightarrow b$

(5) ‘his heirs’ $\Rightarrow c$

(6) ‘Mahseiah son of Jedaniah’ $\Rightarrow d$

(7) ‘20 karsh of pure silver’ $\Rightarrow n$

It is worth noting that Mahseiah is called Jew now, because for the court decision it was necessary to have his oath and his religious identity plaid a significant role there. Among the eight witnesses, the five are Jews and the others are Babylonian and Persian (lines 19-21).

In accordance with axioms (1) – (7) each next court can conclude by **MP** and **SR**:

Dargamana son of Khvarshaina brought a lawsuit against Mahseiah son of Jedaniah that the land belongs to Dargamana. Then from (2), (3), (4), (5), (6), (7) it follows that Dargamana and his heirs should incur a penalty of twenty karsh.

The next document is compiled in Greek and dated to 29 June-28 July, 284 B.C. It is a testament and voluntary disposition of property belonging to the spouses Dionysios, a Temnian, and Kallista, a Temnian. According to this testament, if anything happens to Dionysios, then he leaves all his own belongings to Kallista: [[die ('Dionysios')]] $\Rightarrow$ [[inherit ('Kallista', 'Dionysios', 'all his own belongings')]]. Dionysios also inherits all from the deceased spouse (lines 3-5): [[die ('Kallista')]] $\Rightarrow$ [[inherit ('Dionysios', 'Kallista', 'all her own belongings')]]. If anything then happens to Dionysios after he inherited the Kallista's property, then he leaves the belongings to all his own sons: ([[inherit ('Dionysios', 'Kallista', 'all her own belongings')]] & [[die ('Dionysios')]]) $\Rightarrow$ [[inherit ('sons of Dionysios', 'Dionysios', 'all her own belongings')]]. If anything then happens to Kallista after she inherited the Dionysios' property, then she leaves the belongings to all her own sons (lines 6-8): ([[inherit ('Kallista', 'Dionysios', 'all her own belongings')]] & [[die ('Kallista')]]) $\Rightarrow$ [[inherit ('sons of Kallista', 'Kallista', 'all her own belongings')]]. Nevertheless, there is the following exception in the inheritance awarded to sons: Bakchios, Herakleides, and Metrodoros may receive something from Dionysios and Kallista for their labor while their father and mother are alive, but if they are married, the belongings of Dionysios and Kallista will be shared in common by all the sons (lines 8-10): $\neg$ [[married ('Bakchios')]] $\Rightarrow$ [[pay ('Dionysios' & 'Kallista', 'Bakchios', 'shares for the labor')]]; $\neg$ [[married ('Herakleides')]] $\Rightarrow$ [[pay ('Dionysios' & 'Kallista', 'Herakleides', 'shares for the labor')]]; $\neg$ [[married ('Metrodoros')]] $\Rightarrow$ [[pay ('Dionysios' & 'Kallista', 'Metrodoros', 'shares for the labor')]]. If Dionysios or Kallista owe a debt, all the sons in common should repay their debts: [[loan ('Dionysios' $\vee$ 'Kallista', Ay, n)]] $\Rightarrow$ [[repay-debt ('sons of Dionysios and Kallista', Ay, n)]]. If any one of the sons denies in repaying their debts, he should repay one thousand drachmas of silver (lines 12-13): ([[loan ('Dionysios' $\vee$ 'Kallista', Ay, n)]] & $\neg$ [[repay-debt ('son of Dionysios and Kallista', Ay, n)]]) $\Rightarrow$ [[compensate ('son of Dionysios and Kallista', 'Dionysios' & 'Kallista', '1000 drachmas')]]. Please see:

RECTO

[*Date*]

[1]In the 40th year of the reign of Ptolemy, in the month of Gorpiaios, Menelaos son of Lagos being priest.

[*Title*]

Con[2]tract and acknowledgment.

[*Parties*]

Dionysios, a Temnian, has composed this with his wi[3]fe [2]Kallista, a Temnian.

[*Testament I*]

[3]If anything should happen to Dionysios, he leaves all his own belongings to Kallista and she is in control of [4]all the belongings as long as she lives. If anything should happen to Kallista while Dionysios is alive, [5]Dionysios is in control of the belongings.

[*Testament II*]

And if anything (then) happens to Dionysios, let him leave the belongings [6]to all his own sons. In the same way, let Kallista, if anything should (then) happen to her, leave the [7]belongings to all the sons, except for the shares which [8]Bakchios, Herakleides, and Metrodoros [7]may receive from Dionysios and Kallista for their labor [8]while their father and mother are alive. [9]But if Bakchios, Herakleides, and Metrodoros are [8]married and registered, let the belongings of Dionysios and Kallista be (shared) [10]in common by all the sons.

[*Obligation of Heirs*]

If Dionysios or Kallista, while alive, should be in need or owe a debt [11]let all the sons in common feed them and all join in repaying their debts.

[*Penalty*]

If any one of them [12]should not be willing to support them or join in repaying their debts or should not join in burying them, let him repay one thousand drachmas of silver, [13]and let there be requisition from the one who is insubordinate and does not act in accordance with what is written.

[*Release of Obligation*]

If [14]Dionysios or Kallista should leave any debt, let it be permitted to the sons not to enter into (possession of the inheritance) if they do not wish to [15]when Dionysios and Kallista have died.

[*Validity*]

Let this contract be valid in every respect everywhere, [16]wherever it may be brought, as if the covenant had been made there.

[*Guardian of document*]

They have deposited the contract willingly [17]with the contract keeper, Herakleitos.

[*Witnesses*]

Witnesses:

Polykrates an Arcadian; Androsthenes a Coan; [18]Noumenios a Cretan; Simonides a Maronean; Lysis and Herakleitos Temnians.

VERSO

of Dio of Bakchios of [Poly] of Noume of Lysis nysios (sealings) of Kallista (sealings) c[rates] nios (sealings) of Metrodoros of Simonides of Herakleides of Herakletos of Androsthenes [16, p. 412-413].

Symbolically:

(1) [[die (a)]] $\Rightarrow$ [[inherit (b, a, c)]]

(2) [[die (b)]] $\Rightarrow$ [[inherit (a, b, c)]]

(3) ([[inherit (b, a, c)]] & [[die (b)]]) $\Rightarrow$ [[inherit (d, b, c)]]

(4) ([[inherit (a, b, c)]] & [[die (a)]]) $\Rightarrow$ [[inherit (d', a, c)]]

(5) $\neg$ [[married (e)]] $\Rightarrow$ [[pay (a & b, e, f)]]

(6) [[loan ($a \vee b$, Ay, n)]] $\Rightarrow$ [[repay-debt (d & d', Ay, n)]]

(7) ([[loan ($a \vee b$, Ay, n)]] & $\neg$ [[repay-debt (g, Ay, n)]]) $\Rightarrow$ [[compensate (g, a & b, m)]]

(8) ‘Dionysios’ $\Rightarrow a$

(9) ‘Dionysios’ $\Rightarrow e$

(10) ‘Kallista’ $\Rightarrow b$

(11) ‘all own belongings’ $\Rightarrow c$

(12) ‘sons of Kallista’ $\Rightarrow d$

(13) ‘sons of Dionysios’ $\Rightarrow d'$

(14) ‘Bakchios’ or ‘Herakleides’ or ‘Metrodoros’ $\Rightarrow e$

(15) ‘one of the sons of Dionysios or Kallista’ $\Rightarrow g$

(16) ‘1000 drachmas’ $\Rightarrow m$

From axioms (1) – (16) each next court can conclude by **MP** and **SR**:

Dionysios died. Then Kallista inherits.

Kallista died. Then Dionysios inherits.

Kallista inherited and died. Then her sons inherit.

Dionysios inherited and died. Then his sons inherit.

Bakchios or Herakleides or Metrodoros are not merried. Then each of them before his marriage can receive something from Dionysios and Kallista for his labor.

Dionysios or Kallista owed a debt. Then their sons should repay the debt.

Dionysios or Kallista owed a debt and one of the sons denied to repay the debt. Then he should compensate 1000 drachmas.

It is possible also to draw more complex conclusions by applying more complex inference rules.

There are 52 Greek papyri found in Elephantine and dated from 310 B.C. to 613 C.E. If we say about legal documents like the testament cited above, then their structure are taken and borrowed from the Aramaic documents which were much earlier there: date, parties, the crux of the matter by means of implications, witnesses ($\sigma\upsilon\gamma\gamma\rho\alpha\varphi o\varphi\iota\lambda\alpha\xi$, 'keeper of contracts'), and sealings.

7 Conclusions

The Ancient Greek logic was not the first. This logic was invented within the Semitic legal culture. For instance, the *Law Code of Gortyn* [27], the only preserved code of the Greeks, was written in the way of the *Code of Hammurabi* with many citations from Semitic codes [25]. Hence, we can assume that the Greek logic is a continuation of the Semitic one. In fact, we know only four logical systems of the Ancient Greeks: (i) the Aristotelian syllogistic (Aristotle's *Prior Analytics*); (ii) the Aristotelian modal logic (Aristotle's *On Interpretation*); (iii) the Stoic propositional logic (Hans Von Arnim's *Stoicorum Veterum Fragmenta*); (iv) the Stoic modal logic (Cicero's *On the Nature of the Gods*, *On Divination*, *On Fate*). These logical systems can be formalized within the modern symbolic logic. Nevertheless, we have no evidences how the Greeks used them for inferring. Even in the Aristotle's texts and in the Stoic fragments there are no evidences of applying these systems in long deductions. In other words, Aristotle and the Stoics did not propose their systems in an axiomatic form (where axioms are given in codes or by facts confirmed by courts). In contrast, in the Neo-Babylonian trial records we face really long deductions within an axiomatic form. It is a direct evidence for the following: (i) at least at the time of the Neo-Babylonians there existed an axiomatic system described in this paper, but so probably that this system was invented from the very beginning of the Sumerian-Akkadian legality; (ii) this axiomatic system was implemented in justice to serve the people.

Recently, the European Commission has formulated e-justice (electronic justice) as a promising way of developing the open society. But the Neo-Babylonians had implemented such an e-justice more than 2,500 years ago very efficiently. Let us remember what e-justice is. It is thought up to use technology, information and communication to improve access of citizens to justice and to make judicial action more effective. In other words, e-justice is a logical formalization of justice as such allowing us to implement different expert systems to make justice effective and transparent.

Thus, the Neo-Babylonians invented a symbolic logic in an axiomatic form to make justice effective and transparent. In this paper, I have considered this logic as closer to their trial records as possible, but their axiomatic system can be formalized differently, including the form of sequent calculus.

For trial deductions of the Neo-Babylonians we can define the following sequent calculus. Let $S_1.S_2.S_3.\cdots.S_n \hookrightarrow P_1.P_2.P_3.\cdots.P_n$ mean that $(S_1 \wedge S_2 \wedge S_3 \wedge \cdots \wedge S_n) \Rightarrow (P_1 \vee P_2 \vee P_3 \vee \cdots \vee P_n)$ is true, where $S_1, S_2, S_3, \ldots, S_n, P_1, P_2, P_3, \ldots, P_n$ are metavariables running over formulas defined in Section 4 for formalizing Neo-Babylonian legal proceedings. The only axiom of this sequent calculus is $S \hookrightarrow S$. We can use a standard system of natural deduction with the following additional inference rules:

$$\frac{A \Rightarrow B.S[B] \hookrightarrow P[B]}{S[A] \hookrightarrow P[A]};$$

$$\frac{A \Rightarrow \neg B.S[B] \hookrightarrow P[B]}{S[\neg A] \hookrightarrow P[\neg A]}.$$

Thus, the symbolic logic in the form of axiomatic system existed before Aristotle and the Stoics and it was invented within the Sumerian-Akkadian legal culture [25].

References

[1] Abraham, M., Belfer, I., Gabbay, D., Schild, U. *Fuzzy Logic and Quantum States in Talmudic Reasoning*. London: College Publications, 2015.

[2] Abraham, M., Gabbay, D., Hazut, G., Maruvka, Y.E., Schild, U. *The Textual Inference Rules Klal uPrat How the Talmud defines Sets*. London: College Publications, 2010.

[3] Abraham, M., Gabbay, D., Schild, U. *Non-deductive Inferences in the Talmud*. London: College Publications, 2010.

[4] Abraham, M., Gabbay, D., Schild, U. *Principles of Talmudic Logic*. London: College Publications, 2013.

[5] Albright, W.F. Palestinian Inscriptions: A Letter from the Time of Josiah, in: *Ancient Near Eastern Texts Relating to the Old Testament*. Princeton: Princeton University Press, 1969.

[6] Botta, A. F. *The Aramaic and Egyptian legal traditions at Elephantine: an Egyptological approach*. London; New York: T&T Clark, 2009.

[7] Cicero, M. Tullius. *The Orations of Marcus Tullius Cicero*, literally translated by C. D. Yonge, B. A. London. George Bell & Sons, York Street, Covent Garden. 1891.

[8] Ciceronis, M. Tulli. *Topica*. Edited with a translation introduction, and commentary by Tobias Reinhardt. Oxford University Press, 2003.

[9] Cowley, A. E. *Aramaic Papyri of the Fifth Century B.C.* Oxford: Clarendon Press, 1923.

[10] Gabbay, U. *The exegetical terminology of Akkadian commentaries*. Series: Culture and history of the ancient Near East. Vol. 8. Brill: Leiden, Boston, 2016.

[11] Holtz, Shalom E. *Neo-Babylonian court*. Series: Cuneiform monographs. Brill: Leiden, Boston, 2009.

[12] Holtz, Shalom E. *Neo-Babylonian Trial Records*. Society of Biblical Literature Atlanta, 2014.

[13] Joannès, F. (ed.) *Rendre la justice en Mésopotamie. Archives judiciaires du Proche-Orient ancien (IIIe-Ier millénaires avant J.-C.)*. Saint-Denis: Presses Universitaires de Vincennes, "Temps et espaces", 2000.

[14] Lenzi, A. *Akkadian prayers and hymns: a reader*. Series: Ancient near east monographs. Atlanta: Society of Biblical Literature, 2011.

[15] Poebel, A. The Beginning of the Fourteenth Tablet of Harra Hubullu, *The American Journal of Semitic Languages and Literatures*, 52(2), 1936, pp. 111-114.

[16] Porten, B. *The Elephantine Papyri in English. Three Millennia of Cross-Cultural Continuity and Change*. Second Revised Edition. Society of Biblical Literature. Atlanta, 2011.

[17] Pritchard, J. B. (ed.) *Ancient Near Eastern Texts Relating to the Old Testament with Supplement*. Princeton: Princeton University Press, 1969.

[18] Rochberg, F. *In the path of the Moon: Babylonian celestial divination and its legacy*. Series: Studies in ancient magic and divination. Brill: Leiden, Boston, 2010.

[19] Roth, M. T. *Law collections from Mesopotamia and Asia Minor*. With a contribution by Harry A. Hoffner, Ir.; edited by Piotr Michalowski. Scholars Press Atlanta, Georgia. 1995.

[20] Schumann, A. (ed.) *Judaic Logic*. Gorgias Press, 2010.

[21] Schumann, A. (ed.) *Philosophy and History of Talmudic Logic*. London: College Publications, 2017.

[22] Schumann, A. (ed.) *Pragmatic Studies in Judaism*. Gorgias Press, 2010.

[23] Schumann, A. Preface, *History and Philosophy of Logic*, 32(1):1–8, 2011.

[24] Schumann, A. *Talmudic Logic*. London: College Publications, 2012.

[25] Schumann, A. On the Babylonian Origin of Symbolic Logic, *Studia Humana*, 6(2): 126–154, 2017.

[26] Smelik, K. *Writings from Ancient Israel*, Westminster: John Knox Press, 1991.

[27] *The Law Code of Gortyn*. Edited by Willetts, Ronald F. De Gruyter, 1967, pp. 37–50.

[28] Veenhof, Klaas R. In Accordance with the Words of the Stele: Evidence for Old Assyrian Legislation, *Chicago-Kent Law Review*. Volume 70, Issue 4, 1995, pp. 1717–1744.

[29] Westbrook, R. *Law from the Tigris to the Tiber: the writings of Raymond Westbrook. Raymond Westbrook*; edited by Bruce Wells and Rachel Magdalene. Volume 1 The Shared Tradition. Indiana: Eisenbrauns Winona Lake, 2009.

 Received 6 October 2017

Weight Prediction on Missing Links in Social Networks – a Cross-Entropy-Based Approach –

Wilhelm Rödder
FernUniversität in Hagen, Department of Operations Research
wilhelm.roedder@fernuni-hagen.de

Andreas Dellnitz
FernUniversität in Hagen, Department of Quantitative Methods
andreas.dellnitz@fernuni-hagen.de

Ivan Gartner
Faculdade de Economia, Administração e Contabilidade, Universidade de Brasília
irgartner@hotmail.com

Sebastian Litzinger
FernUniversität in Hagen, Department of Quantitative Methods
sebastian.litzinger@fernuni-hagen.de

Abstract

A social network (SN) is a group of actors and their mutual relations. Sociologists try to answer the question why networked actors in our society are more successful than others and how this networking works. Directed or undirected graphs, hyper- or multigraphs are a suitable means to visualize social relations. Social networks with directed and weighted links among actors need sophisticated instruments for analyses. We model these links as probabilistic conditioned propositions. Then for any actor i the model permits the estimation of transfer probabilities to all actors j, may they be linked to i or not. When future sociological research wants to interconnect missing links, some of the respective weights cannot be chosen at will but must fall in certain intervals. They must be in accordance with former conditional-logical net structure.

We are grateful to the anonymous reviewer whose comments and suggestions helped greatly improve the quality of our paper and enhanced the presentation of our results.

To achieve this goal, cross-entropy-driven knowledge bases are applied. For a middle-size network we demonstrate the new findings.

Keywords: social networks; weight prediction; conditionals; cross-entropy; missing links

1 Introduction

1.1 Motivation

A social network (SN) colloquially means a group of actors and their mutual relations. Sociologists try to answer the question why networked actors in our society are more successful than others and how this networking works. Already in 1895 Émile Durkheim stated that “the whole is more than the sum of its parts” [6]. Social network analysis is the analysis of this “more”. A good survey of historical developments in the field give Jansen [10] and Scott [28].

Jakob Moreno migrated from Vienna to the United States in 1925 and in the following years studied what he called “social configurations”. His principal innovation was the sociogram, a graphical representation of the social fabric [18]. Centrality, prestige, influence and power of actors since then can be measured by graph theoretical indices, cf. again [10] and [28]. The more mathematically oriented reader finds further details in Newman’s compendium “Networks” [19].

Directed or undirected graphs, hyper- or multigraphs are a suitable means to visualize social relations. Social networks with weighted links came up already in the 1950s, cf. Katz [12]. The idea was continued and deepened by Bonacich [1] and Bonacich and Lloyd [2]. The transfer of messages, attitudes or votes might be attenuated; attenuation is “the force of a probability of effectiveness of a single link” [12, p. 41]. And already Katz noticed that this attenuation might not be the same on all links.

In this paper, we focus on directed and weighted graphs, the weights being such probabilities of effectiveness. Borgatti [3] relates on possible flows in networks: goods, money, messages, e-mails, attitudes, infections, information, among others. All these flows might be subject to certain types of losses: goods might rot, money might disappear, messages, e-mails or information might be affected by noise, the transfer of attitudes or infections might be attenuated.

The mathematical treatment of such losses is not easy, however:

- What is the likelihood of actor j to receive a message via a path from a distant actor i? Is it just the product of attenuation factors along the path? If it is not, there might be latent links between some actors, which contradict such result.

- How to treat the even more complicated case for more than one path connecting the actors and again with possible hidden links?

- What is the transfer from i to j in a highly meshed network?

Answering these questions in static social networks is difficult, and even more difficult in dynamic networks. Currently latent links might be revealed, previously observed links and/or vertices might disappear. Recent literature of the social network community has begun to focus on these issues. Here, two mainstreams are the prediction of hitherto missing links and/or the prediction of weights on missing links. In this article, we study an axiomatically justified unbiased form of probabilistic weight prediction in directed networks. Before doing so, we give a literature overview on the aforementioned prediction problems.

1.2 Related work

In Social Network Analysis (SNA) the field of link prediction and weight prediction has grown extensively over the last two decades. Here, the two dominating questions are where to date missing links might appear in a dynamical context and/or which weights might appear.

We will now sketch relevant literature in a chronological order. In [41] the authors present a link prediction application for web sites using Markov chains. Here, link prediction is based on the navigational behavior of visitors, where historic transitions are used to recommend the next step in their digital journey. [16] offers a more traditional approach to link prediction. The authors seek to predict future co-authorships in research contexts. To this end, they consider a training period and a test period and compare various proximity-based prediction scores. In [14] the authors provide a framework for estimating graph parameters via the transformation of a graph's algebraic spectrum. Their methods are applicable to undirected, unweighted, weighted, and bipartite graphs. The authors in [39] validate 9 similarity scores for link prediction employing the well-known accuracy measure AUC. Furthermore, they suggest a generalization of similarity scores in order to improve prediction accuracy. In [17] it is examined whether weight-based similarity scores increase the accuracy of link prediction. In [38] the authors grab this idea once more and apply it to link weight prediction this time. Are the aforementioned methods useful in weighted multiplex networks? To answer this question, the precision of respective algorithms for missing links in a target layer is determined in [29].

Combining classical similarity scores with information-theoretical indices leads to a sophisticated link prediction tool as shown in [40]. A survey of similarity-based link prediction algorithms is provided in [34].

An up-and-coming representation of networks is based on graphs but models an arrow as conditioned proposition: if – then. Whenever a message, an immaterial good or an attitude is to be transferred in a net, such modeling is convenient [25] [22] [8]. A recent paper also considers uncertain transfers [24]. Such uncertain transfers on net links in our context are realized by probabilistic conditionals, see again [12]. In this article, we focus on the calculation of conditional probabilities also for missing links, i. e. probabilistic weight prediction. To ascertain the resilience of such predictions, one needs a sophisticated concept, the concept of Minimum Cross-Entropy (ME).

Section 2 introduces syntax and a probabilistic concept of relations in networks (2.1), shows indeterminacy intervals on missing links for some small examples (2.2), and develops the ME concept for networks (2.3). 2.4 provides an algorithm for calculating indeterminacy intervals for bigger nets. Section 3 performs SNA under ME for a Kronecker network and Section 4 gives a résumé and depicts possible steps of future research.

2 Preliminaries

2.1 Syntax, net frame and net load

In order to transparently present our cross-entropy-based approach we need to rely on certain concepts, which will be introduced in this section.

We consider a set of n actors $a_1, \ldots, a_n$. To each actor a_i there is attached a binary variable V_i with attributes $V_i = v_i$ and $v_i = i/\bar{i}$, $v = (v_1, \ldots, v_n)$ are respective configurations. For pairs of actors, $V_j = j \mid V_i = i$ are called conditionals, sometimes we write $V_j \mid V_i$ or just $j \mid i$ for short; $\mid$ is the conditional operator. For a substantial discussion of conditionals see [5] or [22]. The semantics is as follows. $V_i = i/\bar{i}$ – or $i/\bar{i}$ for short – stands for the proposition that a_i knows the message/does not know it. Conditionals enable possible message transfer: if a_i is informed, then probably a_j also is. To make further developments more intelligible we start with a theoretical construct, a so-called *net frame*. In this net frame all direct transfers between pairs of actors are allowed; i. e. we allow all $V_j = j \mid V_i = i$ for $i, j = 1, \ldots, n$ and $i \neq j$. Graph theoretically speaking this frame is a complete directed graph; between each pair of actors there are back and forth arrows.

A sociological survey provides transfer probabilities (TPs) p_{ij} for some pairs (a_i, a_j), for others it does not – resulting in a (weighted) social network embedded in the net frame. Thus the net consists of $N \subseteq \{1, \ldots, n\} \times \{1, \ldots, n\}$ conditionals plus respective probabilities:

$$V_j = j \mid V_i = i \text{ with } p_{ij} \text{ for } (i, j) \in N. \tag{1}$$

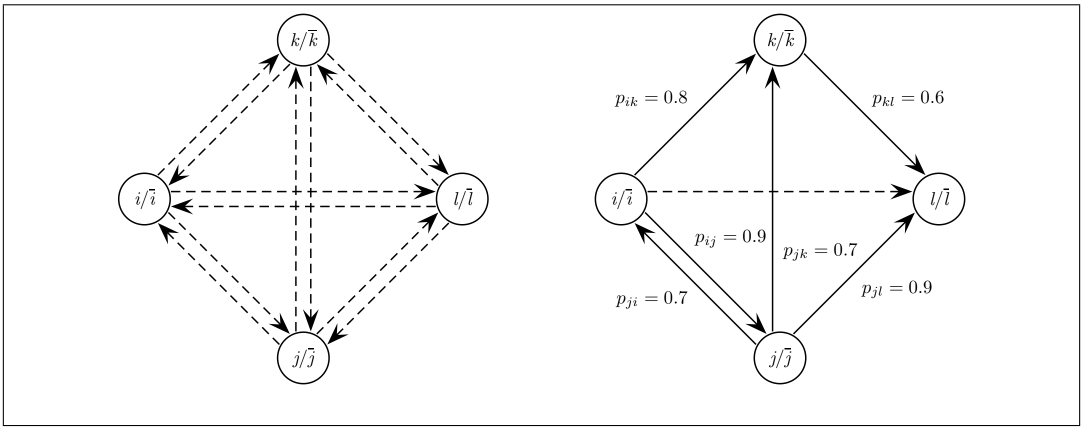


Figure 1: Net frame and social network; possible transfers ($\dashrightarrow$) and transfers ($\longrightarrow$) within the SN.

Figure 1 visualizes net frame (left), and a net (right). A message can flow from actor a_i via a_j or a_k to a_l. While the link from a_i to a_l is currently missing, a later sociological research might find out that a_i can inform a_l directly, however. This paper includes such possible future net interconnections. It tries to answer the question what is our knowledge about TPs of to date missing links. Are these TPs fixed by the actual net structure or are they still undetermined?

These questions could be answered if we had a global probability distribution Q on the set of all configurations $\mathcal{V} = \{v\}$. This distribution would permit the calculation of all conditional probabilities in the net and hence all transfer probabilities for missing links. And each such transfer probability is either certain, uncertain, or null.

Therefore, we seek for a Q on $\mathcal{V}$ which respects the p_{ij}:

$$\mathsf{Q}(V_j = j \mid V_i = i) = p_{ij} \text{ for } (i, j) \in N. \tag{2}$$

Such a global distribution Q we call *net load*.

Q is not fully determined, when the social fabric and its structure are surveyed only partly. For some links the probabilities are known, other links just do not exist. Instead of one Q we might now have a set of possible Qs which are compatible with (2). Consequently, not all transfer probabilities from actors to actors are exact numbers but rather ambiguous: indeterminacy instead of uncertainty.

2.2 Weights on missing links: elementary considerations

To make the concepts of uncertainty and indeterminacy more transparent, we study the following example.

Example 1.

i) In Figure 2, $\mathsf{Q}(j \mid i) = p_{ij} = 1$ and $\mathsf{Q}(k \mid j) = p_{jk} = 1$ implies $\mathsf{Q}(k \mid i) = 1$ for all Q. Evidently, such TPs are transitive.

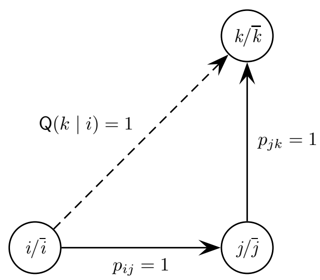


Figure 2: Fully determined TP

ii) To estimate the transfer probability for the missing link from a_i to a_k in Figure 3 solve the following two optimization problems:

$$l = \min\ \mathsf{Q}(k \mid i) \text{ s.t. } \mathsf{Q}(j \mid i) = 1, \mathsf{Q}(k \mid j) = 0.8 \text{ and}$$
$$u = \max\ \mathsf{Q}(k \mid i) \text{ s.t. } \mathsf{Q}(j \mid i) = 1, \mathsf{Q}(k \mid j) = 0.8.$$

The optimal values of the objective functions are $l = 0$ and $u = 1$, respectively; the unknown TP from a_i to a_k can still vary in the interval $[l = 0,\ u = 1]$, and thus is absolutely indetermined.
For more details on these two fractional programming problems cf. Appendix A.

iii) Solving analogously for the net in Figure 4 the equations

$$l = \min\ \mathsf{Q}(k \mid i) \text{ s.t. } \mathsf{Q}(j \mid i) = 0.8, \mathsf{Q}(k \mid j) = 1 \text{ and}$$
$$u = \max\ \mathsf{Q}(k \mid i) \text{ s.t. } \mathsf{Q}(j \mid i) = 0.8, \mathsf{Q}(k \mid j) = 1$$

yields $l = 0.8$, $u = 1$. The transfer probability for the missing link from a_i to a_k is partially indetermined but lies in $[0.8, 1]$.

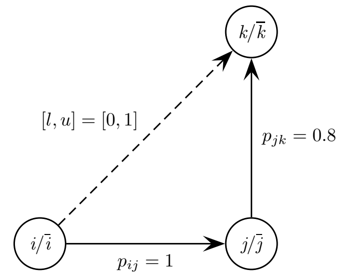


Figure 3: Absolutely indetermined TP

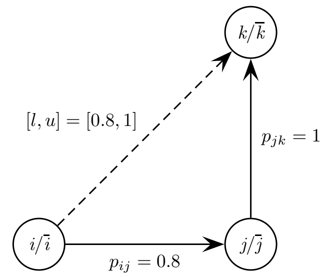


Figure 4: Partially indetermined TP for three actors

iv) For the net in Figure 5 the solution of the optimization problems

$$\begin{aligned} l = \min\ \mathsf{Q}(l \mid i) \text{ s. t. } & \mathsf{Q}(j \mid i) = 0.9, \mathsf{Q}(k \mid i) = 0.8, \mathsf{Q}(i \mid j) = 0.7, \\ & \mathsf{Q}(k \mid j) = 0.7, \mathsf{Q}(l \mid j) = 0.9, \mathsf{Q}(l \mid k) = 0.6 \text{ and} \\ u = \max\ \mathsf{Q}(l \mid i) \text{ s. t. } & \mathsf{Q}(j \mid i) = 0.9, \mathsf{Q}(k \mid i) = 0.8, \mathsf{Q}(i \mid j) = 0.7, \\ & \mathsf{Q}(k \mid j) = 0.7, \mathsf{Q}(l \mid j) = 0.9, \mathsf{Q}(l \mid k) = 0.6 \end{aligned}$$

yields $l = 0.77$ and $u = 1$, resulting in a TP interval [0.77, 1].

v) A Q respecting all given link probabilities like in Figure 6 must satisfy

$$\mathsf{Q}(j \mid i) = \mathsf{Q}(i \mid j) = 1, \mathsf{Q}(k \mid i) = 0.7, \mathsf{Q}(k \mid j) = 0.8.$$

As one would expect when inspecting the network's graphical representation, the inconsistent constraints yield $\mathsf{Q}(i) = \mathsf{Q}(j) = 0$ and hence $\mathsf{Q}(k \mid i) = \mathsf{Q}(k \mid j) = \frac{0}{0}$.

The results of Example 1 need some comments, we feel. The transitivity in i) meets our intuition and the easy proof is left to the reader. At least for these authors

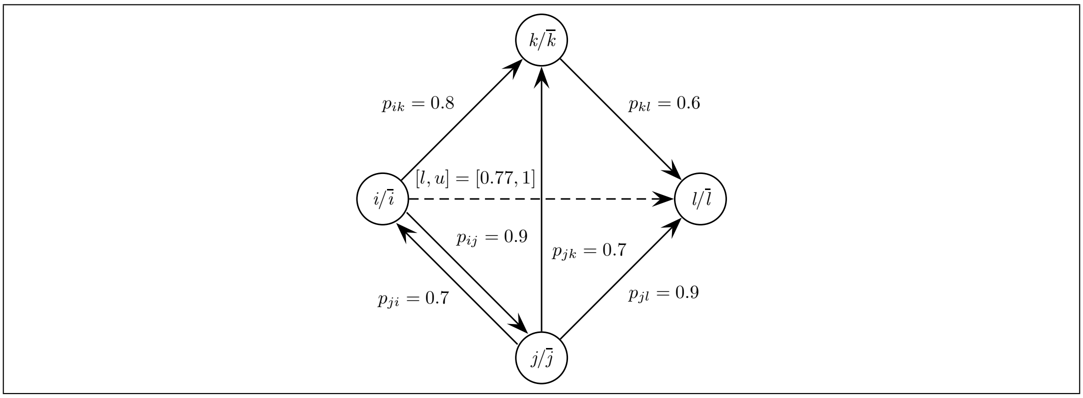


Figure 5: Partially indetermined TP for four actors

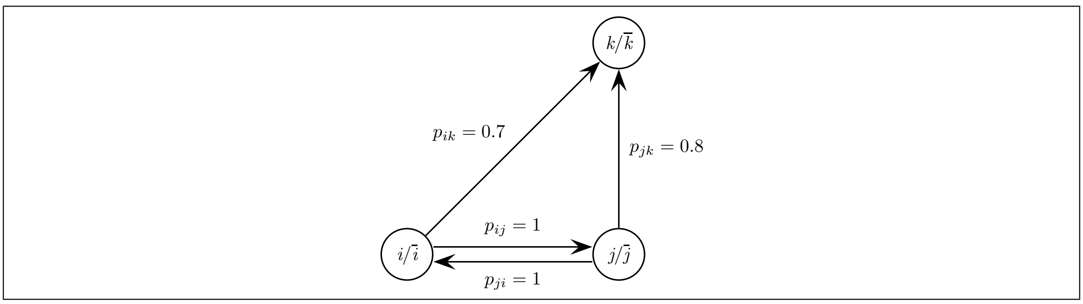


Figure 6: Inconsistent TPs

the results of ii) to iv) were not self-evident when we solved the respective optimization problems. Despite certain transfer from a_i to a_j and an 80% transfer from a_j to a_k the direct transfer probability from a_i to a_k is still absolutely indetermined and can vary in the interval $[0, 1]$. The respective statement for iii) is less viewy but still unexpected and guessing the TP from a_i to a_l in iv) is utterly impossible. Example v) shows that an imprudent assignment of TPs can cause contradictions.

These observations give rise to the question whether there might be a more precise estimation of TPs than the ones in Example 1. There is such an estimation using the concept of minimum relative entropy. More on that in the next section.

2.3 ME load

Entropy is well established in SNA, not so relative or cross-entropy. Entropy is an information-theoretical concept which accompanied us in thermodynamics, in communication theory, and in artificial intelligence. For a short overview, see Section

2 in [4, p. 2]. Many authors employed entropy to analyze social networks, see [36], [20], [9], [37], [30], [11], just to mention a few.

Cross-entropy as an instrument for analyzing networks is applied less frequently, [26], [4]. Cross-entropy stems from Shannon's information theory. Whereas for an optimal coding entropy measures the average information of signal transmission from a sender to a receiver, cross-entropy measures the information gain for the correct against a wrong coding [35, pp. 50–51]. The optimal code is a function of the signals' frequency and if this frequency changes, the code must be adapted to the new situation in order to realize this gain. It is called cross-entropy, cross-entropy thus is a dissimilarity measure for different frequencies. With the "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy" by [32] the concept became admissible at court also in artificial intelligence, see also [31]. In "Characterizing the principle of minimum cross-entropy within a conditional-logical framework" [13] the author combines the ME concept with a conditional-logical framework. Formulating four axioms she develops cross-entropy as the only functional to adapt a probability distribution to new conditioned information, preserving the prior distribution as far as possible under adaptation.

$$\begin{aligned} &\mathsf{Q}^*(\mathsf{P}) = \arg\min R(\mathsf{Q},\mathsf{P}) = \sum_v \mathsf{Q}(v)\log_2\frac{\mathsf{Q}(v)}{\mathsf{P}(v)} \\ &\text{s. t. } \mathsf{Q}(V_j = j \mid V_i = i) = p_{ij},\ (i,j)\in N. \end{aligned} \tag{3}$$

Here P is the prior distribution, $\mathsf{Q}(V_j = j \mid V_i = i) = p_{ij}, (i,j) \in N$ is new conditional information and $R(\mathsf{Q},\mathsf{P})$ is cross-entropy from P to Q; cross-entropy is not symmetrical.

$\mathsf{Q}^*(\mathsf{P})$ in (3) is the adaptation of P to the $p_{ij},\ (i,j) \in N$. If $\mathsf{P} = \mathsf{P}^0$, the uniform distribution, we simply write Q^* instead of $\mathsf{Q}^*(\mathsf{P}^0)$. Q^* then is the only unbiased adaptation of the uniform P^0 to the actual known network structure. P^0 on $\mathcal{V}$ represents ignorance; all transfer (and not transfer) probabilities in the net frame are still unknown. Hence Q^* preserves this ignorance as far as possible respecting the new conditional structure. Consequently, the therefrom derived weight predictions of missing links are mandatory in this conditional-logical framework.

As is well-known, minimization like in (3) for $\mathsf{P} = \mathsf{P}^0$ is equivalent to maximization of entropy $H = -\sum_v \mathsf{Q}(v) \cdot \log_2 \mathsf{Q}(v)$. So Q^* might be called maximum entropy load as well as ME load of the network.

For solving (3) software tools like [27], [7], [33] are available. For further calculations we use the latter, called SPIRIT. To make things more transparent we revisit Example 1 from Section 2.2 and calculate weight predictions under ME.

Example 1. (continued)

Ad i) As $\mathsf{Q}(k \mid i) = 1$ for all Q we also have $\mathsf{Q}^*(k \mid i) = 1$. Transitivity also holds under ME adaptation.

Ad ii) Solving

$$\mathsf{Q}^* = \arg\min R(\mathsf{Q}, \mathsf{P}^0) \text{ s.t. } \mathsf{Q}(j \mid i) = 1,\ \mathsf{Q}(k \mid j) = 0.8$$

and then calculating the TP from a_i to a_k yields $\mathsf{Q}^*(k \mid i) = 0.8$. For a detailed formulation of the optimization problem see Appendix B.

Ad iii) Solving

$$\mathsf{Q}^* = \arg\min R(\mathsf{Q}, \mathsf{P}^0) \text{ s.t. } \mathsf{Q}(j \mid i) = 0.8,\ \mathsf{Q}(k \mid j) = 1$$

and then calculating the TP from a_i to a_k yields $\mathsf{Q}^*(k \mid i) = 0.9$.

Ad iv) Solving

$$\begin{aligned}\mathsf{Q}^* = \arg\min R(\mathsf{Q}, \mathsf{P}^0) \text{ s.t. } & \mathsf{Q}(j \mid i) = 0.9,\ \mathsf{Q}(k \mid i) = 0.8,\ \mathsf{Q}(i \mid j) = 0.7,\\ & \mathsf{Q}(k \mid j) = 0.7,\ \mathsf{Q}(l \mid j) = 0.9,\ \mathsf{Q}(l \mid k) = 0.6\end{aligned}$$

and then calculating the TP from a_i to a_l yields $\mathsf{Q}^*(l \mid i) = 0.84$.

Ad v) $\mathsf{Q}^* = \arg\min R(\mathsf{Q}, \mathsf{P}^0)$
s.t. $\mathsf{Q}(j \mid i) = \mathsf{Q}(i \mid j) = 1$, $\mathsf{Q}(k \mid i) = 0.7$, $\mathsf{Q}(k \mid j) = 0.8$
yields a system error due to the inconsistent restrictions (see above).

To analyse these results we compare them to those of Example 1 in Section 2.2. Now the former indeterminacy in ii) – iv) disappears, the TPs are fixed numbers. They are point estimates rather than indeterminacy intervals. They are the unbiased weights for respective missing links under ME. For an up to now missing link, the best weight in this ME environment is the respective fixed number. Please note that this number always is an element of the respective indeterminacy interval, which was calculated by fractional programming as shown in Section 2.2.

The knowledge of Q^* – the proper knowledge base – permits the calculation of unbiased weights TP for all missing links $j \mid i$ in the network. To turn the argument on its head, it also permits an unbiased estimation of actors a_j's reception probability (RP) once a_i sends a message or a good, of course. And SPIRIT offers a user-friendly feature for these calculations, see Section 3.

Independent of the direction of the bifocal perspective – TP or RP – the resilience of these probabilities is still a problem. As we know from Example 1 in Section 2.2 and Example 1 (continued) in the current section, point estimates are still accompanied by indeterminacy on missing links. How to tackle this problem even for bigger nets is topic of the next section.

2.4 Indeterminacy intervals in SPIRIT

The point estimates of transfer probabilities for missing links are mandatory in our ME framework, we said. Nevertheless, the knowledge of indeterminacy intervals on missing links is advantageous for the analyst. Whenever a future sociological research reveals a new link, this research also might provide an empiric prediction of the respective transfer probability. But is this prediction in accordance with the former net? Does the prediction fall in the Indeterminacy Interval (II)? If yes, it is a welcome refinement of the net structure. If not, it is conflicting with earlier knowledge and must lead to a recheck of all hitherto collected data. Due to the fact that future TPs are accepted or rejected based on the IIs, the algorithm's accuracy is of paramount importance for our method; the determination of the IIs is 100 % exact.

For the little nets of Example 1 in Section 2.2 we determined the respective IIs by means of fractional programming, for bigger nets this is impossible. Problem ii) showed up 8 decision variables – see Appendix A – for the 64 actors of Section 3 there are $2^{64} \approx 1.84 \cdot 10^{19}$ variables. This makes fractional programming infeasible.

The expert system shell SPIRIT allows for a fast and effective calculation of IIs by means of cross-entropy-driven knowledge bases. More details you find in [25] or [21]. The following algorithm shows the gist of the method.

II for the conditional $\boldsymbol{V_j = j \mid V_i = i}$

1. Solve (3) for $\mathsf{P} = \mathsf{P}^0$ with the only restriction $\mathsf{Q}(V_j = j \mid V_i = i) = \epsilon$ for a sufficiently small $\epsilon > 0$. Result $\overline{\mathsf{Q}}^*$.

2. Solve (3) for $\mathsf{P} = \overline{\mathsf{Q}}^*$. Result $\overline{\mathsf{Q}}^{**}(\overline{\mathsf{Q}}^*)$.

3. Make $ll = \overline{\mathsf{Q}}^{**}(V_j = j \mid V_i = i)$.

4. Solve (3) for $\mathsf{P} = \mathsf{P}^0$ with the only restriction $\mathsf{Q}(V_j = j \mid V_i = i) = 1 - \epsilon$ for a sufficiently small $\epsilon > 0$. Result $\overline{\overline{\mathsf{Q}}}^*$.

5. Solve (3) for $\mathsf{P} = \overline{\overline{\mathsf{Q}}}^*$. Result $\overline{\overline{\mathsf{Q}}}^{**}(\overline{\overline{\mathsf{Q}}}^*)$.

6. Make $ul = \overline{\overline{\mathsf{Q}}}^{**}(V_j = j \mid V_i = i)$.

7. $[ll, ul]$ is the II.

In this algorithm global distributions are generated which attribute probability $\epsilon \approx 0$ and accordingly $(1 - \epsilon) \approx 1$ to the conditional under consideration $V_j = j \mid V_i = i$, first. The respective distributions then are adapted to the TPs provided by the sociologist. These adaptions maintain ϵ and $1 - \epsilon$, respectively, as far as possible and hence yield lower limit (ll) and upper limit (ul) of the indeterminacy interval II for this conditional. This is fractional goal programming in an ME environment.

In the next section we show some IIs in a medium size network.

3 ME-analyses in a middle-size network

To demonstrate some of the results developed so far we opted for a synthetic network with 64 actors and 107 weighted directed links, as it permits a clear visual representation of results like in Figures D.1 and D.2.

The procedure of network construction is as follows:

- We construct a stochastic Kronecker graph with the 2×2 initiator matrix $\left(\begin{smallmatrix} .8 & .6 \\ .5 & .3 \end{smallmatrix}\right)$ and make 5 iterations. This generates a 64×64 Kronecker matrix.

- In a second step we determine an adjacency matrix using uniform random numbers. Make a directed link if and only if the respective entry in the Kronecker matrix exceeds the random number. In our case, this yielded 107 directed links, three of the 64 actors were isolated.

Kronecker graphs are a sophisticated tool for emulating real social structures. The specific choice of the initiator matrix guarantees that characteristics of real world social networks can be found in our example as well. For more details on this network construction concept cf. [15, Sections 3.3 – 3.5].

Now that the actors and the directed links are determined, transfer probabilities TP must be assigned to all existing links. We opted again for uniform random numbers.

- For each of the 107 links $j \mid i$ attach a uniform random number and make it the transfer probability. This means, we suppose that link weights are independent of network topology.

Once the shell SPIRIT is informed about all these parameters, it takes less than one second to build Q^* on $\mathcal{V}$. Table C.1 in Appendix C lists the 107 conditionals and respective TPs. Figure D.1 in Appendix D shows the 64 actors as (rectangular) nodes v1 to v64. For each node vi the system provides the respective marginal distribution of Q*. The bars visualize these values. Note that in node vi the 1-bar stands for $V_i = i$ and the 0-bar for $V_i = \bar{i}$. Because of space restrictions we "open" only nodes v9 and v43 and leave the rest "closed".

Now the system is ready to calculate the TPs Q*$(j \mid i)$ also for unlinked nodes vi, vj. A mouse click on the 1-bar of node vi makes its probability equal 1 and the probability of the 0-bar equal 0. SPIRIT executes a proportional fitting regarding the new marginal. It hence provides Q*$(j \mid i)$ for all vj, linked to vi or not. This option of the shell therefore permits a comfortable calculation of all unbiased weights even for missing links. These point estimates are mandatory in our conditional-logical framework, see again Section 2.3. We show the result for the two "opened" nodes v9 and v43 in Figure D.2 in Appendix D. Clicking on the 1-bar in v9 alters the marginal probabilities – on the 1-bar to 1 and on the 0-bar to 0. In v43 we find a 0.85 probability of actor a_{43} to receive the message if a_9 sends it.

Such point estimates now shall be complemented by respective indeterminacy intervals. For selected couples of actors Table 1 shows sender, receiver, transfer probability, and indeterminacy interval.

Sender	Receiver		Receiver		Receiver	
	TP	II	TP	II	TP	II
v6		v58		v9		v11
	0.47	[0.47,0.47]	0.38	[0.27,0.38]	0.46	[0.34,1]
v6		v25		v50		v60
	0.42	[0.16,0.79]	0.09	[0,0.27]	0.53	[0,1]
v9		v39		v43		v42
	0.04	[0.04,0.04]	0.85	[0.84,1]	0.16	[0.04,0.56]
v9		v3		v17		v21
	0.56	[0.05,1]	0.17	[0,0.54]	0.49	[0,1]
v1		v61		v6		v27
	0.75	[0.75,0.75]	0.75	[0.54,1]	0.57	[0.2,0.95]
v1		v49		v50		v58
	0.02	[0,0.07]	0.14	[0,0.33]	0.45	[0,1]

Table 1: Indeterminacy intervals for selected pairs of actors

For sender v9 this reads:

- The conditional $V_{39} \mid V_9$ [0.04] in Table C.1 fixes the transfer probability from actor a_9 to actor a_{39} and hence there is no indeterminacy at all.

- From actor a_9 to actor a_{43} there exists a directed path via actor a_{33}. This does not fix the transfer probability for the missing link but at least keeps indeterminacy in the interval $[0.84, 1]$.
- From actor a_9 to actors a_{42} or a_3 the connections are not clearly arranged, resulting in indeterminacy intervals $[0.04, 0.56]$ and $[0.05, 1]$.
- Finally, missing links to a_{17} or a_{21} would permit TPs in the intervals $[0, 0.54]$ and $[0, 1]$, respectively, the latter being the indeterminacy maximum.

To repeat the gist of the matter: A link from a_9 to a_{43} – missing in the current network – *must* have a TP within $[0.84, 1]$, otherwise the net becomes inconsistent. Indeterminacy intervals are a powerful instrument for the analyst to control dynamic structuring of the network. In this process, (missing) link weights are not just numbers in a descriptive model but permit sophisticated calculations in an explicatory model and consequently are a new kind of SNA.

4 Résumé and the road ahead

Networks appear in many different scientific fields: biology, chemistry, informatics, telecommunication engineering, medicine and sociology, among others. Sociology scrutinizes the advantage an actor or a group of actors take out of the social network. A modern form of representing such nets are graphs. Graph theoretical indices then permit the characterization of the actors' position, prestige, power, embeddedness etc. in the net.

Recent developments are based on propositional logics. An arrow in the graph now becomes a conditional proposition: if actor i knows a message or has a certain attitude, then also actor j does. And this with a suitable probability, the link weight.

With ongoing sociological research into the social fabric, to date missing links come to the fore and their weights must be consistent with earlier knowledge about the net.

The here presented method offers cross-entropy-driven unbiased estimates for weights of missing links. And above all it informs the researcher about limits in which these coefficients must fall. In a highly netted social fabric future link weights are not at will but rather dependent on former net structure. After theoretical developments a synthetic middle-size network of 64 actors and 107 links, which displays essential properties of real-world networks, undergoes such analyses. A next step will be to employ our method for estimating message transfer probabilities in social media.

If link weights in a net are not consistent with each other, messages or attitudes cannot flow; the model is infeasible. Is there a way to assist the model builder in the complicated process of finding a feasible model? To answer this question is also one of our future intents.

Net structure is not always static but rather dynamic. What happens when actors disappear or links fail? What are the consequences for the remaining net? The study of this issue for weighted nets is promising.

Actors can be humans, animals, technical objects or – of course – corporations. Entropy-based analyses of banking activities pledge interesting results. For first steps in that direction cf. [8], [23].

Appendix A

To determine the lower bound of the TP interval minimize the conditional probability $\mathsf{Q}(k \mid i)$ given the constraints below; for the upper bound maximize the conditional probability $\mathsf{Q}(k \mid i)$ subject to the same restrictions.

$$
\begin{aligned}
& l = \min \frac{\mathsf{Q}(ijk) + \mathsf{Q}(i\bar{j}k)}{\mathsf{Q}(ijk) + \mathsf{Q}(ij\bar{k}) + \mathsf{Q}(i\bar{j}k) + \mathsf{Q}(i\bar{j}\bar{k})} \text{ and} \\
& u = \max \frac{\mathsf{Q}(ijk) + \mathsf{Q}(i\bar{j}k)}{\mathsf{Q}(ijk) + \mathsf{Q}(ij\bar{k}) + \mathsf{Q}(i\bar{j}k) + \mathsf{Q}(i\bar{j}\bar{k})}, \text{ respectively,} \\
\text{s. t. } & \mathsf{Q}(j \mid i) = 1, \\
& \mathsf{Q}(k \mid j) = 0.8, \\
& \text{and normalization condition for probabilities.}
\end{aligned}
$$

Please note that

$$
\begin{aligned}
\mathsf{Q}(j \mid i) = 1 \text{ iff } \mathsf{Q}(ij) - 1\mathsf{Q}(i) = 0 \text{ iff } & 0\mathsf{Q}(ijk) + 0\mathsf{Q}(ij\bar{k}) - 1\mathsf{Q}(i\bar{j}k) \\
& - 1\mathsf{Q}(i\bar{j}\bar{k}) + 0\mathsf{Q}(\bar{i}jk) + 0\mathsf{Q}(\bar{i}j\bar{k}) + 0\mathsf{Q}(\bar{i}\bar{j}k) + 0\mathsf{Q}(\bar{i}\bar{j}\bar{k}) = 0
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathsf{Q}(k \mid j) = 0.8 \text{ iff } \mathsf{Q}(jk) - 0.8\mathsf{Q}(j) = 0 \text{ iff } 0.2\mathsf{Q}(ijk) - 0.8\mathsf{Q}(ij\bar{k}) \\
& \quad + 0\mathsf{Q}(i\bar{j}k) + 0\mathsf{Q}(i\bar{j}\bar{k}) + 0.2\mathsf{Q}(\bar{i}jk) - 0.8\mathsf{Q}(\bar{i}j\bar{k}) + 0\mathsf{Q}(\bar{i}\bar{j}k) - 0\mathsf{Q}(\bar{i}\bar{j}\bar{k}) = 0.
\end{aligned}
$$

For the sake of transparency, restrictions are given in table form. The whole model now reads:

$$
\begin{aligned}
l &= \min \frac{\mathsf{Q}(ijk) + \mathsf{Q}(i\bar{j}k)}{\mathsf{Q}(ijk) + \mathsf{Q}(ij\bar{k}) + \mathsf{Q}(i\bar{j}k) + \mathsf{Q}(i\bar{j}\bar{k})}, \\
u &= \max \frac{\mathsf{Q}(ijk) + \mathsf{Q}(i\bar{j}k)}{\mathsf{Q}(ijk) + \mathsf{Q}(ij\bar{k}) + \mathsf{Q}(i\bar{j}k) + \mathsf{Q}(i\bar{j}\bar{k})},
\end{aligned}
$$

s. t.	$\mathsf{Q}(ijk)$	$\mathsf{Q}(ij\bar{k})$	$\mathsf{Q}(i\bar{j}k)$	$\mathsf{Q}(i\bar{j}\bar{k})$	$\mathsf{Q}(\bar{i}jk)$	$\mathsf{Q}(\bar{i}j\bar{k})$	$\mathsf{Q}(\bar{i}\bar{j}k)$	$\mathsf{Q}(\bar{i}\bar{j}\bar{k})$		
	0	0	-1	-1	0	0	0	0	= 0	$\mathsf{Q}(j \mid i) = 1$
	0.2	-0.8	0	0	0.2	-0.8	0	0	= 0	$\mathsf{Q}(k \mid j) = 0.8$
	1	1	1	1	1	1	1	1	= 1	normalization

Appendix B

$$\mathsf{Q}^* = \arg\min R(\mathsf{Q}, \mathsf{P}^0) = \arg\min \sum_{v_i v_j v_k} \mathsf{Q}(v_i v_j v_k) \cdot \frac{\log_2 \mathsf{Q}(v_i v_j v_k)}{\log_2 \mathsf{P}^0(v_i v_j v_k)}$$

s. t. restrictions as in Appendix A.

This yields
$\mathsf{Q}^*(ijk) = 0.25, \mathsf{Q}^*(ij\bar{k}) = 0.06, \mathsf{Q}^*(i\bar{j}k) = 0, \mathsf{Q}^*(i\bar{j}\bar{k}) = 0,$
$\mathsf{Q}^*(\bar{i}jk) = 0.25, \mathsf{Q}^*(\bar{i}j\bar{k}) = 0.06, \mathsf{Q}^*(\bar{i}\bar{j}k) = 0.19, \mathsf{Q}^*(\bar{i}\bar{j}\bar{k}) = 0.19,$

and hence $\mathsf{Q}^*(k \mid i) = 0.8$.

Appendix C

V_{11}	V_1	[0.96]	V_{25}	V_1	[0.19]	V_{28}	V_1	[0.27]	V_{30}	V_1	[0.10]
V_{35}	V_1	[0.04]	V_{61}	V_1	[0.75]	V_3	V_2	[0.69]	V_{10}	V_2	[0.71]
V_{59}	V_2	[0.27]	V_{21}	V_3	[0.48]	V_{25}	V_3	[0.44]	V_{43}	V_3	[0.45]
V_{51}	V_3	[0.48]	V_{31}	V_4	[0.70]	V_{55}	V_4	[0.55]	V_9	V_5	[0.24]
V_{19}	V_5	[0.65]	V_{40}	V_5	[0.44]	V_1	V_6	[0.37]	V_{17}	V_6	[0.11]
V_{27}	V_6	[0.73]	V_{33}	V_6	[0.30]	V_{43}	V_6	[0.99]	V_{45}	V_6	[0.58]
V_{55}	V_6	[0.82]	V_{58}	V_6	[0.47]	V_{64}	V_7	[0.42]	V_6	V_9	[0.87]
V_{26}	V_9	[0.46]	V_{33}	V_9	[0.81]	V_{39}	V_9	[0.04]	V_{21}	V_{11}	[0.70]
V_{40}	V_{11}	[0.34]	V_{46}	V_{11}	[0.22]	V_{49}	V_{11}	[0.01]	V_{50}	V_{11}	[0.34]
V_7	V_{12}	[0.69]	V_{58}	V_{12}	[0.22]	V_1	V_{13}	[0.30]	V_{27}	V_{13}	[0.15]
V_1	V_{15}	[0.34]	V_{33}	V_{15}	[0.35]	V_{60}	V_{15}	[0.40]	V_{51}	V_{16}	[0.64]
V_{10}	V_{17}	[0.81]	V_{21}	V_{17}	[0.59]	V_{39}	V_{17}	[0.20]	V_{60}	V_{17}	[0.39]
V_{13}	V_{18}	[0.03]	V_{21}	V_{18}	[0.37]	V_{36}	V_{18}	[0.54]	V_{37}	V_{18}	[0.13]
V_{57}	V_{18}	[0.07]	V_9	V_{19}	[0.29]	V_{29}	V_{19}	[0.36]	V_{51}	V_{22}	[0.02]
V_{22}	V_{23}	[0.57]	V_9	V_{25}	[0.38]	V_{58}	V_{25}	[0.43]	V_{34}	V_{26}	[0.16]
V_{35}	V_{29}	[0.25]	V_{63}	V_{30}	[0.59]	V_6	V_{31}	[0.77]	V_{34}	V_{31}	[0.63]
V_{43}	V_{33}	[0.31]	V_{46}	V_{33}	[0.72]	V_{55}	V_{33}	[0.45]	V_{60}	V_{33}	[0.39]
V_5	V_{34}	[0.14]	V_8	V_{34}	[0.40]	V_4	V_{35}	[0.63]	V_{11}	V_{35}	[0.11]
V_{20}	V_{35}	[0.03]	V_{45}	V_{35}	[0.24]	V_{63}	V_{36}	[0.29]	V_1	V_{38}	[0.10]
V_{36}	V_{38}	[0.80]	V_{40}	V_{39}	[0.64]	V_{60}	V_{40}	[0.83]	V_5	V_{41}	[0.45]
V_{24}	V_{41}	[0.68]	V_{25}	V_{41}	[0.42]	V_{26}	V_{41}	[0.24]	V_3	V_{42}	[0.99]
V_9	V_{42}	[0.13]	V_{59}	V_{42}	[0.83]	V_{50}	V_{43}	[0.08]	V_{51}	V_{43}	[0.43]
V_{38}	V_{44}	[0.93]	V_{18}	V_{49}	[0.91]	V_{44}	V_{49}	[0.17]	V_{14}	V_{50}	[0.72]
V_{25}	V_{50}	[0.92]	V_{27}	V_{50}	[0.28]	V_{25}	V_{52}	[0.40]	V_{55}	V_{53}	[0.72]
V_{34}	V_{54}	[0.92]	V_{39}	V_{54}	[0.17]	V_{62}	V_{54}	[0.85]	V_{43}	V_{56}	[0.64]
V_{57}	V_{58}	[0.71]	V_4	V_{59}	[0.23]	V_7	V_{61}	[0.76]	V_{12}	V_{61}	[0.37]
V_{29}	V_{61}	[0.81]	V_{18}	V_{62}	[0.34]	V_{44}	V_{63}	[0.59]			

Table C.1: Conditionals and TPs for the Kronecker network

Appendix D

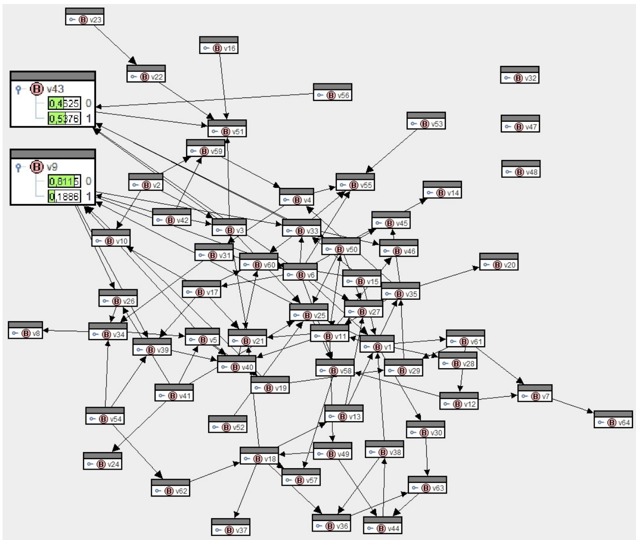


Figure D.1: Net with 64 actors

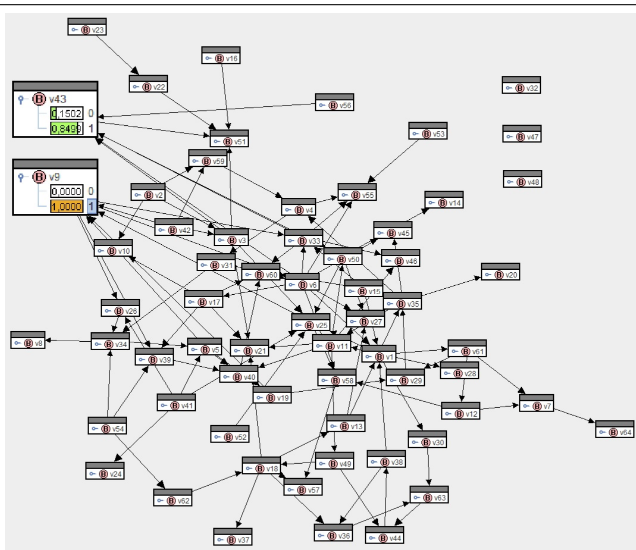


Figure D.2: Net with v9 as sender and v43 as receiver

References

[1] Ph. Bonacich. Power and centrality: A family of measures. *The American Journal of Sociology*, 92(5):1170–1182, 1987.

[2] Ph. Bonacich and P. Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3):191–201, 2001.

[3] S. P. Borgatti. Centrality and network flow. *Social Networks*, 27:55–71, 2005.

[4] D. Brenner, A. Dellnitz, F. Kulmann, and W. Rödder. Compressing strongly connected subgroups in social networks: An entropy-based approach. *The Journal of Mathematical Sociology*, 41(2):84–103, 2017.

[5] P.G. Calabrese. Deduction and inference using conditional logic and probability. In I.R. Goodman, M.M. Gupta, H.T. Nguyen, and G.S. Rogers, editors, *Conditional Logic in Expert Systems*, pages 71–100. North-Holland, Amsterdam, 1991.

[6] É. Durkheim. *Die Regeln der soziologischen Methode*. Luchterhand, Neuwied/Berlin, 1961 (first 1895).

[7] M. Finthammer, C. Beierle, B. Berger, and G. Kern-Isberner. Probabilistic Reasoning at Optimum Entropy with the MEcore System. In *Proceedings of the Twenty-Second International FLAIRS Conference*, pages 535–540, 2009.

[8] I. Gartner. Multi-attribute Utility Model Based on the Maximum Entropy Principle Applied in the Evaluation of the Financial Performance of Brazilian Banks. In P. Guarnieri, editor, *Decision Models in Engineering and Management*, pages 29–55. Springer International Publishing, Cham, 2015.

[9] T. Hoßfeld, V. Burger, H. Hinrichsen, M. Hirth, and P. Tran-Gia. On the computation of entropy production in stationary social networks. *Social Network Analysis and Mining*, 4(1):190, 2014.

[10] D. Jansen. *Einführung in die Netzwerkanalyse*. VS Verlag, Berlin, 2006.

[11] P. Jia, A. MirTabatabaei, N.E. Friedkin, and F. Bullo. Opinion Dynamics and the Evolution of Social Power in Influence Networks. *SIAM Review*, 57(3):367–397, 2015.

[12] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[13] G. Kern-Isberner. Characterizing the principle of minimum cross-entropy within a conditional-logical framework. *Artificial Intelligence*, 98(1–2):169–208, 1998.

[14] J. Kunegis and A. Lommatzsch. Learning Spectral Graph Transformations for Link Prediction. In L. Bottou and M. Littman, editors, *Proceedings of the 26th International Conference on Machine Learning*, pages 561–568, Montreal, 2009.

[15] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani. Kronecker graphs: An approach to modeling networks. *Journal of Machine Learning Research*, 11:985–1042, 2010.

[16] D. Liben-Nowell and J. Kleinberg. The Link Prediction Problem for Social Networks. *Journal of the Association for Information Science and Technology*, 58(7):1019–1031, 2007.

[17] L. Lü and T. Zhou. Link prediction in weighted networks: The role of weak ties. *Europhysics Letters*, 89(1):18001, 2010.

[18] J.L. Moreno. *Who Shall Survive: A New Approach to the Problem of Human Interrelations.* Nervous and Mental Disease Publishing Co., Washington, D.C., 1934.

[19] M.E.J. Newman. *Networks: An introduction.* Oxford University Press, Oxford, 2012.

[20] D. Ortiz-Arroyo. Discovering Sets of Key Players in Social Networks. In A. Abraham, A-E. Hassanien, and V. Snásel, editors, *Computational Social Network Analysis: Trends, Tools and Research Advances*, pages 27–47. Springer, London, 2010.

[21] E. Reucher. *Modellbildung bei Unsicherheit und Ungewißheit in konditionalen Strukturen.* Logos Verlag, Berlin, 2002.

[22] W. Rödder, D. Brenner, and F. Kulmann. Entropy based evaluation of net structures – deployed in Social Network Analysis. *Expert Systems with Applications*, 41(17):7968–7979, 2014.

[23] W. Rödder, I. Gartner, and S. Rudolph. An entropy-driven expert system shell applied to portfolio selection. *Expert Systems with Applications*, 37(12):7509–7520, 2010.

[24] W. Rödder, F. Kulmann, and A. Dellnitz. A new rationality in network analysis – status of actors in a conditional-logical framework –. In C. Beierle, G. Brewka, and M. Thimm, editors, *Computational Models of Rationality*, volume 20, pages 348–364. College Publications, 2016.

[25] W. Rödder, E. Reucher, and F. Kulmann. Features of the Expert-System-Shell SPIRIT. *Logic Journal of IGPL*, 14(3):483–500, 2006.

[26] M. Rosvall and C.T. Bergstrom. An information-theoretic framework for resolving community structure in complex networks. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 104(18), pages 7327–7331, 2007.

[27] M. Schramm and W. Ertel. Reasoning with Probabilities and Maximum Entropy: The System PIT and its Application in LEXMED. In *Operations Research Proceedings*, pages 274–280, 1999.

[28] J. Scott. *Social Network Analysis.* Sage Publications, London, 2000.

[29] S. Sharma and A. Singh. An efficient method for link prediction in weighted multiplex networks. *Computational Social Networks*, 3(1):7, 2016.

[30] J. Shetty and J. Adibi. Discovering important nodes through graph entropy – the case of Enron email database. In *Proceedings of the 3rd international workshop on Link discovery*, pages 74–81, 2005.

[31] J. Shore. Relative entropy, probabilistic inference, and ai. *Machine Intelligence and Pattern Recognition*, 4:211–215, 1986.

[32] J. Shore and R. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, 26(1):26–37, 1980.

[33] SPIRIT, 2011. Last accessed on 2017-12-08.

[34] P. Srilatha and R. Manjula. Similarity Index based Link Prediction Algorithms in Social Networks: A Survey. *Journal of Telecommunications and Information Technology*,

2:87–94, 2016.

[35] F. Topsœ. *Informationstheorie.* Teubner Studienbücher Mathematik, Stuttgart, 1974.

[36] F. Tutzauer. Entropy as a measure of centrality in networks characterized by path-transfer flow. *Social Networks*, 29:249–265, 2007.

[37] G. Ver Steeg and A. Galstyan. Information transfer in social media. In *Proceedings of the 21st international conference on World Wide Web*, pages 509–518, 2012.

[38] J. Zhao, L. Miao, J. Yang, H. Fang, Q-M. Zhang, M. Nie, P. Holme, and T. Zhou. Prediction of Links and Weights in Networks by Reliable Routes. *Scientific Reports*, 5:12261, 2015.

[39] T. Zhou, L. Lü, and Y-C. Zhang. Predicting missing links via local information. *The European Physical Journal B*, 71(4):623–630, 2009.

[40] B. Zhu and Y. Xia. Link Prediction in Weighted Networks: A Weighted Mutual Information Model. *PLoS ONE*, 11(2):e0148265, 2016.

[41] J. Zhu, J. Hong, and J.G. Hughes. Using Markov Chains for Link Prediction in Adaptive Web Sites. In D. Bustard, W. Liu, and R. Sterritt, editors, *Soft-Ware 2002: Computing in an Imperfect World*, pages 60–73, Berlin, 2002. Springer.

Received 13 February 2018

AlphaGo's Decision Making

Woosuk Park
School of Humanities and Social Sciences, KAIST
e_wspark@kaist.ac.kr

Sungyong Kim
Graduate School of Culture Technology, KAIST
simonksy@kaist.ac.kr

Keunhyoung Luke Kim
Graduate School of Culture Technology, KAIST
dilu@kaist.ac.kr

Jeounghoon Kim*
School of Humanities and Social Sciences, Graduate School of Culture Technology, KAIST
miru@kaist.ac.kr

Abstract

In this paper, we study the similarities and differences between the process of decision making in humans and AlphaGo in playing Baduk (Go, Weiqi). Previous discussions of unique or unconventional moves of AlphaGo ignored how AlphaGo tends to play in different situations: (1) when AlphaGo is leading the game, (2) when she is falling behind, and (3) when the situation of the game is close enough. Nor did they pay due attention to the problem of strategic choice of moves of AlphaGo. We argue that (1) that AlphaGo tends to play very thick and safe enclosing moves when she is leading the game, (2) that she tends to play do-or-die (all-or-nothing or gambling) moves that are backed up by very carefully calculated scheming strategy, when there is no hope to win the game, and (3) that she tends to figure out creative moves in order to take the initiative, when the game is close enough. After sharpening the concept of

*corresponding author at KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea.
We are grateful to the anonymous reviewers whose detailed comments and wonderful suggestions helped greatly improve the manuscript.

strategy itself, we also argue that there is sufficient ground to ascribe strategic reasoning to AlphaGo. Based on DeepMind AlphaGo team's monumental paper in *Nature* [24] we will check to what extent our results are compatible with AlphaGo's structure and its operating principles. What is most striking in our examination of AlphaGo's decision making is that her features can be better explained by prospect theory [14] rather than by expected utility theory. In order to test this hypothesis, we analyze many examples from AlphaGo's games. We conclude by a brief discussion of the possible implications of the present study and the remaining urgent problems for future study.

Keywords: AlphaGo, Computer Go, Decision making, Prospect theory, Rationality

> Google DeepMind provides us with strong impetus to develop general purpose artificial intelligence. For, the entire world was shocked by AlphaGo's undisputed victory over top human Baduk (Go, Weiqi) players. As is well known, chess was mastered by the computer in May, 1997, when Deep Blue defeated Gary Kasparov. Despite the extensive studies in computer Go for more than a half century, it turns out to be extremely difficult to develop a computer program that can compete with professional Baduk players. It was only about three years ago when it arrived at the level of playing on a par with advanced amateur players. Nobody anticipated even then such a rapid progress is possible for computer Baduk as to win games against professional Baduk players. Everything changed by AlphaGo, and by now no one doubts the superiority of AI Baduk players over humans. After having defeated Fan Hui, professional 2 dan, in November 2015, AlphaGo won the challenge match against Lee Sedol in March 2016 with the score of four to one. Ke Jie, who is currently the world champion, was also defeated by AlphaGo in May 2017. Meanwhile, DeepMind revealed the secret of their success in their two papers dealing with AlphaGo and AlphaGoZero published in *Nature* [16, 17; 26 also related].

Still, there is a serious problem left to be solved. We do not understand why AlphaGo makes the decision as she does in each and every move. So, no matter how superior decisions AlphaGo makes, we cannot safely rely on them, unless we can explain the reason why she makes them. We may cite cases such as Nvidia's self-driving cars that clearly show the seriousness of the problem of explainability of artificial intelligence. No matter how distinguished Nvidia's autonomous vehicle, we may not allow it to be on the market, as long as we do not understand its process of decision making. Rather, we tend to be more uncomfortable and suspicious about it for that very reason. In this paper, we study the similarities and differences between the process of decision making in humans and AlphaGo in playing Baduk.

For convenience, we may categorize AlphaGo's games into the following seven groups: (1) AlphaGo (v. 13) vs. Fan Hui, (2) 5 games between AlphaGoLee (v. 18) and Lee Sedol, (3) 60 on-line games of AlphaGoMaster with top players of Japan, Taiwan, China, and Korea, (4) 3 games between AlphaGoMaster and Ke Jie, (5) the game between AlphaGoMaster and a team consisting of top Chinese players. (6) 3 self-play games of AlphaGoLee (v. 18) with Fan Hui's commentary based on Gu Li and Zhou Ruiyang, (7) 50 self-play games of AlphaGoMaster without Fan Hui's commentary. We need to distinguish AlphaGoZero's games (AlphaGoZero vs. AlphaGoMaster, and self-play games of AlphaGoZero) from all these games of AlphaGo. These games have somewhat different values depending not only on the characteristics of the content but also on whether the official commentary from DeepMind AlphaGo team, i.e., Fan Hui's commentaries. The necessity of treating each version of AlphaGo as unique individuals has been felt quite a time ago, as AlphaGo showed such a rapid progress. Since AlphaGoZero already existed when AlphaGoMaster retired immediately after having defeated Ke Jie with the score of 3:0, and the second *Nature* paper of DeepMind AlphaGo team was already submitted to the journal, it becomes more important to distinguish carefully the different versions of AlphaGo. DeepMind AlphaGo team calls (2) "AlphaGoLee", and (3), (4), (5) and (7) "AlphaGoMaster", thereby distinguishing them from the current version, i.e., AlphaGoZero. The web-site (http://www.aphago-games.com) calls (1) v. 13, and calls (2) and (6) v. 18. Though we may discuss AlphaGoZero, if necessary, our study is basically a study of AlphaGo before the appearance of AlphaGoZero.

The five games between AlphaGoLee (v. 18) and Lee Sedol has the crucial value that they allow us to fathom AlphaGo's assessments of the game developments through Fan Hui's commentaries published on-line in English, Chinese, and Korean. Since these games were played already more than a year and a half ago, they may not fully represent the current status of AlphaGo that has evolved constantly making incredibly rapid progress. Indeed, AlphaGo won all her on-line games with the top human players and all three games against the current world champion, Ke Jie. As a consequence, we can safely count the current AlphaGo as armed with the much superior ability of assessing the games more sophisticated process in her decision making compared to her former self, i.e. AlphaGo that had the challenge match with Lee Sedol. In fact, DeepMind AlphaGo team announced the retirement of AlphaGo on the ground that it would be meaningless to compete with top human players in view of the fact that even in two stone handicap games humans cannot rival AlphaGo.

AlphaGoMaster's 60 on-line games with the top players of Japan, Taiwan, China and Korea, the three games between AlphaGo and Ke Jie, and the 53 self-play games of AlphaGo are most valuable source for estimating the assessments of the strongest Baduk player of our time. In particular, the three self-play games of AlphaGo with Fan Hui's commentaries are the most precious source of information in that thereby DeepMind AlphaGo team's analysis of the games open to the public presents many clues for understanding how AlphaGo thinks, makes decisions, and plays. However, for other games of AlphaGo with no commentaries of Fan Hui, there is certain limitation in understanding AlphaGo's reasoning only by referring to human points of view and common knowledge.

There is an interesting contrast between AlphaGo's games against humans and AlphaGo's self-play games. In most of the games between AlphaGo and humans, AlphaGo took the definite lead rather early. On the other hand, almost all the self-play games of AlphaGo were extremely close games hard to judge whether Black or White had the superior status. Even in those games ended by resignation, according to professional players, most of them were quite close games in such a way that if there was no resignation, they would have been won by Black or White only by a half or one and a half point. Now, we need to examine what implications such a contrast might have for reconstructing AlphaGo's judging and decision making process. The fact that AlphaGo took the early lead in her games with humans makes it possible to surmise that her decision making would have become much easier in the later phases of the games. Once she took the superior status, she does not have to attempt gambling moves or to struggle to make tough decisions whether to attack or defend, or whether to reduce the opponent's territory or invade it deeply. On the other hand, in AlphaGo's self-play games, she was obliged continuously throughout the games to make extremely difficult decisions for each move.

Based on careful analysis of AlphaGo's game records, our discussion will proceed as follows. In section 1, we start our inquiry by examining critically On and Jeong's study [19] of AlphaGo's unconventional moves in order to set the stage. Here, we will criticize two of the most troublesome aspects of their study. i.e., (1) that they fail to consider the relationship between the assessment of the status of the game and the choice of moves, and (2) they merely focus on some unfamiliar or strange individual moves never attempting to consider whether a sequence of moves was selected strategically at a higher dimension. By focusing on the former, in section 2, we will discuss how AlphaGo tends to play in different situations: (1) when AlphaGo is leading the game, (2) when she is falling behind, and (3) when the situation of the game is close enough. We will tentatively conclude (1) that AlphaGo tends to play very thick and safe enclosing moves when she is leading the game, (2) that she tends to play do-or-die (all-or-nothing or gambling) moves that are backed up by very carefully calculated scheming strategy, when there is no hope to win the game, and (3) that she tends to figure out creative moves in order to take the initiative, when the game is close enough. In section 3, we will discuss the problem of strategic choice of moves, which is another lacuna in On and Jeong's study. Above all, we will apply to AlphaGo's Baduk John Woods' previous research [37] on the logical foundations of strategic reasoning and Woosuk Park's concept of strategy [22], which was introduced to avoid the problems in the game theoretic concept of strategy. We will suggest from this that AlphaGo demonstrates extremely sophisticated creative strategies especially when the game is close. There is a need to reevaluate the value of AlphaGo' strange moves, including fortuitous errors or tricky plays appearing when she is almost losing the game as audacious do-or-die moves that are executed in the middle of the subtle strategic operation intending to change the direction of the games. In section 4, based on DeepMind AlphaGo team's first monumental paper in *Nature*, we will check to what extent our results are compatible with AlphaGo's structure and its operating principles. What is most striking in our examination of AlphaGo's decision making in sections 2, 3, and 4 is that her features can be better explained by prospect theory [14] rather than by expected utility theory. In order to test this hypothesis, in section 5, we analyze a few examples from AlphaGo's games. Such a case study will confirm our tentative conclusions concerning the relationship between AlphaGo's assessment of the game and her choice of moves as well as AlphaGo's strategic reasoning. It will give also some further support for our hypothesis that AlphaGo's decision making can be better explained in terms of prospect theory than expected utility theory. In section 6, brief discussion of the possible implications of the present study and the remaining urgent problems for future study will follow.

1 AlphaGo's Unconventional Moves

On and Jeong [19] categorized AlphaGo's moves that are unfamiliar to human's point of views into the following five groups: (1) brute attaching moves, (2) frequent shoulder hits, (3) moves not recommended by the textbooks, (4) 3-3 point invasion without any allies, (5) vulgar moves [19, 16f]. There is no doubt that this timely research would be the foundation stone for all subsequent studies. However, there are also its intrinsic limitations. Above all, as the authors themselves point out, there is the need for comprehensive study covering not just the opening phase but also the middle game and the ending game, and for "studying by what mechanism AlphaGo plays such unique moves in addition to noting her unique techniques" [19]. Our study can also be understood as a small attempt to meet such a need. In this vein, we would like to start our inquiry by criticizing the limitations in their research.

It seems rather natural and almost universal to grasp AlphaGo's unique traits and propensities by starting from AlphaGo's move that appear strange to human point of view. Now, let us focus on one of the fatal problems in On and Jeong's study, i.e., the lack of considering the relationship between the choice of moves and the assessment of games. Among the categories identified by On and Jeong, (1) (2), and (4) are so salient that it is hard not to be observed. As a consequence, extensive commentaries have been written on these peculiar characteristics of AlphaGo in the opening phase of the game.

On and Jeong point out that AlphaGo (in terms of patterns or haengma) plays without any hesitation "moves not found in textbooks", "moves that would bring up master's severe criticism, if played by a beginning amateur". According to them there are more than ten games in which such moves appear.

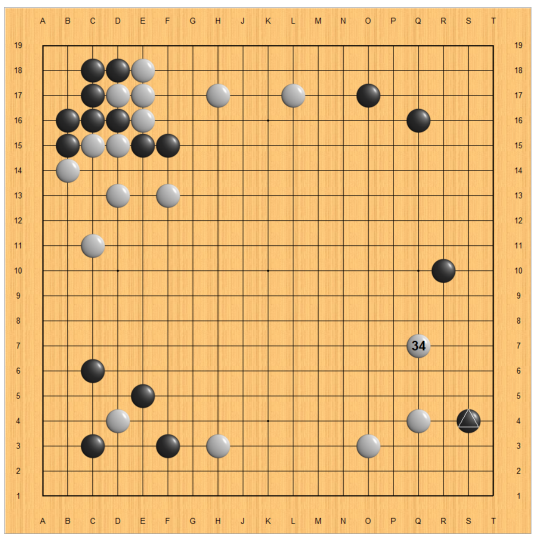


Figure 1. AlphaGo (White) vs. Liu Yuhang (Black) (∼34, number displayed only for 34)

Regarding White 34 in figure 1[1], On and Jeong comment that it is "a unique idea not found in regular patterns" and that it is "an ambiguous move in the center not directly responding to Black's low flying move". In view of the fact that, more frequently, humans would play 3-3 point to protect the corner, or push down Black

[1] This game was played at 12:06:51 on December 30, 2016 at the web site called Tygem. In DeepMind's homepage, it is uploaded as a game with Liu Yuhang, professional 1 dan, China.

more aggressively by diagonal move or knight's move, it is understandable why On and Jeong characterize the move as "an ambiguous move in the center". However, if the evaluation involved in the word "ambiguous" has the implication that it should be criticized because it is neither an attack nor a defense, it could be a somewhat hasty judgment. For, it is highly likely that such an ambiguity was intended with full consciousness. Insofar as we need to assess the whole situation on the board by evaluating the value of the move with perceptive attention to the arrangements of other stones, there is definitely an aspect of the ambiguity of the move that would make the opponent harder to respond.

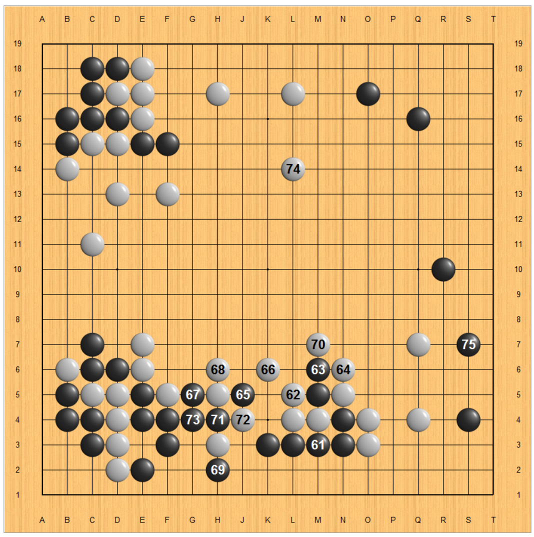


Figure 2. AlphaGo (White) vs. Liu Yuhang (Black) (∼75, numbers displayed for 61∼75)

At the time of Black 75 in figure 2, White's influence is overwhelming. If so, White's sacrificial strategy, which allowed the double pincer and the enclosure by knight's move at the left bottom corner, turns out to be very effective. From Black's point of view, it is necessary to check White's influence and to get a huge territory on the right side at the same time. But the problem seems to be that no matter what happens it is almost impossible to pay komi. Black 75 reveals Black's problem rather vividly. If it were not for White 34 (in figure 1), Black would have made a decision as to which area should be more emphasized at a much earlier stage: the right top corner or the right bottom corner. After having failed to make such a decision in timely fashion, Black has to play reluctantly such a move as Black 75, which seems to aim at connecting the allies in right top corner and the right bottom corner. It is also clear that White 34, which is placed closer to the center than the conventional moves pressing the Black's group at the right bottom corner, contributes to maximize White's influence stemming from the left top corner to the center by assisting it from the opposite side.

Another example of (3) On and Jeong cite is White 36 in figure 3[2] They write: "It would be hard to find real examples of this move in actual games of professional players, since moves like this one in the given situation are not just bad but almost meaningless".

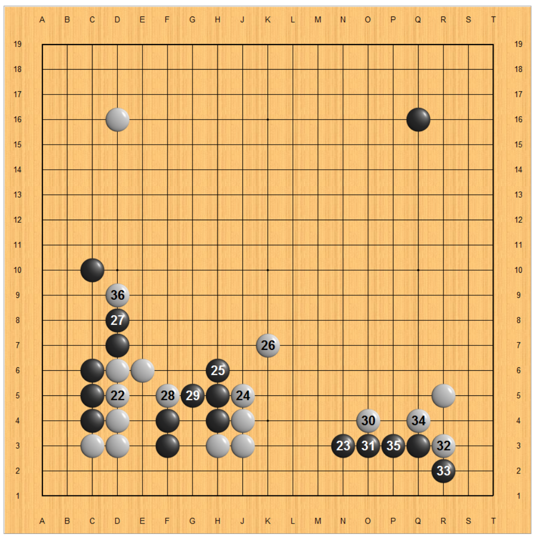


Figure 3. AlphaGo (White) vs Tang Weixing (Black) (∼36, numbers displayed for 22∼36)

Tang Weixing resigned at White 186. But the game was irrevocably leaning toward White at the time of White 66 in figure 4. As an effect of attaching move of White 36, White got a huge territory on the left side. White's solid territory on the left side and the left top corner is more than 30 points. Further, White in the bottom side and the right bottom corner has rather a thick shape, if not a remarkable influence. On the other hand, Black cannot be proud of territory or

[2]White: Master 9P, Black: Tang Weixing 9P, 2016-12-31 21:18:19

influence. In retrospect, Black 37 was an over play that determined the destiny of the game once and for all. In playing this protesting hane, Tang did not realize that White 36 was a brilliant strategic move based on deep thought.

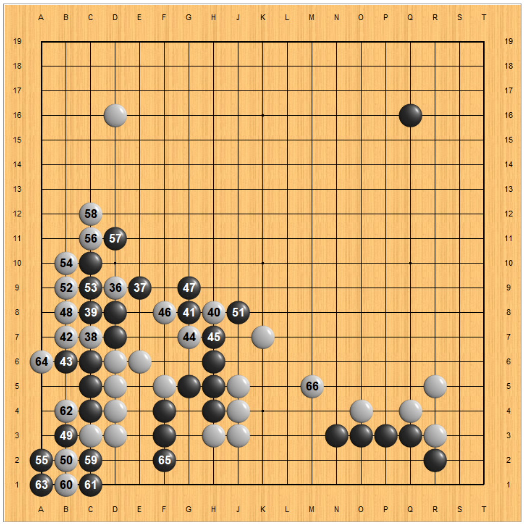


Figure 4. AlphaGo (White) vs Tang Weixing (Black) (∼68, numbers displayed for 36∼68)

Even at a casual look, On and Jeong's classification betrays a weakness that it mixes up different layers. For (1), (2), and (4) are items in terms of moves of particular patterns, while (3) and (5) are items that can be captured by the evaluation according to the characteristics of the moves at hand on a higher level. The most troubling substantial problem is, however, that they focus on moves rarely found in games played by humans only by observing the similarities in shapes without examining why a move of certain pattern was chosen in which specific phase and situation, without full appreciation of the non-monotonic logical character of Baduk. Our criticism can be further substantiated in two respects. First, the mutual dependence between the assessment of the game and the choice of moves was completely ignored. Even though the same moves can have different meaning and values depending on whether the game is leaning toward one side or quite close, such connection between the assessment of the game and the choice of move is unduly ignored. Second, they consider only the choice of single, independent moves without considering the higher order problem of selecting the sequence of moves such as the choice of strategies or tactics. For example, unless one presupposes the context of sacrificial strategy, the choice of allowing many stones to be killed might appear strange and incomprehensible.

2 The Relationship between the Assessment of Games and the Choice of Moves

The two examples, which were categorized as non-standard moves of AlphaGo and cited above, are from AlphaGo's on-line games with top players in Japan, Taiwan, China, and Korea. Since Fan Hui's commentaries on these games have not been published, it is difficult to be sure about under what assessment of the game those moves were played by AlphaGo. However, these examples provide us with a point of departure for raising the following meaningful questions.

When AlphaGo is leading the game, what tendency does she show in making decisions? Does she try to fortify her lead in territory? Or does she try to minimize the risk insofar as she keeps the lead? Since AlphaGo won almost all games against humans, there are more than enough examples. Of course, there must be a winner and loser even in her self-play games. So, the appropriate examples belonging to this category can in principle be found in them too, though they are extremely close games.

On the other hand, when AlphaGo's status is inferior, what tendency in her choice of moves is revealed? Does she prefer to play the strongest moves in order to make up? Or, does she patiently wait for the chance by trying to decrease the gap? It would be more difficult to find examples of this kind. For, there is almost no such example in AlphaGo's games against humans, and AlphaGo's self-play games are mostly quite close. Below, we will examine the situation in AlphaGo's third self-play game, where balance is destroyed and win rate is leaning toward one side.

Ultimately, we will present a working hypothesis that AlphaGo tends to play the most creative moves that look strange and unfamiliar to humans when the status of the game has not been leaning toward to one player. In arguing for such a hypothesis, our focus will be laid on what kind of creative moves AlphaGo play in a dilemma-like situation where she is forced to make a tough decision.

2.1 When AlphaGo is leading the game

As a nice example of the case where AlphaGo chose a thick and safe move that confirms her lead, we may cite the two such moves in the first game in AlphaGo vs. Lee Sedol challenge match [9, 32, 40].

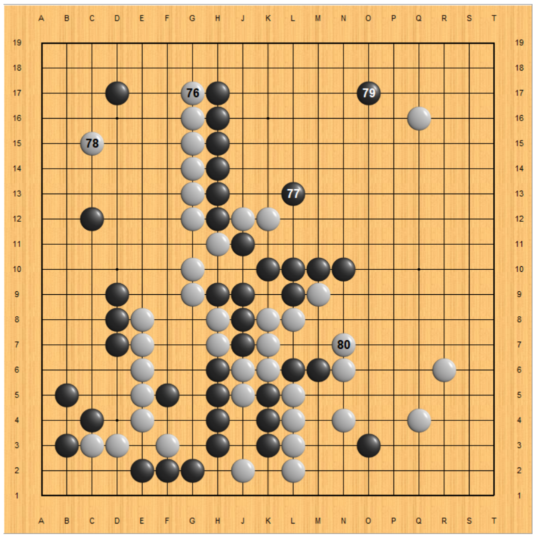


Figure 5. AlphaGo (White) vs. Lee Sedol (Black) (∼80, numbers displayed for 76∼80)

Here is Fan Hui's impression of some such moves of AlphaGo:

> "... When I saw AlphaGo's response at 80, I wrote down in my notebook: Statement of victory!... At this point, AlphaGo showed a 74% win rate, confirming my impressions. Apparently, AlphaGo thought that move 79 was not the best, preferring diagram 13 instead" [11].

The majority opinion among professional players in Korea coincides with that of ours in that White 80 is viewed as a move that can be played only when one is quite convinced of winning the game. However, there is a subtle difference among the commentators regarding whether such an assessment is correct. For example, Shin Jinseo, professional 5 dan, expressed his view as follows:

> "Shin Jinseo said that the right move, even for AlphaGo, would be the knight move at the right top corner. According to the common sense, that is right. But we should be aware of the fact that AlphaGo was guarding against any possible danger in order to raise the win rate" [16, p. 133].

Somewhat similar view as this is also found in the following report:

> "In fact, White 80 shocked the professional players who were examining the game together. There was an evaluation that "If White loses, this must be losing move.... There is no hurry to play White 80. It is indeed difficult for humans how to judge it. Nevertheless, AlphaGo never hesitated in choosing to play White 80. She must have finished the calculation for the ultimate result of the game. The calculation of the machine, which is different from humans' intuition" [39, pp. 48-49].

One interesting point is that Black 77 was identified as to blame for allowing such a move as White 80 that confirms the sure lead. Further, according to a promising examination, Black 77 was analyzed as highly likely resulting from a wrong assessment of the game. For example, Korean National Team thinks that Black 77 should have been a forcing move peeping at the tiger's mouth [16, p. 128]. A bit more severe criticism is found in Kim Yeong Sam, 9 dan professional player:

> "Kim Yeong Sam 9 dan disliked Black 77. . . Black 77 is a kind of move that announces one's own victory. Is Lee Sedol optimistic about the game as favorable to Black? Is there really some problem in Lee Sedol in the capability of assessing the games?" [39, p. 47].

White 116 in the same game (figure 6) provides us with a bit more certain example than White 80. The fact that the win rate of White was raised enormously at this scene seems to make it more plausible that AlphaGo chose White 116 as a thick move that would insure the victory. Fan Hui's comment is also consistent with our interpretation: "When AlphaGo took the 3-3 point at 116, its win rate rose to 82%. A value this high means AlphaGo already considers the game won, even if the absolute difference is small" [11].

It is somewhat surprising that the majority of professional players in Korea assessed the situation in opposite fashion, even when AlphaGo's win rate was raised up to 82%. Yang [39] witnesses well the majority view at this situation assessed the game as favorable to Black: "[Until Black 115] the general opinion was leaning toward Black". However, there was a different point of view too. Park Yeong Hoon 9 dan was the typical case. He is respected for his superior ability in the assessment of the game and calculation, possibly the best after the legendary Lee Changho. Park was examining the game together with his colleagues in a study room, and the conclusion was that "I do not know whether Black is leading the game" [39, pp. 58-59]. It is surprising to know that it was only one person, i.e., Park, who raised an objection to the majority view. Equally surprising is that his objection was expressed in such a cautious way, possibly considering the atmosphere.

Though Kam [15] resolutely judges the situation as favorable to White, he is also very cautious in phrasing:

> "Locally, there is no loss for Black. However, White not only secured the life of a large group but also exploited fully the sente by defending the right bottom corner, which was then the best in terms of territory. As result, the game is favorable to White, however slightly" [15, p. 142].

It is highly suggestive that all this assessment was made in the situation where AlphaGo's win rate in her self-assessment was 82%.

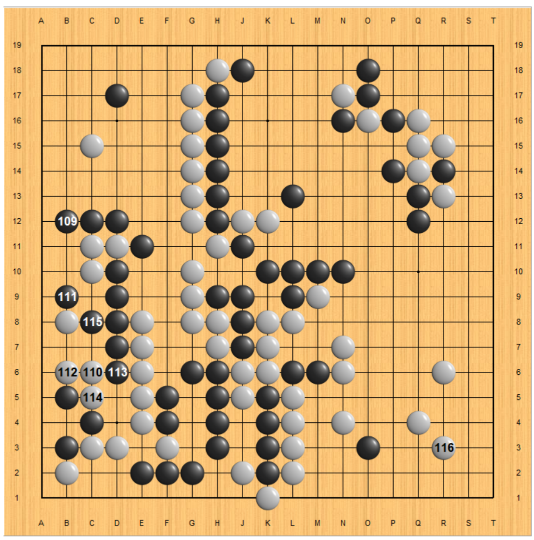


Figure 6. AlphaGo (White) vs. Lee Sedol (Black) (~116, numbers displayed for 109~116)

2.2 When AlphaGo is falling behind

Above all, we need to point out that it is almost impossible to find examples in which AlphaGo was in inferior status throughout her games with human players. As a result, we have to find pertinent examples from AlphaGo's self-play games. However, here is also a hurdle. Even when the games ended with resignation, it seems that most of them were quite close games in which the balance between Black and White was sustained continuously. Now we realize that at this moment

AlphaGo's three self-play games with Fan Hui's commentaries have utmost values. Furthermore, we can see that among the three games, the third game is the one we have been looking for. For, in this game, one player's win rate arrives up to 80%.

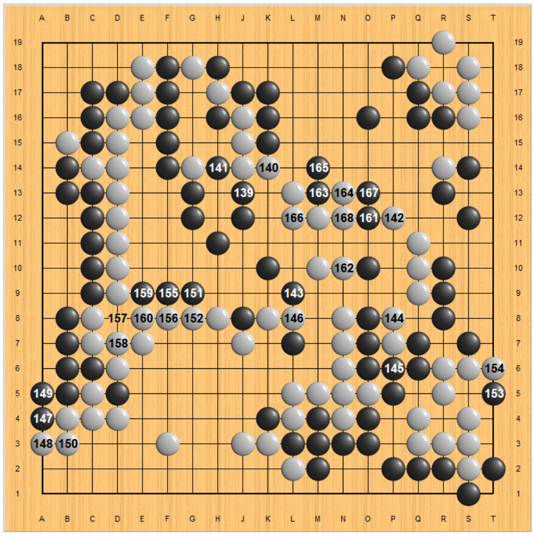


Figure 7. The 3rd AlphaGo vs. AlphaGo (~168, numbers displayed for 139~168)

According to Fan Hui, at the time of 142 (figure 7), White's win rate was raised to 65%. "Black 153 is an absolute loss, probably a reflection of Black's desperation. 161 is the last real try for a comeback, but White refutes it with the tesuji of 162... By move 168, White's win rate has arisen to 80%. Black has no chance of victory" [10]. Unless White responds with White 154 to Black 153, the whole group in the right bottom corner would be dead. Therefore, Black 153 secures sente. As long as we confine our interest on the right bottom corner, since there is no difficulty for White to be safe by responding by White 154, playing the enforcing move Black 153 seems an unnecessary loss. Isn't it, then, merely a tricky play expecting the opponent's confusion and mistakes? Probably AlphaGo would not attempt such a tricky play. So, we should understand the sacrificial move of Black 153, which swallows an apparent loss, as Fan Hui seems to believe, as the first step for "the last real try for a comeback" in order to separate by Black 161 the few White stones in the right side from both the White's group in the right bottom corner and that in the center. However, thanks to the nice move of White 152, the six stones of White in the right side can be connected to White's group in the center by appealing to the paired move aiming at the tiger's mouth. Here is the need to examine carefully whether there is any common characteristic or pattern whenever AlphaGo attempts do-or-die moves in situations when she is desperately in the inferior status.

2.3 When the Game is close

2.3.1 AlphaGo's Third Self-Play Game[3]

As noted above, since AlphaGo's self-play games tend to sustain balance between Black and White throughout the entire games, the instances that explicitly mention the win rates revealing the assessments of the players in Fan Hui's commentary are comparatively rare compared to the games between AlphaGo and humans. In Fan Hui's commentary on AlphaGo's third self-play game, at least five times such an explicit note on the win rate is found. For that reason, this game has a special value as a truly exceptional case. In particular, from this game one can see not only the process of changing situation from balanced status to situations in which one player becomes enjoying the superior status, but also how both winning and losing players select their moves once the balance is destroyed.

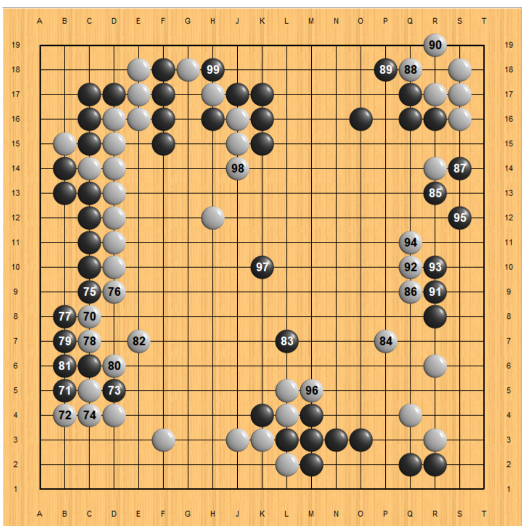


Figure 8. The 3rd AlphaGo vs. AlphaGo (∼99, numbers displayed for 70∼99)

[3]Unlike the previous two games that were played with the speed of one move within five seconds, the third game was played with speed of one or two minutes per move. (Fan Hui's comments available only on-line.)

Another notably important fact is that, in most of the game between AlphaGo and humans, the balance between Black and White was completely destroyed around the 100th move. Further, again in many of them, even around at 50th move, one player took the definite lead. In view of all this, it is remarkable that, in AlphaGo's third self-play game with Fan Hui's commentary, there is a tight balance between Black and White from the 70th to 97th move (figure 8). Let us turn to Fan Hui's commentary.

> "Against White 70, Black takes an extremely direct approach. Black first plays the forcing moves at 71 and 73, then connects up the left side, and finally plays a reducing move at 83, aiming to make White's center over concentrated. Although the aim is clear, most players could not tolerate the crudeness of the sequence of the left side. Up to here, this game conveys the feeling that White is playing with masterful lightness, while Black is being dragged around the board. Gu Li and Zhou Ruiyang felt this to such an extent that they declared the game "totally one-sided," almost as if White were playing by itself. Yet AlphaGo's own calm-minded assessment, White has a win rate of just 51.5%, a lead by only the slimmest of margins" [10].

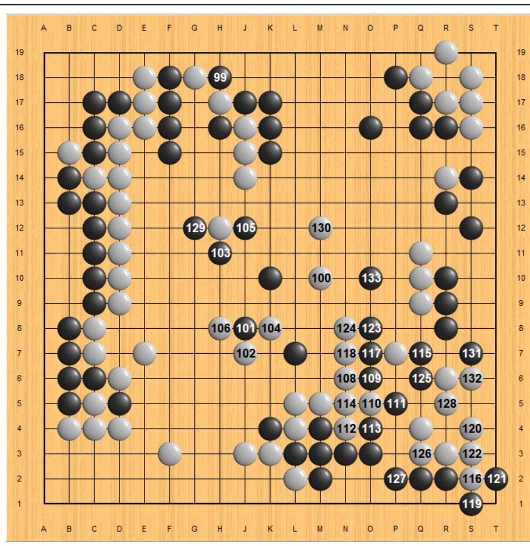


Figure 9. The 3rd AlphaGo vs. AlphaGo (∼133, numbers displayed for 99∼133)

Even at the time of the moves between 99-133 (figure 9), the balance of the game was not yet destroyed. As we can see from the following quotation, the status of the game was not leaning toward Black or White in such a way that Fan Hui counted the small increase of Black's win rate of 56% as a remarkable change:

> "Through 106, a trade has developed, and the outlook has reversed: Black's win rate now stands at 56%. In other words, Black believes that the fight in the middle has been a success. However, this judgment is predicated on Black's ability to further harass White in the center. We will investigate this assumption soon. . . White 126 is a very strange move, and incurs a definite loss of territory. Before this move, connecting at A may have influenced the status of Black's group, but once White provokes 127, the connection becomes completely gote. AlphaGo may like to play the clearest variations, but this move must be called a mistake. Through Black 133, Black's win rate stands at 53%" [10].

And, finally, crucial moment has arrived. At the scene from moves 133 to 139 (figure 10) covering only seven moves, there comes a radical change.

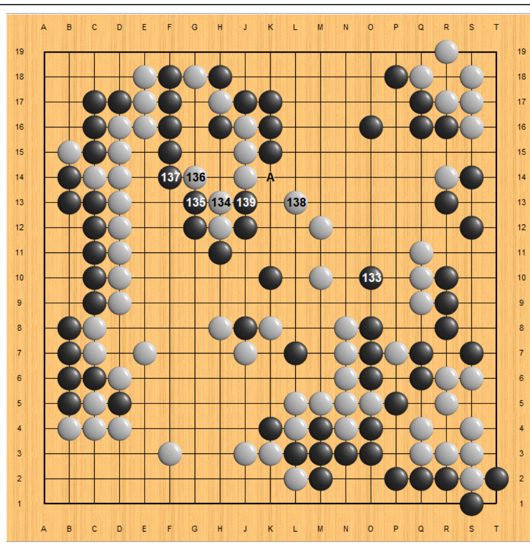


Figure 10. The 3rd AlphaGo vs. AlphaGo (∼139, numbers displayed for 133∼139)

Fan Hui's comments vividly expresses and conveys the tension at this scene:

> "The game has entered a stage of extreme suspense. When Black jumps to 133, Black believes that White will be overwhelmed trying to balance the middle and the right side. More precisely, Black thinks both sides are in danger of dying, and this is the reason behind the splitting move at 133. White also believes that the situation is difficult, but when White hanes at 136, the win rate begins to shift, as though both sides failed to foresee this move. After White plays kosumi at 138, Black has nothing better than the atari at 139! The two pressing questions are: why didn't Black push and cut at A? And why does White appear to give away points with the hane at 136?" [10]

Gu Li and other professional players, after having examined the extremely complicated numerous variations confirmed that there would no good results to Black if it had played the cutting move at A. Fan Hui continues to comment:

> "In light of the above variation, Black ultimately chose to Atari at 139 and this was the point of no return. (*See figure 7*) Thanks to White's extra stone at 140, the kosumi at 142 manages to connect up the whole center. This is a huge loss for Black! Both sides understand the situation now, and White's win rate shoot up to 65%. Black 153 is an absolute loss, probably a reflection of Black's desperation. 161 is the last real try for a comeback, but White refutes it with the tesuji of 162. See diagram 27. By move 168, White's win rate has risen to 80%. Black has no chance of victory" [10] (figure number added).

2.3.2 AlphaGo's First Self-Play Game

In Fan Hui's commentary on AlphaGo's first self-play game, we find also three scenes with explicit win rate (%). Overall, we can find here the example of the changing win rate from 50% to 60%.

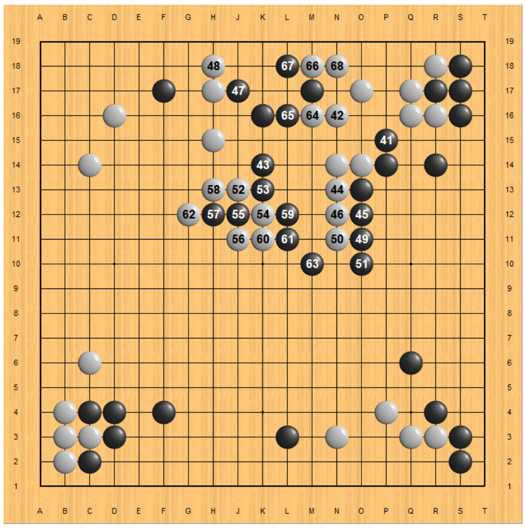


Figure 11. The 1st AlphaGo vs. AlphaGo (∼68, numbers displayed for 41∼68)

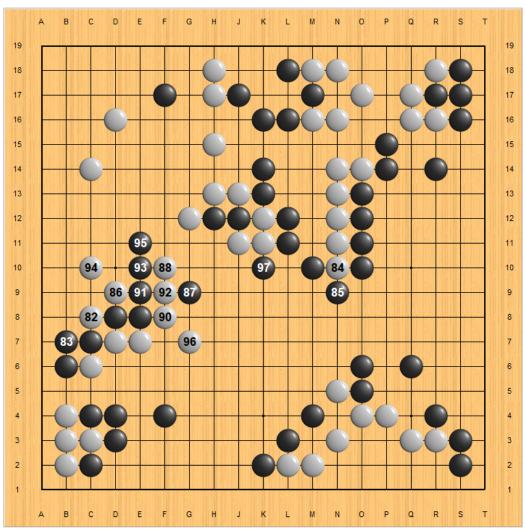


Figure 12. The 1st AlphaGo vs. AlphaGo (∼97, numbers displayed for 82∼97)

> “When Black extends at 51, AlphaGo assesses Black’s win rate at 50%: a completely balanced game... After Black sacrifices the two stones at 55 and 57 in return for the kosumi at 63, Black’s win rate rises to 52%”.
>
> “Black plays a double ladder breaker at 87, and White hangs tough with 88, Black runs out on the left at 89. Here, Black’s win rate reaches 60%! It is clear that Black holds the advantage in this fight” [10].

2.4 Tentative Conclusions

Let us summarize what can be counted as tentative conclusions. AlphaGo chooses very thick moves as if she confirms and celebrates her victory when her win rate is extremely high (approximately when it is higher than 75%) AlphaGo often plays some strange moves when the opponent’s win rate is extremely high. Nevertheless, there is room for counting them as the pursuit and the execution of somewhat creative strategy rather than tricky play. In this sense, more careful analysis of the exact timing and the method of AlphaGo in her launching a do-or-die gambling move. The situations when the balance between Black and White has not been completely destroyed (i.e., the win rate is lower than 60%) can be divided into two different categories: (1) there is still a balance even in the middle game, and (2) when it is too early in the opening phase of the game to judge which party is winning. In the former cases (for example, when there is balance when the game is at between 50th to 150th move), AlphaGo attempts a variety of strategies constantly trying to take the lead. Also, in the latter cases, AlphaGo prefers the moves that counter the opponent’s intention or the moves that could be highly likely surprising and unexpected to the opponent. From this point of view, the widely studied and confirmed propensity of AlphaGo in her decision making such as predilection of playing “shoulder hit” or “invading the corner by playing at 3-3 point” in the opening phase of the game must be the result of extensive prior study of DeepMind of modern patterns (joseki). This conjecture is corroborated by the fact that AlphaGo already attempted some new moves and patterns in connection with the some of the most famous and frequently used patterns such as the Large Avalanche joseki, the Great Slant joseki, or the Magic Sword joseki, most remarkable creative moves of AlphaGo must appear in the situations when there is still balance between Black and White.

3 AlphaGo’s Strategic Decision Making

We criticized in section 1 above On and Jeong’s study of unconventional moves of AlphaGo as filing to consider in what contexts they were played with what intention.

In particular, we pinned down its serious limitations in two respects: (1) it ignored the connection between the assessments of the game and the moves; and (2) it treated only the problem of selecting single moves without considering the problem of selecting a sequence of moves such as selecting a strategy. Since we discussed (1) in section 2, here (in section 3) we discuss (2).

As one can surmise from the fact that it is seriously considered in so many different fields of study, strategic reasoning looms large in many contexts. However, what is a strategy? What is the logic of strategy? Even in game theory, which has dominated all these problems in recent years, it is hard to secure an answer to these ultimate questions. Not to mention the classical game theory, which aimed at the highest mathematical abstraction, even in more recent trends like evolutionary game theory or epistemic game theory, it is rare to find a serious effort to uncover the essence of strategic reasoning. So, it is somewhat fortunate that logicians and game theorists tend to collaborate more actively in recent years. Such a movement is detected unmistakably in the birth of novel areas of study such as game logic or strategy logic through the interaction between epistemic logic and game theory [e.g. 20, 27]. Nevertheless, there is an unbridgeable gap between the concept of strategy in game theory and that of real games such as Baduk or chess.

Cudd [5] claims that game theory (as a part of rational choice theory) should be distinguished from individual decision theory and social choice theory. According to her, game theory is inspired by the following three ideas:

> (1) the idea that rationality is utility maximization; (2) the idea that rational beliefs and rational expectations (that is, of utility) can be formalized using probability theory; and (3) the idea that rational interaction, or interaction among rational agents, is strategic [5, p. 102].

Unlike the first two of these three ideas, the third idea "distinguishes game theory from individual decision theory". She elaborates the idea as

> That in order to act rationally in situations of interaction with other rational agents one must act strategically [5, p. 103].

If Cudd is right, then the importance of the concept of strategy in game theory cannot be too much emphasized. For, it is the differentia of game theory.

Curiously, however, the concept of strategy in game theory has never been seriously examined. In any standard textbook of game theory, of course, we can find virtually the same definition of strategy. For example, Perea [23] defines a strategy for a player i as "a complete plan of his choices throughout the game":

> Definition 8.2.1 (Strategy)
> A strategy for player i is a function s_i that assigns to each of his information sets $h \in H_i$ some available choice $s_i(h) \in C_i(h)$, unless h cannot be reached due to some choice $s_i(h')$ at an earlier information set $h' \in H_i$. In the latter case, no choice needs to be specified at h. [Perea (2012), p. 358]

What should be noted is that some such definition of strategy in game theory might have been originated from von Neumann's paper "On the Notion of Games of Strategy" [35]. As Cudd reports, the first formal treatment of strategic games was presented by von Neumann there:

> Von Neumann formalized the notion of strategy by first reducing games of chance, that is, games in which there is a risky event, to games of pure strategy by calculating the expected outcome for each player and for each possible outcome of the risky event. Then a strategy for each player consists in a set of decisions that he makes, one action for each possible decision point contingent upon the information that he has at that point [5, p.121].

Virtually the same definition of strategy is found in the monumental book co-authored by Von Neumann and Morgenstern published in 1944 [36]. In a sub-section entitled "11.1. The Concept of a Strategy and Its Formalization", we read:

> Imagine now that each player k = 1, . . . , n, instead of making each decision as the necessity for it arises, makes up his mind in advance for all possible contingencies; i.e. that the player k begins to play with a complete plan: a plan which specifies what choices he will make in every possible situation, for every possible actual information which he may possess at that moment in conformity with the pattern of information which the rules of the game provide for him for that case. We call such a plan a strategy [36, p. 79].

The problem with such a concept of strategy in game theory is that it is completely free from the reasoning about the opponent. For it is interested in the gains secured to the player even if he or she has no clue about the opponent's choices. This problem is still found even in evolutionary game theory or epistemic game theory that evolved from classical game theory. Even though the interaction between epistemic logic and game theory became rather active around 1990, as the need to consider some problems of epistemic logic such as opponent's desire and belief, the hierarchy of beliefs, common belief and knowledge were appreciated, the weakness of the concept of strategy has not been corrected in all these efforts based on the mathematical approaches of game theory. As a consequence, such a problem enforced even the research group around Johan van Benthem [34], who has been the leader in strategy logic or game logic to turn to cognitive science to revise the basic approach [8].

Park [22], after having criticized the problems in the concept of strategy implicitly assumed in game theory, and thereby in game logic or strategy logic, and attempts to reestablish a concept of strategy" from the descriptive point of view of the theories of war, whether it be that of Sun Tzu or Carl von Clausewitz. Based on John Woods' study [38] of the so-called CLM approach (Carl von Clausewitz, Edward N. Luttwak, and Henry Minzberg) from the history of the theories of war, a tentative desideratum for any good strategy (strategy∗) was suggested as follows: as the CLM approach:

To sum up, we have found three desiderata for any strategy∗, i.e., any good strategy, should satisfy.

1. It is not necessarily the case that a strategy is found in any game.
2. There must be an intriguing relationship between it and the tactics supporting it.
3. It should be inconsistency-robust.

Park [21] wants to show ultimately that, if we apply this concept of strategy∗ to Baduk, some examples of the most brilliant strategies in the history of Baduk can be explained successfully. (1) is from the criticism of the fictional character of the concept of strategy in game theory. (2) is aiming at eliminating the troubling cause of understanding a strategy as merely a *tactics writ large* by mixing up strategies and tactics without clearly distinguishing between them. Also, (3) is not only a necessary requirement due to the sheer size of the information to be processed in wars or Baduk games, but also making explicit the spirit of "Abandon small to save big" emphasized by both in Warcraft and Baduk.

As a representative example of such a strategy∗, Park cites the game between Honginbo Jowa and Inseki, played on December 2, 14, and 24, 1815.

What is remarkable in these two diagrams is that the two sacrificial tactics used by Inseki early in the top side and later in the right bottom corner were organically synthesized in such a way that a super large scale strategy governing the whole board and building a backbone of the entire story of the game was executed.

Apparently, at the time of White 106 in figure 13, despite the sacrificial play performed at the top side, Black failed to block White's group's thrusting from the top side to the center. Perhaps, for that reason, Jowa allowed Inseki to play another sacrificial tactic at the right bottom corner. However, at the time of Black 143 in figure 14, it turns out that Jowa underestimated the power of Inseki's sacrificial play. Although Inseki lost this game due to the counterattack starting with White 144 and the subtle move of White 170, what Inseki showed us by this strategy suggests to us a significant insight as to "what kind of a plan can be called a great strategy". Park [21] highlights the insight that we should ask "When is a strategy?" rather than "What is a strategy?" as the most important lesson from applying the concepts of strategy and tactics to this example. In this game, according to Park's view, Inseki's strategy started at Black 105 and Black 143.

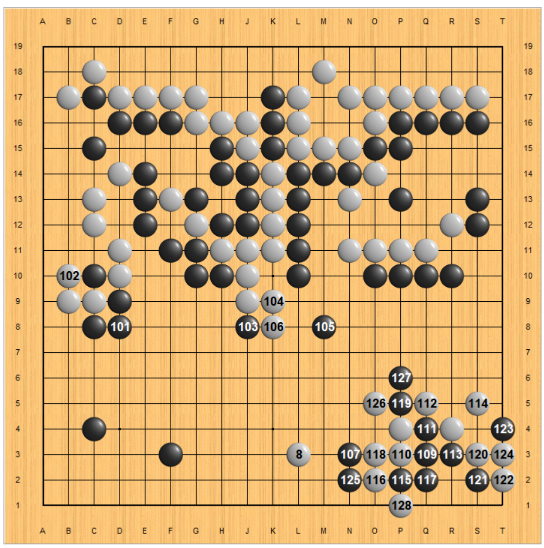


Figure 13. Honginbo Jowa vs. Inseki (∼128, numbers displayed for 101∼128)

Let us apply such a concept of strategy to AlphaGo's Baduk. The 27th of the self-play games of AlphaGoMaster is a rare example of a large scale sacrificial play that reminds us of the game between Jowa and Inseki discussed above. Figure 15 shows us that the entire group of White at the left top corner is dead at the time of Black 57. Even at the time of Black 47, it was already virtually dead, other things being equal. By the time of Black 57, the combat which started at the left top corner is over, since the Black stones surrounding the group of Whites escaped safely toward the bottom side.

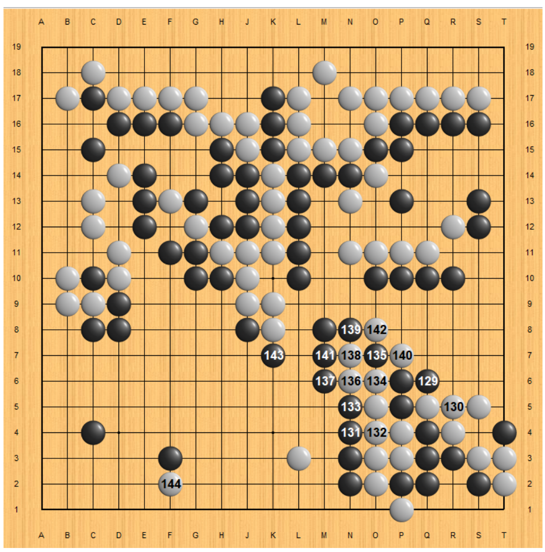


Figure 14. Honginbo Jowa (White) vs. Inseki (Black) (∼144, numbers displayed for 129∼144)

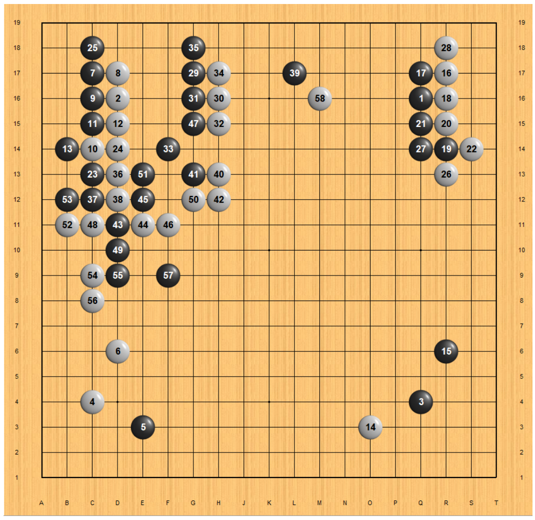


Figure 15. The 27$^{\text{th}}$ AlphaGo vs. AlphaGo (1∼57)

However, if we examine the situation starting with White 58 and developed to figure 16, it is not the case that White gave up the huge group on the left top corner without any compensation. Rather, White seems to develop a grand scale sacrificial strategy.

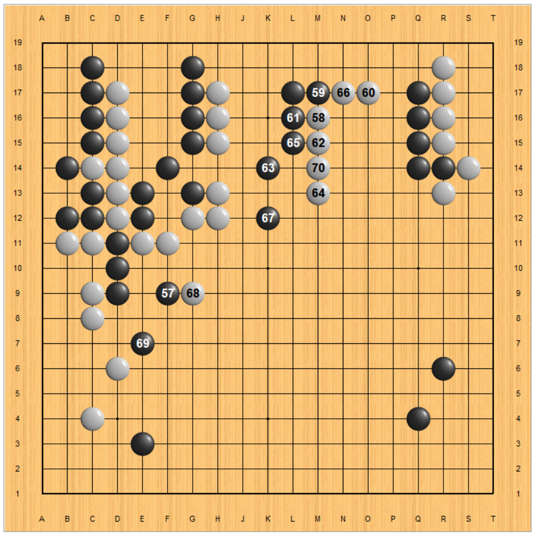


Figure 16. The 27$^{\text{th}}$ AlphaGo vs. AlphaGo (∼70, numbers displayed for 57∼70)

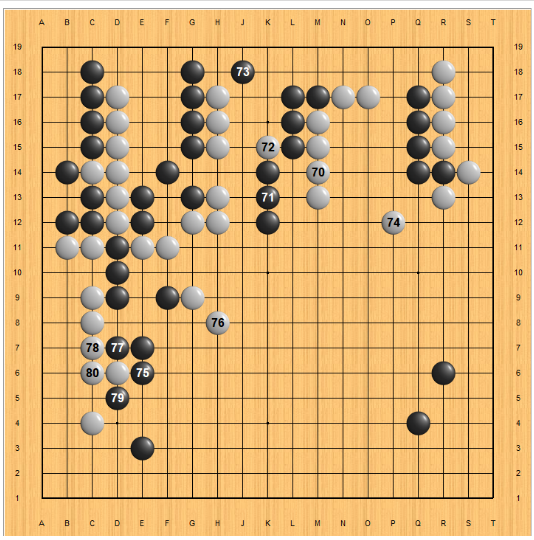


Figure 17. The 27$^{\text{th}}$ AlphaGo vs. AlphaGo (∼80, numbers displayed for 70∼80)

Indeed, such a hypothesis is confirmed to be almost actualized by the time of the figure 17. Since this game consists of extremely complicated fights in which several groups of Black and White are involved, we need exact and detailed examination and analysis to arrive at a convincing conclusion. However, we seem to have enough information for our purpose, i.e., understanding AlphaGo's strategy. It is hard to determine whether White had the grand scale sacrificial strategy of attacking the group of Blacks at the right top corner by sacrificing the group of White at the left top corner even at the outset, or White was forced to play moves securing compensation as much as possible only after it got a huge loss at the left top corner, which can be interpreted retrospectively as if White had a sacrificial strategy. However, except for the problem of determining when such a large scale sacrificial strategy began and ended, all the other elements that would constitute the necessary and sufficient conditions for confirming the existence of a strategy∗ are clearly presented. It is not necessary that all games exemplify a strategy∗. But this game already satisfied the condition (1), for it contains enough strategic components to be called a strategy∗. Also, the condition (3) for the existence of a strategy∗ is satisfied, for there was a sacrificial operation, which by definition presupposes inconsistency-robustness. Finally, the condition (2) is sufficiently satisfied. White was threatening Black to escape the group of Blacks lying between the groups of White at the left top corner and at the right top corner by connecting it with its allies at the left top corner. In this process, a huge group of Blacks is surrounded, and White was executing a grand scale sacrificial strategy by combining the sacrificial tactic used at the left top corner and the tactical technique of attacking the group of Blacks at the top side.

It seems difficult to find examples of such a grand scale sacrificial strategy that were planned initially and actually executed in the Baduk games played by humans. For, the immediate loss could be too much to swallow. At the same time, however, we cannot ignore the fact that, when for some reasons such as simple mistakes or misjudgment one's group has been captured, sacrificial strategy is a strategy that naturally suggests itself as an option. In that sense, we cannot exclude the possibility that, the cases of enforced sacrificial strategy might present more appropriate examples for comparing human and AlphaGo's Baduk than the cases of sacrificial strategy set up when the game is close. From this perspective, the situation of the end game of AlphaGoMaster's third self-play game (figure 7) seems to present an appropriate example.

Black 153 shown in figure 7, when assessed independently of the entire situation, is certainly a loss that gives up one stone (Black 153) to White's group in the right bottom corner (which is alive, ceteris paribus) without any compensation. If such a move appears in games played by humans, it could be counted as a move to earn time. Or, it could be counted as a tricky play that expects the opponent to ignore (or bypass) the possibility that, for some reasons, if White does not respond to Black 153, then Black can capture the entire group of White in the right bottom corner. In fact, some such moves to earn time or tricky plays are frequently found in games between humans, especially in amateurs, when the fate of the game is apparently leaning toward the opponent. Black 153 has been mentioned as an unusual move, in that sense, for such a move that allows an obvious loss is rarely found in games played by top level professional players. They instinctively avoid such moves. Also, it is extremely rarely to find simple tricky plays in those highly advanced players. As a consequence, there seems to be a necessity to reinterpret Black 153 as being backed up by some legitimate and reasonable thought rather than as simple mistakes, tricky moves, or moves to earn time.

4 Some Unique Features of AlphaGo's Decision Making in Terms of Her Mechanism

Before AlphaGo, the progress in computer Go was extremely slow, despite the several decades' effort. Even after chess was conquered by Deep Blue, it has been the prevailing opinion that, due to the difference between Baduk and chess in the sheer size of the possible moves, it would take at least several more decades for computer Go programs to play at the level of professional Baduk players. Monte Carlo Search was a landmark in that, by 2015, not only it upgraded computer Go one step further rivaling the most advanced amateur Baduk players, but also prepared the way toward the appearance of AlphaGo. Assuming that there is a gap (at least 2 or 3 handicap stones) between advanced amateurs and the professional Baduk players, however, a long period of time seemed inevitable to see the true tournament between computer Baduk and top human players. Therefore, AlphaGo's one-sided (4:1) victory over Lee Sedol, who was the best player in the world for years, was indeed a remarkable achievement of computer Baduk, a revolutionary event deserves to be called "AlphaGo Shock". Then, wherein lies the crucial difference between AlphaGo and the previous computer Go programs before AlphaGo?

The monumental paper of DeepMind AlphaGo team, published in January 2016 in *Nature*, provides us with an ample explanation of how AlphaGo is constituted and how it works [24]. AlphaGo consists of basically the three components: (1) policy networks, (2) value network, and (3) Monte Carlo Tree Search (MCTS). The algorithm used in computer Go prior to AlphaGo was based on MCTS and the policy function learned from the behaviors of human players [e.g. 17]. On the other hand, what is new in AlphaGo can be found in that AlphaGo exploits extensively reinforcement learning [e.g. 30], and uses DNN (Deep Neural Network) in policy function and value function. In effect, the reason why AlphaGo is distinguished can be found in the efficient combination of its components, i.e., supervised learning, reinforcement learning, value function, and MCTS. Let us see how DeepMind AlphaGo team views the matter in the first *Nature* paper.

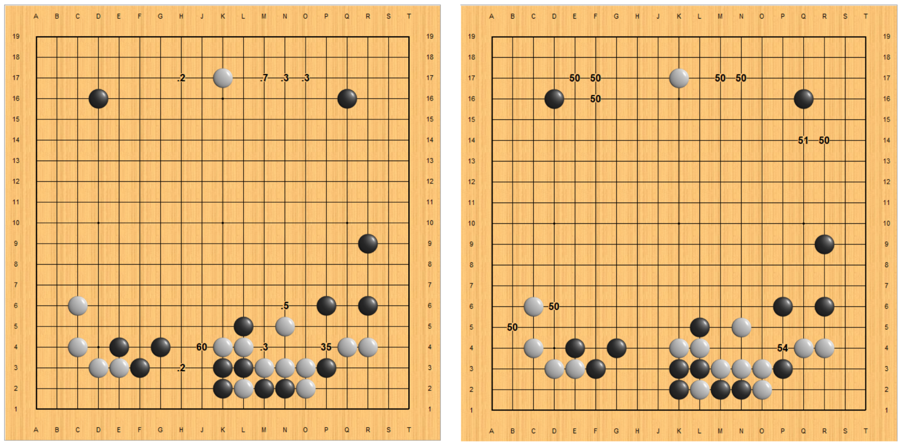
According to DeepMind, AlphaGo evaluated thousand times fewer positions in her game against Fan Hui than Deep Blue did in its match against Kasparov:

> "compensating by selecting those positions *more intelligently*, using the policy network, and evaluating them *more precisely*, using the value network – *an approach that is perhaps closer to how humans play*" [24, p. 489] (Emphases added)

The most interesting parts in the quote are italicized, for, as will be made clearer below, these suggest a very useful perspective to understand the roles and functions of policy networks and the value network. Here, let it suffice to note that such a perspective that counts "what is more intelligent" as "what is closer to how humans play" will be a very important point of reference in discussing the similarities and differences between human decision making and AlphaGo's decision making.

Before the monumental challenge match with Lee Sedol, AlphaGo played 5 official games and a few more informal games with Fan Hui, who is a professional 2 dan player. Though Fan Hui lost all five official games, he managed to win a couple of games in unofficial games against AlphaGo. The game records used by DeepMind's groundbreaking article published in Nature were from one of those unofficial games between Fan Hui and AlphaGo.

What is most shocking is that there is a dazzling contrast between the policy network and the value network. In the policy network shown in figure 18(a), the focus is on the move that scored 60 and the move scored 35. For, all the other moves scored less than 1. AlphaGo's judgment that humans prefer the former move to the latter move seems correct. It could be the case that most Baduk players, including the advanced ones, would play the former move without serious consideration. The move not only guarantees ample territory but also promises to secure sente. For, Black can capture two White stones by ladder, unless White responds to Black's move that scored 60. Even if one considers the invading move in the right bottom corner that scored 35, it would rarely be executed, since it is not so attractive. For, as shown in figure 19, though it is a quite nice move destroying White's territory in the right bottom corner, it is not a fatal move threatening White's group. There is even a worry due the uncertainty involved in case White counterattacks by thrusting a wedging move, which was in fact the choice Fan Hui made in the actual game. Now, we can see that in the value network shown in figure 18(b) the invading move in the right bottom corner got the highest score 54, and there are many other moves that scored 50, while the hane move in the bottom side got extremely low evaluation even failing to get serious consideration.



(a) (b)

Figure 18. (a) Policy network (b) Value network (modified from figure 5(a, d) in [24])

I would like to draw the readers' attention to the fact that this scene is one of the cases where a very serious problem of choice was given to AlphaGo at the early phase (around 30th move) of the game. The balance between Black and White has not yet been destroyed, and there are several nice moves one might want to play, whether they be beneficial for better opening strategy or vital for forthcoming battles. Humans would prefer moves that guarantee the moves that promise sure territorial advantage as well as securing sente so that they could move on to the remaining good positions. Of course, AlphaGo also understand all this. Nevertheless, AlphaGo selected a move the outcome of which is relatively uncertain compared to the move humans prefer. How are we to explain this interesting decision? Let us remember that, in view of the discussion above, this scene is an example of a case where AlphaGo is facing a difficult problem of choice in a situation where there is still a balance between Black and White. In other words, this scene does not represent a situation where one should consider the moves that would sustain one's superior status or a situation where one should consider challenging moves to change one's inferior status drastically. Above, we also pointed out that there is the best chance for AlphaGo to play creative moves when there is still a balance between Black and White. Let us emphasize the point that here “creative moves" does not refer to the strange moves in the sense that playing such a move is not easy for them even when they have the superior status, or the strange moves in the sense that human would never choose such a move when they have absolutely inferior status, for that would apparently result unnecessary losses. AlphaGo’s move in this situation is an unfamiliar, strange, creative move in the sense that it is not only a kind of a move that would allow AlphaGo to take the lead in a balanced situation but also that a kind of a move that it is not easy for humans to play.

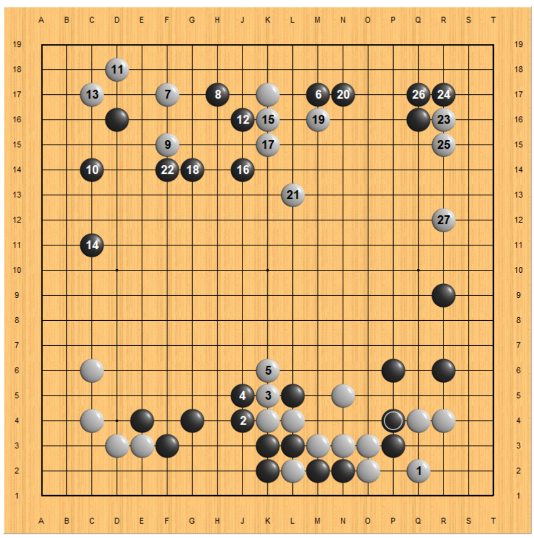


Figure 19. A variation reference to figure 18 (modified from figure 5(f) in [24])

Figure 19 shows what AlphaGo’s anticipation of the variation resulting from Black’s move destroying White’s moyo in the right bottom corner. When compared with the variation expected from the hane move on the bottom side, their difference becomes salient. And, we can agree that AlphaGo’s choice was correct. The situation that would follow from the hane move enforcing 3 and 5 in figure 19, which would usually be preferred by humans, is almost similar to that of figure 19, except for the shape of right bottom corner. But the difference is overwhelmingly huge, and, as there is such a difference, the latter, i.e., choosing to destroy White’s moyo in the right bottom corner, as AlphaGo did, is more favorable to Black. In the former case, everything is exactly the same, except for the fact that there is no more room for playing the move, marked by O, that would destroy White's moyo in the right bottom corner. For, if O is played, since White has built a great wall by 3 and 5, all black stones conscripted to destroy White's moyo in the right bottom corner would be captured by White's splitting move. On the other hand, if Black takes the subsequent sequence of moves, after having played O first, as in figure 19, enforcing White's best reply on White 1, Black already successfully reduces White's territory in the right bottom corner. Further, even if two Black stones are captured by White's splitting move, it would be favorable to Black in as much as it has enforced White to invest one more move by White 1.

5 The Connection between AlphaGo's Assessment of the Game and her Unconventional Moves in Terms of Prospect Theory

For a long time, expected utility theory has been accepted widely as the model for human decision making [12, p. 1; 3, p. 5]. However, expected utility theory was severely attacked by prospect theory suggested by Kahneman and Tversky [14, 33]. This challenge from experimental psychology shows that the classical decision theory is not adequate either as a descriptive theory or as a normative theory of human decision making. For, Kahneman and Tversky show once and for all that some of the fundamental axioms of classical decision theory, for example, the assumption that human preference is transitive, are often violated by humans. Curiously, as Thagard and Millgram acutely point out, in spite of such a refutation by prospect theory the classical decision theory based on expected utility theory is still a foundation stone in economics [31].

Now, after briefly summarizing the basic tenets of prospect theory, we would like to confirm the results regarding the mechanism of AlphaGo's decision making in the previous sections by some further concrete examples in terms of prospect theory. For, all these results betray somewhat unexpectedly similarity with some basic tenets of prospect theory.

5.1 Prospect Theory as a mechanism of human decision making

As we saw above, there are two core components that make the backbone of Google DeepMind's AlphaGo: policy networks and a value network. The policy networks provide guidance regarding which action to choose next, while the value network provides the estimate of the value of the current state of the game. Here, the policy networks are essentially neural networks trained using the professional Go players' data. This is based on the premise that human Go players make the best possible rational choices and that AlphaGo can do the same by learning from the human rationality. This imposes an utmost importance on scrutinizing the problem of human decision making.

The human decision making has long been explained by the expected utility theory which assumes that all reasonable people would make rational choices based on the product of expected outcome and probability. Although the domain of rational choices is not limited to any particular class of consequences, it has long been accepted as a normative model of economic behavior. Let us consider the following hypothetical choice problem.

A 80% chance to win $4000, 20% chance to win nothing.

B 100% chance to win $3000.

(phrases modified from [14])

According to the expected value theory, the utilities of outcomes are weighted by their probabilities. Thus, since the expected utilities of A and B are $0.8 \times 4000 = 3200$ and $1.0 \times 3000 = 3000$ respectively, the theory suggests that any reasonable person would choose A to pursue higher expected gain. However, when respondents were asked to imagine the above choice problem and indicate the decision, the majority preferred B over A. This experiment shows the typical 'certainty effect' that when faced with positive prospect choices, people overweight outcomes that are considered certain, relative to outcomes which are merely probable.

However, when the positive prospects are replaced with negative values, that is, the signs of the outcomes are reversed so that gains are now losses, the respondents' preferences become the complete opposite. Named as the 'reflection effect', the subjective expected value of prospects around 0 reverses as if the preferences are forming the mirror image. For instance, let's look at the following choice problem which is basically A and B with negative signs.

C 80% chance to lose $4000, 20% chance to lose nothing.

D 100% chance to lose $3000.

(phrases modified from [14])

Again, according to the expected value theory, the expected utilities of outcomes of C and D are –3200 and –3000 each, which makes it only rational for any reasonable person to choose D in order to avoid higher loss. However, that was not the case with actual respondents. In fact, over 90% of respondents chose C instead of going with D showing a risk-seeking preference in order to avoid a smaller but certain loss.

Based on these empirical effects and other findings which appear to invalidate the expected utility theory, Kahneman and Tversky suggested the prospect theory in 1979. They argued that people overweight the outcomes of certain events when it comes to decision making and proposed the following propositions.

1. In the positive domain (gains), people show risk-aversion preference.

2. in the negative domain (losses), people show risk-seeking preference.

That is, people have a tendency to prefer a smaller but sure gain over a larger gain that is merely probable in positive outcomes, while merely probable larger loss is preferred over a smaller certain loss in negative domain.

This tendency is well depicted in the value function which maps each outcome and its values. A hypothetical value function for changes of wealth depicted in the figure indicates that the function is concave above the reference point and convex below it. It implies that the value of gains increase in positive direction logarithmically, shrinking the subjective gain (satisfaction) on the high outcomes, while similarly the subjective sense of loss (dissatisfaction) on higher losses becomes relatively dull.

As we have seen from the prospect theory, the human decision making is better explained by subjective value estimation rather than crude mathematical calculations. Now, it is worth noting that AlphaGo's policy network was trained with human Go data which is thousands of collected Go players' choices that are bound to be under the framework of prospect theory. Therefore, although AlphaGo's decision making mechanism won't be the same with that of professional Go players (or any human being for that matter) since there are additional neural network components that work in tandem such as the value network with reinforcement learning, it is worth analyzing AlphaGo's choice of moves in terms of prospect theory.

5.2 Understanding AlphaGo's Decision Making in Terms of Prospect Theory

We saw above that, according to prospect theory, humans show behavior of (1) risk aversion preference in the positive domain (gains), while (2) risk-seeking in the negative domain (losses). But we noted above in section 2 exactly the same behavior in AlphaGo's decision making. We exemplified the behavior of (1) by the moves White 80 and White 116 in the first match between AlphaGoLee and Lee Sedol (figure 5 and 6). Let us not forget the fact that the win rate of White at the time of White 80 was 74%, and the win rate of White at the time of White 116 was 82%. It is evident that these moves clearly show the positive certainty effect, i.e., the preference to surely attainable positive result in prospect theory, As the example that clearly shows the behavior of (2) we suggested the moves starting with Black 153 in the third self-play game of AlphaGoMaster (figure 7). These moves nicely show the negative certainty effect, i.e., the aversion for the sure losses in prospect theory.

By prospect theory, White 80 and White 116 in the first game between AlphaGo and Lee Sedol (figure 5 and 6) can be interpreted beautifully as risk aversion in a favorable situation. Also, by prospect theory, moves from Black 153 in the third self-play game of AlphaGoMaster (figure 19) can be interpreted well as risk seeking. Now, let us focus on the fact that these examples were identified by professional players and the commentators as unconventional moves of AlphaGo differentiated from peculiarly human moves. Why did they were believed to be unconventional moves of AlphaGo differentiated from peculiarly human moves?

Whether it be as a descriptive theory or as a normative theory, the status of expected utility theory was seriously challenged by prospect theory. In particular, expected utility theory as a descriptive theory for human process of decision making must be damaged severely. On the other hand, it may not have been too damaging to expected utility theory as a normative theory for human decision making process, even if humans actually make decisions in accordance with the model of prospect theory rather than that of expected utility theory, still it would be possible to claim that rational decision making must be made in terms of expected utility theory.

Given all this, what does it mean that AlphaGo shows behaviors in decision making more akin to prospect theory than expected utility theory? The reason why professional players and the commentators were so surprised by the so-called unconventional moves of AlphaGo might be found in the fact that they were unable to understand them from the point of view of the model of decision making based on expected utility theory, which guarantees the rationality implicitly assumed by them. In other words, since AlphaGo won Baduk games over humans, AlphaGo's decisions are rational in terms of expected utility theory, and this should be explainable by expected utility theory. However, they had difficulties in providing us with some such explanations as to the so-called unconventional moves of AlphaGo. If AlphaGo exploits the model of decision making based on prospect theory, and if the winner's model of decision making is surely securing rationality, the only possible conclusion seems to be that prospect theory, both as a descriptive and as a normative theory, is the decision-making model of AlphaGo. In the following section, we shall discuss this important central problem in more detailed and systematic fashion. Here, prior to such a discussion, we shall examine a few cases so that we may fathom to what extent AlphaGo's decision making behaviors can be explainable and understandable in terms of prospect theory.

5.3 Case 1: White 34 in AlphaGo's Third Self-Play Game

Fan Hui's comments indicate clearly that White 34 in figure 20 is an unfamiliar move hard to see in human players' games:

> "Here, Black and White have reached an agreement of sorts. Black hanes on the outside, and White hanes and extends at 32 for a very comfortable position. However, just as we were thinking White would be satisfied with the extension at A, White pleasantly surprised us by invading directly at 34! Could it be that AlphaGo does not understand the importance of a hane at the head of two stones? Truly, AlphaGo never ceases to amaze. One would never find a tenuki like this in a game between professionals. What will happen if Black plays the hane now? See figure 21" [10].

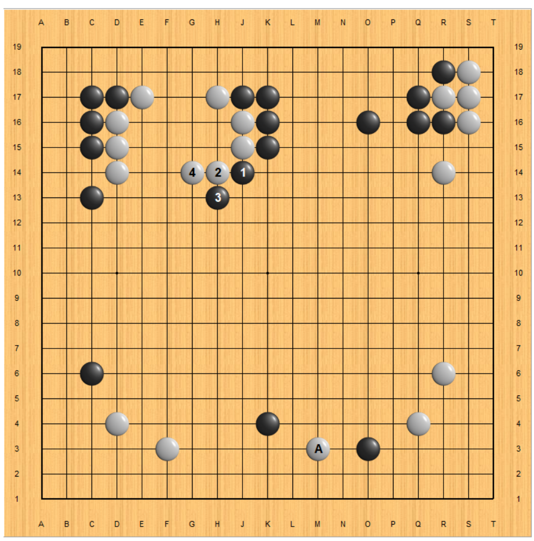


Figure 20. The 3rd AlphaGo vs. AlphaGo (~34, numbers displayed for 28~34)

Regarding the variation in figure 21, Fan Hui continues to explain:

> "When Black hanes at the head of two stones, White must hane in return, and Black's double hane looks very comfortable. But if we look closer,

are Black's profits actually that significant? White is alive with territory, and the aji of the cutting points diminishes Black's advantages in the middle. Furthermore, White has already invaded at the bottom. Has Black truly profited? Despite this, Zhou Ruiyang emphasized strongly, "If it were me, I would have extended." Wrong or right, AlphaGo has once again opened our minds to a new perspective. Perhaps we really can play this way. One thing that becomes obvious from Fan Hui's comments is that, even when they find AlphaGo's unconventional moves reasonable, top human players are still reluctant to accept them by heart" [10].

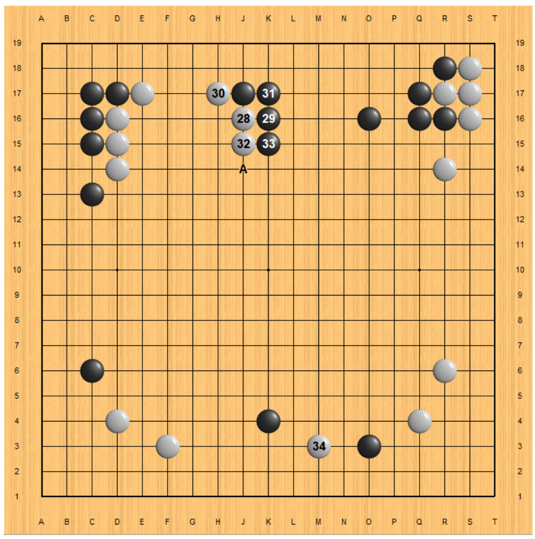


Figure 21. A variation reference to figure 20

What becomes clear from Fan Hui's commentary here is that top human players of Baduk show strong aversion to accept AlphaGo's unconventional moves even after having understood their cogency. Why? Which, among the rival theories, i.e., expected utility theory and prospect theory, provides us with a better explanation of the aversion of professional players? Since the situation is still in the early phase of the game, it seems rather difficult to claim that prospect theory is better than expected utility theory for explanation. It seems prudent and appropriate to focus on the possibility that humans' thinking, especially in the professional players, is governed unconsciously by expected utility theory. Above all, Fan Hui, by his question

"Are Black's profits actually that significant?" shows inadvertently that all Baduk player might in fact assume that judging profits and losses is the essential element in decision making. Of course, we should leave open the problem of determining whether there is any dependence between the contemporary scientific theories and the theories of Baduk as a future agenda. However, please note the fact that most of the introductory Baduk textbooks explain the reason why we proceed from the corner and move on to the sides, and then to the center part of the board in the opening phase on the ground that we can efficiently occupy more territory by using a fewer number of moves in the order of corner, side, and the center, This fact indicates the possibility that expected utility's prolonged dominance as the model of human decision making might explain why the ideas tantamount to the axioms of expected utility theory have played the role of norms or regulative ideals in the decision making of Baduk players [4, p. 7-13; 13, p. 173]. Also, Zhou Ruiyang's repeated expression of aversion and reluctance by uttering "If it were me, I would have extended." could be understood as confessing the difficulties involved in the rejection of a well-entrenched belief on a familiar shape such as hane at the head of two stones from the repeated inductive evidence throughout his career as a professional Baduk player by a single counter-example.

5.4 Case 2: Black 13 in AlphaGo's First Self-Play Game

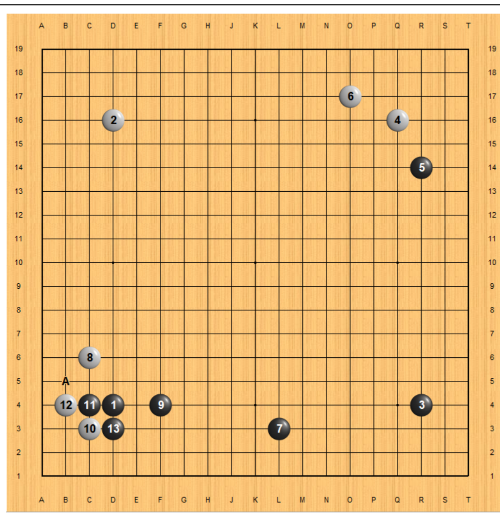


Figure 22. The 1st AlphaGo vs. AlphaGo (∼13)

For this example, Fan Hui's commentary again represents exactly how humans feel: "... Black 13 struck Gu Li and Zhou Ruiyang as highly unusual. They wondered, "Will this be enough?" A professional would normally block at A". The obsessive prejudice that such a lenient move would not lead us to victory make us avoid the moves like Black 13 almost instinctively.

Gu Li and Zhou Ruiyang's question "Will this be enough?" suggests a possibility that expected utility theory and prospect theory may be working together complementing each other unconsciously in their decision making. For, the complex question "Will this be enough?" expresses implicitly that even though we should select moves based on exact judgment on the profits and losses that would be resulted by a sequence of moves in accordance with the model of expected utility theory, we may sometimes, depending on the situation, i.e., in case where we would be surely leading the game even if we swallow a certain amount of loss, follow prospect theory. Since prospect theory may not be widely known among Baduk players, as is expected utility theory, the fact that Baduk lovers can be sympathetic with Gu Li and Zhou Ruiyang's question seems to corroborate the point that the basic idea, which is tantamount to prospect theory, has been also accepted by Baduk players as a part of common sense.

What one may find more frequently in such a situation is shown in figure 23 by Fan Hui:

> "This is one common line of play. Black gains outside influence, while White lives in the corner in sente. White concludes by extending at 12. Generally speaking, while most professionals would slightly prefer this fast-paced opening for White, few would be willing to accept such a simple result for Black as the game line" [10].

Fan Hui's expectation regarding professional players' disposition has a point, though we do not have any statistical evidence. It is extremely difficult to change such a well-entrenched disposition, which has become almost instincts. This certainly provides us with intriguing clue for the future research in formal epistemology on the belief revision in Baduk.

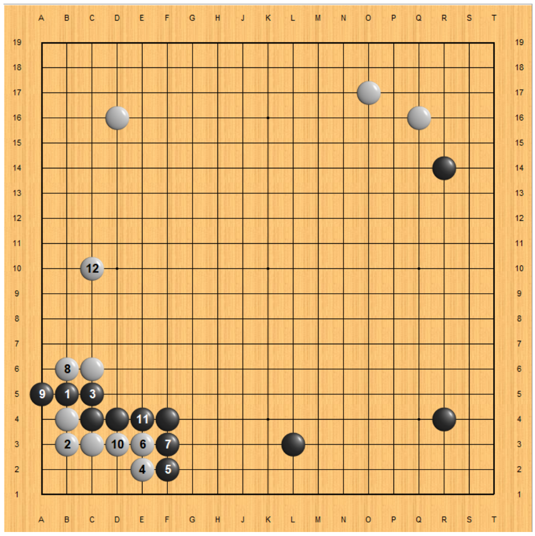


Figure 23. A variation reference to figure 22

5.5 Case 3: White 30 in AlphaGo's First self-play Game

White 30 in figure 24 shows a shape similar to Black 13 in the Case 2 (figure 22). In both cases, White plays a somewhat defensive safe move when it can cut the opponents group into two parts. Due to the entirely different arrangements of the other stones in the neighborhood, however, White 30 looks more surprising. What could be troublesome in Case 2 is that Black's moyo is too wide. On the other hand, the problem is that the Black's moyo is too close to White's group so that the appearance of White's group resulting from White 30 is too stingy. Fan Hui expresses well the astonishment regarding this move as follows:

> "Although a few rare moves have appeared so far, all of them are understandable. After seeing White's turn at 30, however, our shock was palpable. This move is so defensive! No human would play like this! White's playing style looks extremely cautious. Why not the hane?" [10].

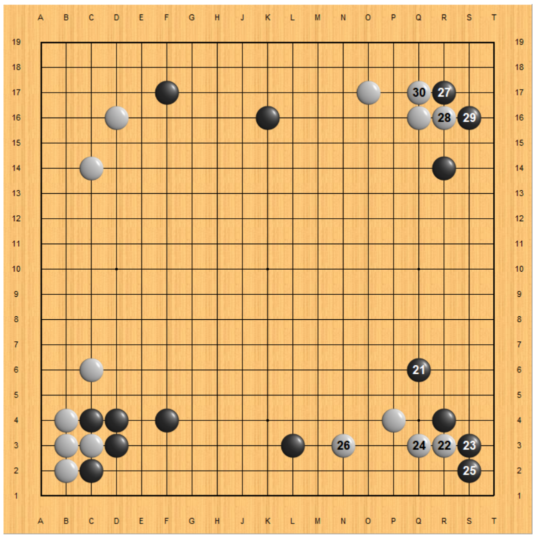


Figure 24. The 1st AlphaGo vs. AlphaGo (∼30, numbers displayed for 21∼30)

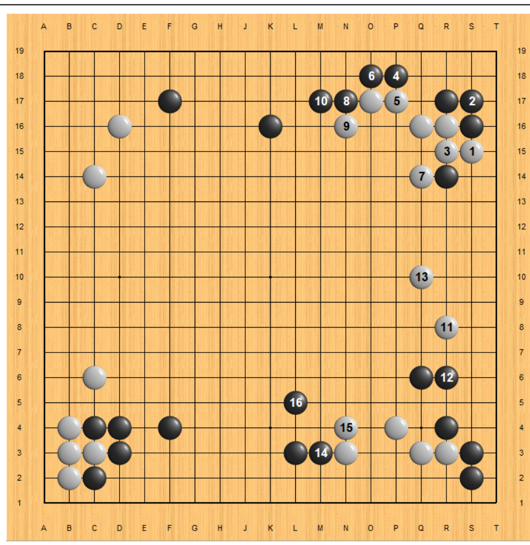


Figure 25. A variation reference to figure 24

As an answer to the question, Fan Hui suggests the following comments:

> "If White were to hane, the ensuing position would resemble figure 25. Black's bump at 14 is a smart, pragmatic way to fix the shape. White's formation on the right looks good, but on closer examination, it becomes apparent that Black is not only doing fine on points, but is also very thick. Is this variation actually good for White? With this analysis, we began to understand White's reasoning, but Gu Li and Zhou Ruiyang emphasized that this would be a very difficult prescription for professional players to swallow" [10].

Why did Gu Li and Zhou Ruiyang think that it would be a very difficult prescription for professional players to swallow? If humans choose moves in accordance with the model of expected utility theory, insofar as they understand and accept the result of figure 25, there would be no difficulty at all to make the same decision as that of AlphaGo. For, just as usual, all they have to do is to follow the calculation on expected utility and proceed as in figure 25. Therefore, it could be the case that Gu Li and Zhou Ruiyang implicitly have claimed that humans do not choose moves in accordance with the model of expected utility theory.

Then, could this example allow better explanation by prospect theory? Not necessarily! Though White 30 in Case 3 (figure 24) is not the move as a result of exact calculation of gain and loss in terms of expected utility theory, it is so difficult to view it as a result of being convinced that White has a superior status in accordance with prospect theory. It seems much more persuasive to explain it by appealing to a folk psychology that the move is too defensive and narrow minded to play.

5.6 Prospect Theory and the Operating Principles of AlphaGo

Above we claimed that the process of AlphaGo's decision making is remarkably similar to the model of decision making in prospect theory. If such a similarity is not merely a coincidence, and if DeepMind AlphaGo Team in fact set up the constituting and Operating principles of AlphaGo a la prospect theory, the implication and the possible effect would be far-reaching. Of course, DeepMind AlphaGo Team never made such an announcement. Nor the value of our claim would become different depending on such a testimony. However, it is not a trivial point that not only the decision making model following the prospect theory provides us with an excellent perspective for interpreting AlphaGo's Baduk games but there is also a structural similarity between such a model and the principles of constitution and operation of AlphaGo. For, in Kahneman and Tversky's 1979 paper we find the following remarks:

> "Prospect theory distinguishes two phases in the choice process: an early phase of editing and a subsequent phase of evaluation. The editing phase consists of a preliminary analysis of the offered prospects, which often yields a simpler representation of these prospects. In the second phase, the edited prospects are evaluated and the prospect of highest value is chosen" [14, p. 274].

We saw above that the efficient combination of the two neural networks, i.e., the policy networks and the value network, is (together with the Monte Carlo tree search) the essential element in understanding how AlphaGo is constituted and how it is operated. Now, we see that the roles and functions of the policy networks and the value network correspond beautifully the phases of editing and evaluating of prospect theory: "The function of the editing phase is to organize and reformulate the options so as to simplify subsequent evaluation and choice" [14].

6 Discussion

Is the process of decision making in humans entirely different from that of AlphaGo? Or, they are basically the same, though there are some differences in degree in efficiency or precision? Our point of departure is also in some sense betrays dual aspects to questions like this. If they are incommensurable, it would be simply impossible to compare them. So, there may be some good reason (if not a faith in their commonalities) why we strive to find some common denominator between them. However, as is revealed in our initial approach to find a clue from allegedly unconventional moves of AlphaGo, we may have been implicitly conceding that there are also significant differences between them. For example, Inchul Bae [1, 2] raises the following incisive question: "If intuition is not peculiar to humans, how is the intuition of artificial intelligence different from that of humans? Is it reducible to the calculating power armed with strong value network?" And, such a way of posing a question is not alien or strange at all to us.

Surprisingly, however, our research indicates the possibility that both humans and AlphaGo make decisions in accordance with the model of decision making based on prospect theory. Even as a descriptive model regarding human decision making, prospect theory is still merely a challenger to the dominant expected utility theory. In such a situation, our research goes one step further to claim that AlphaGo too makes decision in terms of prospect theory. So, it is not so difficult to anticipate some spirited criticisms from all directions. Since AlphaGo turns out to be the undisputed winner in her match with humans, there seems to be enough ground to claim that prospect theory, which AlphaGo follows as a descriptive model, is also a normative model that presents the standard of rationality. By claiming so, we seem to invite more severe criticisms. As a result, instead of presenting ultimate answers to the crucial questions, our research seems to have a potential to provoke more controversies.

The second AlphaGo shock resulting from DeepMind's second monumental paper published in Nature seems to be a torch on the potential arsenal. As was pointed out above, in their first article in Nature, published immediately before the challenge match between Lee Sedol vs. AlphaGo, it was claimed that AlphaGo plays Baduk in similar ways as humans do. Also, AlphaGo's choice of moves were claimed be more intelligent for that reason. In other words, even then, there was a room for thinking that the fact that policy networks of AlphaGo is a neural network developed by supervised learning based on human ways of playing Baduk itself presupposes that human way of playing Baduk is rational, and eo ipso AlphaGo can make rational decision making. Of course, since AlphaGo, as she was combining policy networks developed through supervised learning based on data of human's game records and the neural networks such as policy network and value network trained by reinforcement learning, may not simply copy or imitate human behaviors in decision making. Nevertheless, it was possible to understand that if AlphaGo makes rational decisions, that is ultimately due to her human origin, i.e., rationality of humans [6, 7]. However, everything changed within a year and a half. AlphaGo Zero can give three handicap stones to AlphaGo Master, they say. But AlphaGo Master won games against the current world champion Ke Jie, professional 9 dan. So, it seems natural to conclude that humans are no longer AlphaGo's rivals. It is also natural to find the cause of such a huge gap in skills of Baduk in the qualitative differences between the process of decision making in humans and that of AlphaGo or AlphaGo Zero. The second Nature paper of DeepMind AlphaGo Team seems anxious to emphasize the fact that AlphaGo Zero, unlike AlphaGo, achieved such a high level of playing Baduk within a few days entirely depending on reinforcement learning without any supervised learning. This fact too seems to play a large role in making us to concede implicitly that there is a qualitative difference between the process of decision making in humans and that of artificial intelligence. By now, there is room for wondering whether the basic framework of understanding the situation has been changed in such a direction that everyone tends to trust the rationality of AlphaGo or AlphaGo Zero wholeheartedly, while assuming implicitly that the process of decision making in humans, who are far inferior to AlphaGo or AlphaGo Zero in playing Baduk, must be irrational.

One might want to understand what is going on as follows:

We have presumed or believed that the decision making in artificial intelligence must be rational, though there is no way to understand it. On the other hand, we have thought that, due to the various bias effects studied in psychology, human decision making cannot match the rationality of artificial intelligence. The undisputed victory of AlphaGo against the top Baduk players like Lee Sedol or Ke Jie fortified such a presumption (belief or even prejudice). Indeed, it amplified the worries of humans about losing jobs by artificial intelligence, computers, and robots in the age of 4th industrial revolution.

In this frame of thinking, we may say that artificial intelligence follows the model of decision making based on expected utility theory while humans follow the model of decision making based on prospect theory. In other words, when humans are satisfied by "small but sure gain", artificial intelligence pursues by thorough calculations "the maximum expected utility", and such a difference in decision making between them ultimately results in the unbridgeable gap between their abilities.

There is certainly a consistent and persuasive stream of thought in this possible argument. However, unlike such argument based on some implicit assumptions, we argued above that the process of decision in AlphaGo can be better explained in accordance with the model of decision making based on prospect theory. That means, our research has a potential that can not only appease our worries about life in the age of artificial intelligence but also emphasize the affinity of humans and artificial intelligence, thereby pursue collaborative research on unexplored realm. We also emphasized above the fact that expected utility theory has been accepted as a dominant model for rational decision making. We even cited explicitly some examples that show the strong influence of expected utility theory on the process of decision making of Baduk players (especially professional Baduk players) not only as a descriptive model but also as a normative model. It is well known that it takes incredibly long time for a dominant theory to be challenged and discarded from extensive researches in history of science or sociology of science. Thus, it is not so difficult to understand the tendency to sustain the belief that expected utility theory represents rationality even when many prejudices or fixed ideas are forced to be revised by the two AlphaGo shocks. After all, for some reasons, the implicit assumption that the process of decision making in AlphaGo is entirely different from that of humans may be more deeply entrenched in our thinking. However, can this implicit assumption be supported by any ground?

There is rivalry between expected utility theory and prospect theory as the descriptive model of decision making in humans and AlphaGo. For argument's sake, let us suppose that these are the only options. Now there are four possible positions. (1) the position, according to which both humans and AlphaGo make decisions according to prospect theory. (2) the position, according to which, humans follow the model of prospect theory, while AlphaGo follows expected utility theory. (3) the position, according to which humans follow expected utility theory, while AlphaGo follows prospect theory. (4) the position, according to which both humans and AlphaGo follow expected utility theory. In this article, we have sustained (1) with sufficient supporting grounds.

On the other hand, if these rival hypotheses are considered as the normative model of decision making, there would be also four different possible positions: (1) the position, according to which it is rational for both humans and AlphaGo to make decisions according to prospect theory. (2) the position, according which it is rational for humans to follow the model of prospect theory, while it is rational for AlphaGo to follow that of expected utility theory. (3) the position, according to which it is rational for humans to follow expected utility theory, while it is rational for AlphaGo to follow prospect theory. (4) the position, according to which it is rational for both humans and AlphaGo to follow expected utility theory. In case we believe that AlphaGo's victory over humans guarantees that the process of decision making in AlphaGo is rational, our claim that AlphaGo makes decisions according to the model of decision making based on prospect theory seems to present strong ground for (1), which claims that it is rational for both humans and AlphaGo to follow prospect theory.

Of course, our conclusion that, both as a descriptive model and as a normative model, prospect theory provides us with the model of decision making for both humans and AlphaGo may invite several interpretations, including very severe criticisms. Suppose that one is basically sympathetic with our result but alert to the possible problems. In that case, one may criticize our conclusion as follows.

If one examines carefully AlphaGo's impressive unconventional moves, AlphaGo also seems to makes decisions following a model more appropriate to prospect theory rather than expected utility theory. In other words, contrary to humans' prejudice or implicit belief, the process of decision making in both humans and artificial intelligence is explainable by one and the same prospect theory. Then, as was demonstrated by AlphaGo, prospect theory is not merely descriptive model but also a normative model of rational decision making. That means, prospect theory does not merely explain phenomena retrospectively. There could be an equation that can predict the rational decision making in the future.

It should be explained why there is a gap between humans and AlphaGo, even if both are exploiting the same model of decision making, i.e., prospect theory. Substantial discussion would be possible only on the premise that prospect theory provides us with a normative model with such an equation. In other words, the difference between the rationality of human decision making and that of artificial intelligence *must be sought in their prospect functions. The difference of certain terms or coefficient values suggested by cumulative prospect theory influences the performance of applying the model of decision making, if so, we may say that the shape of AlphaGo's prospect function (i.e., the function determined by terms and coefficients) was superior to Lee Sedol's or Ke Jie's prospect functions.*

Here, two interesting projects emerge:

1. *Could there be domain-specific perfect prospect function? Depending on the domain (e.g. Baduk, music, sock market), shouldn't the best prospect function be different? If so, it might be possible to explain why in some other domains (e.g. arts) artificial intelligence is inferior to human, though AlphaGo wins against humans in Baduk.*

2. *If so, how are we to determine the perfect prospect function in each domain? Can we find them by starting with the equation suggested by cumulative prospect theory as the core and testing it by simulation and the experimental results in each domain?*

Such a constructive criticism would not destroy our results. Rather it points out some respects of our result to be fortified. But one warning is in order. Our research was about AlphaGo, But the criticism unconsciously seems to mix up thoughts about AlphaGo with thoughts about AlphaGoZero. The ultimate equation that secures rationality by prospect theory as the normative model may be possible for AlphaGoZero. However, it would be impossible for AlphaGo. Further, even for AlphaGoZero, such an expectation seems too much. Perhaps logically omniscient agent may not need such an equation. Thus, our discussion aims at only those agents who are not logically omniscient (including AlphaGo and AlphaGoZero).

Further, such a criticism implicitly assumes that, when we say that the normative model secures rationality, the rationality has the perfection that is possible only for the perfect being. Of course, such a usage is not entirely impossible having certain values. The reason why we find the conventional distinction between prescriptive model and normative model in rationality debate may not be irrelevant to this [See 28, 29]. After all, the equation required by the criticism would be a prescriptive model individuated depending on the abilities and resources of a given agent, not normative model universally applicable to all agents. If our critic concedes this point, probably the power of the criticism would be weakened considerably.

If such a criticism is launched from the followers of expected utility theory, one might use to quoque argument, by running the risk of committing an informal fallacy. Suppose that expected utility theory is both a descriptive and normative model of decision making of Baduk players. In that case, nobody would feel the need to claim that, in order to explain the gap between Ke Jie, the current world champion, and a novice, they should use different descriptive model of decision making. Both Ke Jie and the novice believe that expected utility theory is the descriptive and normative model of decision making at the same time. Nevertheless, the difference between them in terms of their abilities and skills is enormous. Explanation of what makes such a huge difference in abilities and skills between them cannot be found not in their different model of decision making. It should be found rather in how well they apply the same model of decision making based on expected utility theory. Likewise, the explanation of why AlphaGo and AlphaGoZero win Baduk games against humans by superior decision making should be found in not in their different decision making model but in how they apply well the same prospect theory to actual games.

Finally, it is not necessary to explain the difference in skills between humans and AlphaGo in their different models of decision making. In fact, after having contrasted the two rival theories of enthymeme, Park [21] tried to analyze situations at each move of Baduk game as the problem of interpreting enthymemes. According to this analysis, AlphaGo consistently applied the superior theory of enthymeme throughout the games, while Lee Sedol, even though sympathetic with the superior theory, was not completely freed from the influence of inferior theory of enthymeme. As a consequence, Lee Sedol lost the games by failing to solve the problem of interpreting enthymemes, i.e., by failing to apply the superior theory consistently. This is not to claim Park's interpretation is the only possible interpretation. This is merely to indicate that there may be room for many other possible interpretations. Another example of some such possible alternative explanation would be the development of the method of testing Baduk skills similar to various intelligence tests.

References

[1] I. Bae (2017a), "Aesthetics of 'AlphaGo's Baduk' through Deleuze", *LeMonde Diplomatique*, May issue. (in Korean)

[2] I. Bae (2017b), "How to Understand AlphaGo's Language?", *LeMonde Diplomatique*, September issue. (in Korean)

[3] J. L. Bermúdez, (2009), *Decision Theory and Rationality*, Oxford: Oxford University Press.

[4] N. C. Cho (1963), *Basic Haengma*, Seoul: BobmunSa. (in Korean)

[5] E. Cudd, (1993), "Game Theory and the History of Ideas and our Rationality", *Economics and Philosophy* 9, 101-133.

[6] Jonathan St. B. T. Evans (2007), *Hypothetical Thinking*, Home and New York: Psychology Press.

[7] Jonathan St. B. T. Evans and David E. Over (1996), *Rationality and Reasoning*, Home and New York: Psychology Press.

[8] S. Ghosh, B. Meijering, & R. Verbrugge, (2014), "Strategic Reasoning: Building Cognitive Models from Logical Formulas", *J Log Lang Inf* (2014) 23:1–29.

[9] M. P. Hong (2016), *AlphaGo Vs. Lee Sedol*, Seoul: Yisang Media (in Korean)

[10] Fan Hui (2016a), Commentary on AlphaGo's Self-Play, Available at https://deepmind.com/research/alphago/match-archive/alphago-games-english/

[11] Fan Hui (2016b), Commentary on AlphaGo Vs. Lee Sedol, Available at https://-deepmind.com/research/alphago/match-archive/alphago-games-english/

[12] R. Jeffrey (1983), *The Logic of Decision*, New York: McGraw-Hill.

[13] S.-H. Jeong et al. (2017), *The Theory of Baduk Techniques*, Seoul: Myongji University Press. (in Korean)

[14] D. Kahneman and A. Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk", *Econometrica* 47.2, 263-291.

[15] D. G. Kam (2016), *Understanding Artificial Intelligence through Baduk*, Seoul: East-Asia Publishing. (in Korean)

[16] K. Koo et l. (2016), *Who Are You?: AlphaGo Anatomized by Korean National Baduk Team* 1 and 2, Seoul: Korea Baduk Association. (in Korean)

[17] B. D. Lee (2017), "The Evolution of Computer Baduk and the Proposal for Its Development", *Monthly Baduk*, March issue. (in Korean)

[18] H. B. Lee (1997), *The Elements of Baduk*, Seoul: JeonwonmoonhwaSa. (in Korean)

[19] S. On and S-H. Jeong (2016), "An Analysis of AlphaGo's Unusual Moves", *Journal of Baduk Studies* 13.2, 11-27. (in Korean)

[20] W. Park (2002), *Philosophy of Baduk*, Seoul: Dongyeon, (in Korean)

[21] W. Park (2016), "Enthymematic Interaction in Baduk", a paper presented in the International Workshop "Logical Foundations of Strategic Reasoning", forthcoming in *IfCoLog.*

[22] W. Park, (2017), "When Is a Strategy in Games?", forthcoming in *IfCoLog.*

[23] A. Perea, (2012). *Epistemic game theory: reasoning and choice.* Cambridge University Press, Cambridge.

[24] D. Silver et al. (2016), "Mastering the Game of Go with Deep Neural Networks and Tree Search", *Nature*, 529, 484-503.

[25] D. Silver et al. (2017), "Mastering the Game of Go without Human Knowledge", *Nature*, 550, 354-371.

[26] S. Singh, A. Okun and A. Jackson (2017), "Learning to Play Go from Scratch", *Nature* 550, 336-337.

[27] B. Skyrms, (1999), "Theories of Counter-factual and Subjunctive Conditionals in Contexts of Strategic Interaction", *Research in Economics* 53, 275-291.

[28] Chris Sraemer (2000), "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk", *Journal of Economic Literature* 38.2, 332-382.

[29] K. E. Stanovich (1999), *Who Is Rational? Studies of Individual Differences in Reason-*

ing, Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

[30] Richard S. Sutton and Andrew G. Barto, (2017), *Reinforcement Learning*, Second Edition, Complete Draft, November 5, 2017, Cambridge, MA: MIT Press.

[31] P. Thagard and E. Millgram (1997), "Inference to the Best Plan: Coherence Theory of Decision", in A. Ram and D. B. Leake (eds.), *Goal-driven Learning*, Cambridge, MA: MIT Press, 439-454.

[32] A. Törmänen, et al. (2017), *Invisible: The Games of AlphaGo*, Hebsacker Verlag, Germany.

[33] A. Tversky and D. Kahneman (2016), "Advances in Prospect Theory: Cumulative Representation of Uncertainty", H. ArlÃş-Costa et al. (eds.), *Readings in Formal Epistemology*, Heidelberg: Springer, pp. 493-519.

[34] J. Van Benthem, (2014). Logic in Games. Cambridge: The MIT Press.

[35] J. Von Neumann, (1928), "Zur Theorie der Gesellschaftsspiele", *Mathematische Annalen* 100, 295-320.

[36] J. Von Neumann, and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ.

[37] J. Woods, (2017a), "The logical foundations of strategic reasoning: Inconsistency management as a test case for logic", forthcoming in *IfCoLog*.

[38] J. Woods, (2017b), "What Strategicians Might Learn from the Common Law: Implicit and Tacit Understandings of the Unwritten", forthcoming in *IfCoLog*.

[39] H. Yang (2016), *78 Divine Move, Human Move*, Seoul: CheoumSa (in Korean)

[40] Yuan Zhou, (2016), *AlphaGo Vs. Lee Sedol*, Richmond, VA: Slate and Shell.

Received 6 April 2018

Strengthening Gossip Protocols using Protocol-Dependent Knowledge

Hans van Ditmarsch
CNRS, LORIA, University of Lorraine, France & ReLaX, Chennai, India
hans.van-ditmarsch@loria.fr

Malvin Gattinger
University of Groningen, The Netherlands
malvin@w4eg.eu

Louwe B. Kuijer
University of Liverpool, United Kingdom
louwe.kuijer@liverpool.ac.uk

Pere Pardo
Ruhr-Universität Bochum, Germany
pere.pardo.v@gmail.com

Abstract

Distributed dynamic gossip is a generalization of the classic telephone problem in which agents communicate to share secrets, with the additional twist that also telephone numbers are exchanged to determine who can call whom. Recent work focused on the success conditions of simple protocols such as "Learn New Secrets" (*LNS*) wherein an agent a may only call another agent b if a does not know b's secret. A protocol execution is successful if all agents get to know all secrets. On partial networks these protocols sometimes fail because they ignore information available to the agents that would allow for better coordination. We study how epistemic protocols for dynamic gossip can be strengthened, using epistemic logic as a simple protocol language with a new operator for protocol-dependent knowledge. We provide definitions of different strengthenings and show that they perform better than *LNS*, but we also prove that there is no strengthening of *LNS* that always terminates successfully. Together, this gives us a better picture of when and how epistemic coordination can help in the dynamic gossip problem in particular and distributed systems in general.

This work is based on chapter 6 entitled "Dynamic Gossip" of Malvin Gattinger's PhD thesis [19]. Malvin Gattinger is the corresponding author, and was affiliated to the University of Amsterdam during part of this work. We would like to thank the anonymous IfCoLog referees for their helpful feedback and suggestions.

1 Introduction

The so-called *gossip problem* is a problem about peer-to-peer information sharing: a number of agents each start with some private information, and the goal is to share this information among all agents, using only peer-to-peer communication channels [38]. For example, the agents could be autonomous sensors that need to pool their individual measurements in order to obtain a joint observation. Or the agents could be distributed copies of a database that can each be edited separately, and that need to synchronize with each other [18, 21, 28].

The example that is typically used in the literature, however, is a bit more frivolous: as the name suggests, the gossip problem is usually represented as a number of people *gossiping* [24, 16, 15]. This term goes back to the oldest sources on the topic, such as [6]. The gossip scenario gives us not only the name of the gossip problem, but also the names of some of the other concepts that are used: the private information that an agent starts out with is called that agent's *secret*, the communication between two agents is called a *telephone call* and an agent a is capable of contacting another agent b if *a knows b's telephone number*.

These terms should not be taken too literally. Results on the gossip problem can, in theory, be used by people that literally just want to exchange gossip by telephone. But we model information exchange in general and ignore all other social and fun aspects of gossip among humans — although these aspects can also be modeled in epistemic logic [30].

For our framework, applications where artificial agents need to synchronize their information are much more likely. For example, recent ideas to improve cryptocurrencies like bitcoin and other blockchain applications focus on the peer-to-peer exchange (gossip) happening in such networks [36] or even aim to replace blockchains with directed graphs storing the history of communication [5]. Epistemic logic can shed new light on the knowledge of agents participating in blockchain protocols [22, 10].

There are many different sets of rules for the gossip problem [24]. For example, calls may be one-on-one, or may be conference calls. Multiple calls may take place in parallel, or must happen sequentially. Agents may only be allowed to exchange one secret per call, or exchange everything they know. Information may go both ways during a call, or only in one direction. We consider only the most commonly studied set of rules: calls are one-on-one, calls are sequential, and the callers exchange all the secrets they know. So if a call between a and b is followed by a call between b and c, then in the second call agent b will also tell agent c the secret of agent a.

The goal of gossip is that every agent knows every secret. An agent who knows all secrets is called an *expert*, so the goal is to turn all agents into experts.

The *classical* gossip problem, studied in the 1970s, assumed a total communication network (anyone could call anyone else from the start), and focused on optimal call sequences, i.e. schedules of calls which spread all the secrets with a minimum number of calls, which happens to be $2n - 4$ for $n \geq 4$ agents [38, 27]. Later, this strong assumption on the network of the gossiping agents was dropped, giving rise to studies on different network topologies (see [24] for a survey), with $2n - 3$ calls sufficing for most networks.

Unfortunately, these results about optimal call sequences only show that such call sequences exist. They do not provide any guidance to the agents about how to achieve an optimal call sequence. Effectively, these solutions assume a central scheduler with knowledge of the entire network, who will come up with an optimal schedule of calls, to be sent to the agents, who will eventually execute it in the correct order. Most results also rely upon synchrony so that agents can execute their calls at the appropriate time (i.e. after some calls have been made, and before some other calls are made).

The requirement that there be a central scheduler that tells the agents exactly what to do, is against the spirit of the peer-to-peer communication that we want to achieve. Computer science has shifted towards the study of *distributed algorithms* for the gossip problem [23, 29]. Indeed, the gossip problem becomes more natural without a central scheduler; the gossiping agents try to do their best with the information they have when deciding whom to call. Unfortunately, this can lead to sequences of calls that are redundant because they contain many calls that are uninformative in the sense that neither agent learns a new secret. Additionally, the algorithm may fail, i.e., it may deadlock, get stuck in a loop or terminate before all information has been exchanged.

For many applications it is not realistic to assume that every agent is capable of contacting every other agent. So we assume that every agent has a set of agents of which they "know the telephone number", their neighbors, so to say, and that they are therefore able to contact. We represent this as a directed graph, with an edge from agent a to agent b if a is capable of calling b.

In classical studies, this graph is typically considered to be unchanging. In more recent work on *dynamic gossip* the agents exchange both the secrets and the numbers of their contacts, therefore increasing the connectivity of the network [16]. We focus on dynamic gossip. In distributed protocols for dynamic gossip all agents decide on their own whom to call, depending on their current information [16], or also depending on the expectation for knowledge growth resulting from the call [15]. The latter requires agents to represent each other's knowledge, and thus epistemic logic.

Different protocols for dynamic gossip are successful in different classes of gossip networks. The main challenge in designing such a protocol is to find a good level of redundancy: we do not want superfluous calls, but the less redundant a gossip protocol, the easier it fails in particular networks. Another challenge is to keep the protocol simple. After all, a protocol that requires the agents to solve a computationally hard problem every time they have to decide whom to call next, would not be practical. There is also a trade-off between the content of the message of which a call consists, and the expected duration of gossip protocols. A nice example of that is [25], wherein the minimum number of calls to achieve the epistemic goal is reduced from quadratic to linear order, however at the price of more 'expensive' messages, not only exchanging secrets but also knowledge about secrets.

A well-studied protocol is "Learn New Secrets" (*LNS*), in which agents are allowed to call someone if and only if they do not know the other's secret. This protocol excludes redundant calls in which neither participant learns any new secrets. As a result of this property, all *LNS* call sequences are finite. For small numbers of agents, it therefore has a shorter expected execution length than the "Any Call" (ANY) protocol that allows arbitrary calls at all times and thus allows infinite call sequences [14]. Additionally, it is easy for agents to check whom they are allowed to call when following *LNS*. However, *LNS* is not always successful. On some graphs it can terminate unsuccessfully, i.e. when some agents do not yet know all secrets. In particular there are graphs where the outcome depends on how the agents choose among allowed calls [16].

Fortunately, it turns out that failure of *LNS* can often be avoided with some forethought by the calling agents. That is, if some of the choices available to the agents lead to success and other choices to failure, it is often possible for the agents to determine in advance which choices are the successful ones. This leads to the idea of *strengthening* a protocol. Suppose that P is a protocol that, depending on the choices of the agents, is sometimes successful and sometimes unsuccessful. A strengthening of P is an addition to P that gives the agents guidance on how to choose among the options that P gives them.

The idea is that such a strengthening can leave good properties of a protocol intact, while reducing the chance of failure. For example, any strengthening of *LNS* will inherit the property that there are no redundant calls: It will still be the case that agents only call other agents if they do not know their secrets.

Let us illustrate this with a small example, also featuring as a running example in the technical sections (see Figure 1 on page 13). There are three agents a, b, c. Agent a knows the number of b, and b and c know each other's number. Calling agents exchange secrets and numbers, which may expand the network, and they apply the *LNS* protocol, wherein you may only call other agents if you do not know their secret. If a calls b, it learns the secret of b and the number of c. All different ways to make further calls now result in all three agents knowing all secrets. If the first call is between b and c (and there are no other first calls than ab, bc, and cb), they learn each other's secret but no new number. The only possible next call now is ab, after which a and b know all secrets but not c. But although a now knows c's number, she is not permitted to call c, as she already learned c's secret by calling b. We are stuck. So, some executions of *LNS* on this graph are successful and others are unsuccessful. Suppose we now strengthen the *LNS* protocol into LNS' such that b and c have to wait before making a call until they are called by another agent. This means that b will first receive a call from a. Then all executions of LNS' are successful on this graph. In fact, there is only *one* remaining execution: $ab; bc; ac$. The protocol LNS' is a *strengthening* of the protocol *LNS*.

The main contributions of this paper are as follows. We define what it means that a gossip protocol is common knowledge between all agents. To that end we propose a logical semantics with an individual knowledge modality for protocol-dependent knowledge. We then define various strengthenings of gossip protocols, both in the logical syntax and in the semantics. This includes a strengthening called uniform backward induction, a form of backward induction applied to (imperfect information) gossip protocol execution trees. We give some general results for strengthenings, but mainly apply our strengthenings to the protocol *LNS*: we investigate some basic gossip graphs (networks) on which we gradually strengthen *LNS* until all its executions are successful on that graph. However, no such strengthening will work for all gossip graphs. This is proved by a counterexample consisting of a six-agent gossip graph, that requires fairly detailed analysis. Some of our results involve the calculation and checking of large numbers of call sequences. For this we use an implementation in Haskell.

Our paper is structured as follows. In Section 2 we introduce the basic definitions to describe gossip graphs and a variant of epistemic logic to be interpreted on them. In particular, Subsection 2.3 introduces a new operator for protocol-dependent knowledge. In Section 3 we define semantic and — using the new operator — syntactic ways to strengthen gossip protocols. We investigate how successful those strengthenings are and study their behavior under iteration. Section 4 contains our main result, that strengthening *LNS* to a strongly successful protocol is impossible. In Section 5 we wrap up and conclude. The Appendix describes the Haskell code used to support our results.

2 Epistemic Logic for Dynamic Gossip Protocols

2.1 Gossip Graphs and Calls

Gossip graphs are used to keep track of who knows which secrets and which telephone numbers.

Definition 1 (Gossip Graph)**.** *Given a finite set of agents A, a* gossip graph *G is a triple (A, N, S) where N and S are binary relations on A such that $I \subseteq S \subseteq N$ where I is the identity relation on A. An* initial gossip graph *is a gossip graph where $S = I$. We write $N_a b$ for $(a, b) \in N$ and N_a for $\{b \in A \mid N_a b\}$, and similarly for the relation S. The set of all initial gossip graphs is denoted by $\mathcal{G}$.*

The relations model the basic knowledge of the agents. Agent a *knows the number* of b iff $N_a b$ and a *knows the secret* of b iff $S_a b$. If we have $N_a b$ and not $S_a b$ we also say that a knows the *pure number* of b.

Definition 2 (Possible Call; Call Execution)**.** *A* call *is an ordered pair of agents $(a, b) \in A \times A$. We usually write ab instead of (a, b). Given a gossip graph G, a call ab is* possible *iff $N_a b$. Given a possible call ab, G^{ab} is the graph (A', N', S') such that $A' := A$, $N'_a := N'_b := N_a \cup N_b$, $S'_a := S'_b := S_a \cup S_b$, and $N'_c := N_c$, $S'_c := S_c$ for $c \neq a, b$. For a sequence of calls $ab; cd; \ldots$ we write σ or τ. The empty sequence is ϵ. A sequence of possible calls is a* possible call sequence. *We extend the notation G^{ab} to possible call sequences by $G^\epsilon := G$ and $G^{\sigma;ab} := (G^\sigma)^{ab}$. Gossip graph G^σ is the result of* executing *σ in G.*

To visualize gossip graphs we draw N with dashed and S with solid arrows. When making calls, the property $S \subseteq N$ is preserved, so we omit the dashed N arrow if there already is a solid S arrow.

Example 3. *Consider the following initial gossip graph G in which a knows the number of b, and b and c know each other's number and no other numbers are known:*

$$\begin{tikzcd} a \arrow[r, dashed] & b \arrow[r, dashed, leftrightarrow] & c \end{tikzcd}$$

Suppose that a calls b. We obtain the gossip graph G^{ab} in which a and b know each other's secret and a now also knows the number of c:

$$\begin{tikzcd} a \arrow[r, leftrightarrow] \arrow[rr, dashed, bend left] & b \arrow[r, dashed, leftrightarrow] & c \end{tikzcd}$$

2.2 Logical Language and Protocols

We now introduce a logical language which we will interpret on gossip graphs. Propositional variables $N_a b$ and $S_a b$ stand for "agent a knows the number of agent b" and "agent a knows the secret of agent b", and $\top$ is the 'always true' proposition. Definitions 4 and 5 are by simultaneous induction, as the language construct $K_a^P \varphi$ refers to a protocol P.

Definition 4 (Language)**.** *We consider the language $\mathcal{L}$ defined by*

$$\begin{array}{lcl} \varphi & ::= & \top \mid N_a b \mid S_a b \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a^P \varphi \mid [\pi]\varphi \\ \pi & ::= & ?\varphi \mid ab \mid (\pi\ ;\ \pi) \mid (\pi \cup \pi) \mid \pi^* \end{array}$$

where $a, b \in A$. Members of $\mathcal{L}$ of type φ are formulas *and those of type π are* programs*.*

Definition 5 (Syntactic protocol)**.** *A* syntactic protocol *P is a program defined by*

$$P := \left(\bigcup_{a \neq b \in A} (?(N_a b \wedge P_{ab}); ab) \right)^* ; ? \bigwedge_{a \neq b \in A} \neg \left(N_a b \wedge P_{ab}\right)$$

where for all $a \neq b \in A$, $P_{ab} \in \mathcal{L}$ is a formula. This formula is called the protocol condition *for call ab of protocol P. The notation P_{ab} means that a and b are designated variables in that formula.*

Other logical connectives and program constructs are defined by abbreviation. Moreover, $N_a bcd$ stands for $N_a b \wedge N_a c \wedge N_a d$, and $N_a B$ for $\bigwedge_{b \in B} N_a b$. We use analogous abbreviations for the relation S. We write Ex_a for $S_a A$. We then say that agent a is an *expert*. Similarly, we write Ex_B for $\bigwedge_{b \in B} Ex_b$, and Ex for Ex_A: all agents are experts.

Construct $[\pi]\varphi$ reads as "after every execution of program π, φ (is true)." For program modalities, we use the standard definition for diamonds: $\langle\pi\rangle\varphi := \neg[\pi]\neg\varphi$, and further: $\pi^0 := ?\top$ and for all $n \in \mathbb{N}$, $\pi^n := \pi^{n-1}; \pi$.

Our protocols are *gossip* protocols, but as we define no other, we omit the word 'gossip'. The word 'syntactic' in syntactic protocol is to distinguish it from the semantic protocol that will be defined later. It is also often omitted.

Our new operator $K_a^P \varphi$ reads as "given the protocol P, agent a knows that φ". Informally, this means that agent a knows that φ on the assumption that it is common knowledge among the agents that they all use the gossip protocol P. The epistemic dual is defined as $\hat{K}_a^P \varphi := \neg K_a^P \neg\varphi$ and can be read as "given the protocol P, agent a considers it possible that φ."

We note that the language is well-defined, in particular K_a^P. The only variable parts of a protocol P are the protocol conditions P_{ab}. Hence, given $|A|$ agents, and the requirement that $a \neq b$, a protocol is determined by its $|A| \cdot (|A| - 1)$ many protocol conditions. We can therefore see the construct $K_a^P \varphi$ as an operator with input $(|A| \cdot (|A| - 1)) + 1$ objects of type formula (namely all these protocol condition formulas plus the formula φ in $K_a^P \varphi$), and as output a more complex object of type formula (namely $K_a^P \varphi$).[1]

Note that this means that all knowledge operators in a call condition P_{ab} of a protocol P must be relative to protocols strictly simpler than P. In particular, the call condition P_{ab} cannot contain the operator K_a^P, although it may contain $K_a^{P'}$ where P' is less complex than P. So the language is incapable of describing the "protocol" X given by "a is allowed to call b if and only if a knows, assuming that X is common knowledge, that b does not know a's secret." This is intentional; the "protocol" X is viciously circular so we do not want our language to be able to represent it.

Example 6. *The "Learn New Secrets" protocol (LNS) is the protocol with protocol conditions $\neg S_a b$ for all $a \neq b \in A$. This prescribes that you are allowed to call any agent whose secret you do not yet know (and whose number you already know). The "Any Call" protocol (ANY) is the protocol with protocol conditions $\top$ for all $a \neq b \in A$. You are allowed to call any agent whose number you know.*

The standard epistemic modality is defined by abbreviation as $K_a \varphi := K_a^{ANY} \varphi$.

2.3 Semantics of Protocol-Dependent Knowledge

We now define how to interpret the language $\mathcal{L}$ on gossip graphs. A *gossip state* is a pair (G, σ) such that G is an initial gossip graph and σ a call sequence possible on G (see Def. 2). We recall that G and σ induce the gossip graph $G^\sigma = (A, N^\sigma, S^\sigma)$. This is called the gossip graph *associated* with gossip state (G, σ). The semantics of $\mathcal{L}$ is with respect to a given initial gossip graph G, and defined on the set of gossip states (G, σ) for all σ possible on G. Definitions 7 and 8 are simultaneously defined.

Definition 7 (Epistemic Relation)**.** *Let an initial gossip graph $G = (A, N, S)$ and a protocol P be given. We inductively define the* epistemic relation $\sim_a^P$ *for agent a over gossip states (G, σ), where $G^\sigma = (A, N^\sigma, S^\sigma)$ are the associated gossip graphs.*

[1] Alternatively one could define a *protocol condition function* $f \colon A^2 \to \mathcal{L}$ and proceed as follows. In the language BNF replace $K_a^P \varphi$ by $K_a(\vec{\varphi_{ab}}, \varphi)$ where $a \neq b$ and $\vec{\varphi_{ab}}$ is a vector representing $|A| \cdot (|A| - 1)$ arguments, and in the definition of protocol replace P_{ab} by $f(a, b)$. That way, Definition 4 precedes Definition 5 and is no longer simultaneously defined. Then, when later defining the semantics of $K_a(\vec{\varphi_{ab}}, \varphi)$, replace all φ_{ab} by $f(a, b)$.

1. $(G, \epsilon) \sim_a^P (G, \epsilon)$;
2. *if* $(G, \sigma) \sim_a^P (G, \tau)$, $N_b^\sigma = N_b^\tau$, $S_b^\sigma = S_b^\tau$, *and* ab *is* P*-permitted at* (G, σ) *and at* (G, τ), *then* $(G, \sigma; ab) \sim_a^P (G, \tau; ab)$;
   if $(G, \sigma) \sim_a^P (G, \tau)$, $N_b^\sigma = N_b^\tau$, $S_b^\sigma = S_b^\tau$, *and* ba *is* P*-permitted at* (G, σ) *and at* (G, τ), *then* $(G, \sigma; ba) \sim_a^P (G, \tau; ba)$;
3. *if* $(G, \sigma) \sim_a^P (G, \tau)$ *and* $c, d, e, f \neq a$ *such that* cd *is* P*-permitted at* (G, σ) *and* ef *is* P*-permitted at* (G, τ), *then* $(G, \sigma; cd) \sim_a^P (G, \tau; ef)$.

Definition 8 (Semantics)**.** *Let initial gossip graph* $G = (A, N, S)$ *be given. We inductively define the interpretation of a formula* $\varphi \in \mathcal{L}$ *on a gossip state* (G, σ), *where* $G^\sigma = (A, N^\sigma, S^\sigma)$ *is the associated gossip graph.*

$$
\begin{array}{lll}
G, \sigma \models \top & & \textit{always} \\
G, \sigma \models N_a b & \textit{iff} & N_a^\sigma b \\
G, \sigma \models S_a b & \textit{iff} & S_a^\sigma b \\
G, \sigma \models \neg\varphi & \textit{iff} & G, \sigma \not\models \varphi \\
G, \sigma \models \varphi \wedge \psi & \textit{iff} & G, \sigma \models \varphi \textit{ and } G, \sigma \models \psi \\
G, \sigma \models K_a^P \varphi & \textit{iff} & G, \sigma' \models \varphi \textit{ for all } (G, \sigma') \sim_a^P (G, \sigma) \\
G, \sigma \models [\pi]\varphi & \textit{iff} & G, \sigma' \models \varphi \textit{ for all } (G, \sigma') \in [\![\pi]\!](G, \sigma)
\end{array}
$$

where $[\![\cdot]\!]$ *is the following interpretation of programs as relations between gossip states. Note that we write* $[\![\pi]\!](G, \sigma)$ *for the set* $\{(G, \sigma') \mid ((G, \sigma), (G, \sigma')) \in [\![\pi]\!]\}$.

$$
\begin{array}{rcl}
[\![?\varphi]\!](G, \sigma) & := & \{(G, \sigma) \mid G, \sigma \models \varphi\} \\
[\![ab]\!](G, \sigma) & := & \{(G, (\sigma; ab)) \mid G, \sigma \models N_a b\} \\
[\![\pi; \pi']\!](G, \sigma) & := & \bigcup\{[\![\pi']\!](G, \sigma') \mid (G, \sigma') \in [\![\pi]\!](G, \sigma)\} \\
[\![\pi \cup \pi']\!](G, \sigma) & := & [\![\pi]\!](G, \sigma) \cup [\![\pi']\!](G, \sigma) \\
[\![\pi^*]\!](G, \sigma) & := & \bigcup\{[\![\pi^n]\!](G, \sigma) \mid n \in \mathbb{N}\}
\end{array}
$$

If $G, \sigma \models P_{ab}$ *we say that* ab *is* P-permitted *at* (G, σ). *A* P-permitted call sequence *consists of* P*-permitted calls.*

Let us first explain why the interpretation of protocol-dependent knowledge is well-defined. The interpretation of $K_a^P \varphi$ in state (G, σ) is a function of the truth of φ in all (G, τ) accessible via $\sim_a^P$. This is standard. Non-standard is that the relation $\sim_a^P$ is a function of the truth of protocol conditions P_{ab} in gossip states including (G, σ). This may seem a slippery slope. However, note that $K_a^P \varphi$ cannot be a subformula of any such P_{ab}, as the language $\mathcal{L}$ is well-defined: knowledge cannot be self-referential. These checks of P_{ab} can therefore be performed without vicious circularity.

Let us now explain an important property of $\sim_a^P$, namely that it only relates two gossip states if both are reachable by the protocol P. So if $(G, \sigma) \sim_a^P (G, \sigma')$ and σ is a P-permitted call sequence, then σ' is P-permitted as well. In other words, a assumes that no one will make any calls that are not P-permitted. The set $\{\sim_a^P \mid a \in A\}$ of relations therefore represents the information state of the agents under the assumption that it is common knowledge that the protocol P will be followed.

Given the logical semantics, a convenient primitive is the following *gossip model*.

Definition 9 (Gossip Model; Execution Tree)**.** *Given an initial gossip graph G, the* gossip model *for G consists of all* gossip states (G, σ) *(where, by definition of gossip states, σ is possible on G), with epistemic relations $\sim_a^P$ between gossip states. The* execution tree *of a protocol P given G is the submodel of the gossip model restricted to the set of those (G, σ) where σ is P-permitted.*

The relation $\sim_a^P$ is an equivalence relation on the restriction of a gossip model to the set of gossip states (G, σ) where σ is P-permitted. This is why we use the symbol $\sim$ for the relation. However, $\sim_a^P$ is typically not an equivalence relation on the entire domain of the gossip model, as $\sim_a^P$ is not reflexive on unreachable gossip states (G, σ).

In our semantics, the modality $[ab]$ can always be evaluated. There are three cases to distinguish. (i) If the call ab is not possible (if a does not know the number of b), then $[\![ab]\!](G, \sigma) = \emptyset$, so that $[ab]\varphi$ is trivially true for all φ. (ii) If the call ab is possible but not P-permitted, then $[\![ab]\!](G, \sigma) = \{(G, \sigma; ab)\}$ but $\sim_a^P (G, \sigma; ab) = \emptyset$, so that in such states $K_a^P \bot$ is true: the agent believes everything including contradictions. In other words, we have that $\neg P_{ab} \to [ab]K_c^P \bot$. (iii) If the call ab is possible and P-permitted, then $[\![ab]\!](G, \sigma) = \{(G, \sigma; ab)\}$ and $\sim_a^P (G, \sigma; ab) \neq \emptyset$ consists of the equivalence class of gossip states that are indistinguishable for agent a after call ab.

In view of the above, one might want to have a modality or program strictly standing for 'call ab is possible and P-permitted'. We can enforce protocol P for call ab by $[?P_{ab}; ab]\varphi$, for "after the P-permitted call ab, φ is true."

Let us now be exact in what sense the gossip model is a Kripke model. Clear enough, the set of gossip states (G, σ) constitute a *domain*, and we can identify the valuation of atomic propositions $N_a b$ (resp. $S_a b$) with the subset of the domain such that $(G, \sigma) \models N_a b$ (resp. $(G, \sigma) \models S_a b$). The relation to the usual accessibility relations of a Kripke model is less clear. For each agent a, we do not have a unique relation $\sim_a$, but parametrized relations $\sim_a^P$; therefore, in a way, there are as many relations for agent a as there are protocols P. These relations $\sim_a^P$ are only implicitly given. Given P, they can be made explicit if a semantic check of $K_a^P \varphi$ so requires.

Gossip models are reminiscent of the history-based models of [34] and of the protocol-generated forest of [9]. A gossip model is a protocol-generated forest (and similarly, the execution trees contained in the gossip model are protocol-generated forests), although a rather small forest, namely consisting of a single tree. An important consequence of this is that the agents initially have *common knowledge of the gossip graph.* For example, in the initial gossip graph of the introduction, depicted in Figure 1, agent a knows that agent c only knows the number of b. Other works consider uncertainty about the initial gossip graph (for example, to represent that agent a is uncertain whether c knows a's number), such that each gossip graph initially considered possible generates its own tree [15].

The gossip states (G, σ) that are the domain elements of the gossip model carry along a *history* of prior calls. This can, in principle, be used in a protocol language to be interpreted on such models, although we do not do this in this work. An example of such a protocol is the "Call Once" protocol described in [16]: call ab is permitted in gossip state (G, σ), if ab and ba do not occur in σ.

With respect to the protocol ANY the gossip model is not restricted. If we only were to consider the protocol ANY, to each agent we can associate a unique epistemic relation $\sim_a^{ANY}$ in the gossip model, for which we might as well write $\sim_a$. We now have a standard Kripke model. This justifies $K_a\varphi$ as a suitable abbreviation of $K_a^{ANY}\varphi$.

Definition 10 (Extension of a protocol)**.** *For any initial gossip graph G and any syntactic protocol P we define the* extension of P on G *by*

$$\begin{array}{lcl} P_0(G) & := & \{\epsilon\} \\ P_{i+1}(G) & := & \{\sigma; ab \mid \sigma \in P_i(G),\ a, b \in A,\ G, \sigma \models P_{ab}\} \\ P(G) & := & \bigcup_{i<\omega} P_k(G) \end{array}$$

The extension of P *is* $\{(G, P(G)) \mid G \in \mathcal{G}\}$*.*

Recall that $\mathcal{G}$ is the set of all initial gossip graphs. We often identify a protocol with its extension. To compare protocols we will write $P \subseteq P'$ iff for all $G \in \mathcal{G}$ we have $P(G) \subseteq P'(G)$.

Definition 11 (Success)**.** *Given an initial gossip graph G and protocol P, a P-permitted call sequence σ is* terminal *iff for all calls ab, $G, \sigma \not\models P_{ab}$. We then also say that the gossip state (G, σ) is* terminal*. A terminal call sequence is* successful *iff after its execution all agents are experts. Otherwise it is* unsuccessful.

- *A protocol P is* strongly successful *on G iff all terminal P-permitted call sequences are successful: $G, \epsilon \models [P]Ex$.*

- *A protocol is* weakly successful *on G iff some terminal P-permitted call sequences are successful: $G, \epsilon \models \langle P \rangle Ex$.*

- *A protocol is* unsuccessful *on G iff no terminal P-permitted call sequences are successful: $G, \epsilon \models [P]\neg Ex$.*

A protocol is strongly successful *iff it is strongly successful on all initial gossip graphs G, and similarly for weakly successful and unsuccessful.*

Instead of 'is successful' we also say '*succeeds*', and instead of 'terminal sequence' we also say that the sequence is *terminating*. Given a gossip graph G and a P-permitted sequence σ we say that the associated gossip graph G^σ is *P-reachable* (from G). A terminal P-permitted sequence is also called an *execution* of P. Given any set X of call sequences, $\overline{X}$ is the subset of the terminal sequences of X.

All our protocols can always be executed. If this is without making any calls, the protocol extension is empty. Being empty does not mean that $[P]\bot$ holds, which is never the case.

Strong success implies weak success, but not vice versa. Formally, we have that $[P]\varphi \to \langle P \rangle \varphi$ is valid for all protocols P, but $\langle P \rangle \varphi \to [P]\varphi$ is not valid in general, because our protocols are typically non-deterministic.

We can distinguish unsuccessful termination (not all agents know all secrets) from successful termination. In other works [16, 2] this distinction cannot be made. In those works termination implies success.

Example 12. *We continue with Example 3. The execution tree of LNS on this graph is shown in Figure 1. We denote calls with gray arrows and the epistemic relation with dotted lines. For example, agent a cannot distinguish whether call bc or cb happened. At the end of each branch the termination of LNS is denoted with ✓ if successful, and × if unsuccessful.*

To illustrate our semantics, for this graph G we have:

- $G, \epsilon \models N_a b \wedge \neg S_a b$ *— the call ab is LNS-permitted at the start.*

- $G, \epsilon \models [ab](S_a b \wedge S_b a)$ *— after the call ab the agents a and b know each other's secret*

- $G, \epsilon \models [ab]\langle ac \rangle \top$ *— after the call ab the call ac is possible.*

- $G, \epsilon \models [ab][LNS]Ex$ *— after the call ab the LNS protocol will always terminate successfully.*

- $G, \epsilon \models [bc \cup cb][LNS]\neg Ex$ *— after the calls bc or cb the LNS protocol will always terminate unsuccessfully.*

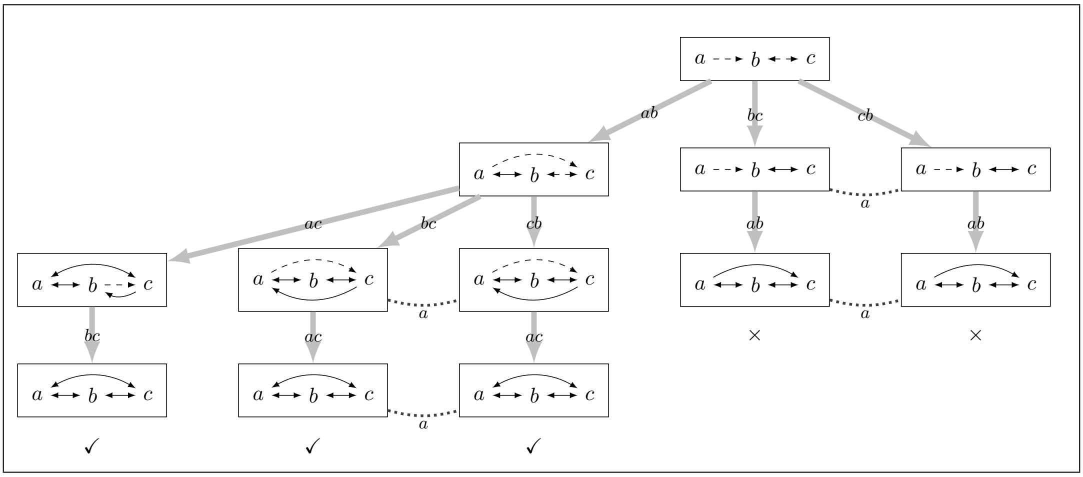


Figure 1: Example of an execution tree for *LNS*.

- $G, \epsilon \models [bc \cup cb]K_a^{LNS}(S_b c \wedge S_c b)$ — *after the calls bc or cb, agent a knows that b and c know each others secret.*

- $G, ab; bc; ac \models \bigwedge_{i \in \{a,b,c\}} K_i^{LNS} Ex$ — *after the call sequence* $ab; bc; ac$ *everyone knows that everyone is an expert.*

We only have epistemic edges for agent a, and those are between states with identical gossip graphs. If there are three agents, then if you are not involved in a call, you know that the other two agents must have called. You may only be uncertain about the direction of that call. But the direction of the call does not matter for the numbers and secrets being exchanged. Hence all agents always know what the current gossip graph is. For a more interesting epistemic relation, see Figure 2 in the Appendix.

2.4 Symmetric and epistemic protocols, and semantic protocols

Given a protocol P, for any $a \neq b$ and $c \neq d$, the protocol conditions P_{ab} and P_{cd} can be different formulas. So a protocol may require different agents to obey different rules. Although there are settings wherein this is interesting to investigate, we want to restrict our investigation to those protocols where there is one protocol condition to rule them all. This is enforced by the requirement of *symmetry*. Another requirement is that the calling agent should know that the protocol condition is satisfied before making a call. That is the requirement that the protocol be *epistemic*. It is indispensable in order to see our protocols as *distributed* gossip protocols.

Definition 13 (Symmetric and epistemic syntactic protocol)**.** *Let a syntactic protocol P be given. Protocol P is* symmetric *iff for every permutation J of agents, we have $\varphi_{J(a)J(b)} = J(\varphi_{ab})$, where $J(\varphi_{ab})$ is the natural extension of J to formulas.*[2] *Protocol P is* epistemic *iff for every $a, b \in A$, the protocol condition $P_{ab} \to K_a^P P_{ab}$ is valid. We henceforth require all our protocols to be symmetric and epistemic.*

Intuitively, a protocol is *epistemic* if callers always know when to make a call, without being given instructions by a central scheduler. This means that whenever P_{ab} is true, so agent a is allowed to call agent b, it must be the case that a knows that P_{ab} is true. In other words, in an epistemic protocol P_{ab} implies $K_a^P P_{ab}$. Furthermore, by Definition 8 knowledge is truthful on the execution tree for protocol P in gossip model. So except in the gossip states that cannot be reached using the protocol P, we also have that $K_a^P P_{ab}$ implies P_{ab}.

If a protocol is *symmetric* the names of the agents are irrelevant and therefore interchangeable. So a symmetric protocol is not allowed to "hard-code" agents to perform certain roles. This means that, for example, we cannot tell agent a to call b, as opposed to c, just because b comes before c in the alphabet. But we can tell a to call b, as opposed to c, on the basis that, say, a knows that b knows five secrets while c only knows two secrets. If a protocol P is symmetric, we can think of the protocol condition as the *unique* protocol condition for P, modulo permutation.

Epistemic and symmetric protocols capture the distributed peer-to-peer nature of the gossip problem.

Example 14. *The protocols ANY and LNS are symmetric and epistemic. For ANY this is trivial. For LNS, observe that agents always know which numbers and secrets they know. A direct consequence of clause (2.) of Definition 7 of the epistemic relation is that for any protocol P, if $(G, \sigma) \sim_a^P (G, \sigma')$, then $N_a^\sigma = N_a^{\sigma'}$ and $S_a^\sigma = S_a^{\sigma'}$. Thus, applying the clause for knowledge $K_a^P \varphi$ of Definition 8, we immediately get that the following formulas are all valid: $N_a b \to K_a^P N_a b$, $\neg N_a b \to K_a^P \neg N_a b$, $S_a b \to K_a^P S_a b$, and $\neg S_a b \to K_a^P \neg S_a b$. Therefore, in particular this holds for $P = LNS$.*

Although the numbers and secrets known by an agent before and after a call may vary, the agent always knows *whether* she knows a given number or secret. Knowledge about other agents having a certain number or a secret is preserved after calls. But, of course, knowledge about other agents *not* having a certain number or secret is not preserved after calls.

[2]Formally: $J(\top) := \top$, $J(N_a b) := N_a b$, $J(S_a b) := S_a b$, $J(\neg\varphi) := \neg J(\varphi)$, $J(\varphi \wedge \psi) := J(\varphi) \wedge J(\psi)$, $J(K_a^P \psi) := K_{J(a)}^{J(P)} J(\psi)$, $J(?\varphi) := ?J(\varphi)$, $J(ab) := J(a)J(b)$, $J(\pi; \pi') := J(\pi); J(\pi')$, $J(\pi \cup \pi') := J(\pi) \cup J(\pi')$, $J(\pi^*) := J(\pi)^*$.

Not all protocols we discuss in this work are definable in the logical language. We therefore need the additional notion of a *semantic protocol*, defined by its extension.

Definition 15 (Semantic protocol)**.** *A* semantic protocol *is a function* $P\colon \mathcal{G} \to \mathcal{P}((A \times A)^*)$ *mapping initial gossip graphs to sets of call sequences. We assume semantic protocols to be closed under subsequences, i.e. for all* G *we want that* $\sigma; ab \in P(G)$ *implies* $\sigma \in P(G)$*. For a* semantic protocol P *we say that a call* ab *is* P-permitted at (G, σ) *iff* $(\sigma; ab) \in P(G)$*.*

Given any syntactic protocol we can view its extension as a semantic protocol. Using this definition of permitted calls for semantic protocols we can apply Definition 7 to get the epistemic relation with respect to a semantic protocol P. Because the relation $\sim_a^P$ depends only on which calls are allowed, the epistemic relation with respect to a (syntactic) protocol P is identical to the epistemic relation with respect to the extension of P.

We also require that semantic protocols are symmetric and epistemic, adapting the definitions of these two properties as follows.

Definition 16 (Symmetric and epistemic semantic protocol)**.** *A semantic protocol* P *is* symmetric *iff for all initial gossip graphs* G *and for all permutations* J *of agents we have* $P(J(G)) = J(P(G))$ *(where* $J(P(G)) := \{J(\sigma) \mid \sigma \in P(G)\}$*). A semantic protocol* P *is* epistemic *iff for all initial gossip graphs* G *and for all* $\sigma \in P(G)$ *we have:* $(\sigma; ab) \in P(G)$ *iff for all* $\tau \sim_a^P \sigma$ *we have* $(\tau; ab) \in P(G)$*.*

It is easy to verify that the syntactic definition of an epistemic protocol agrees with the semantic definition.

Proposition 17. *A syntactic protocol* P *is epistemic if and only if its extension is epistemic.*

Proof. Let Q be the extension of P and note that, as remarked above, the epistemic relations induced by P and Q are identical. Now we have the following chain of equivalences:

$$
\begin{array}{ll}
 & P \text{ is not epistemic} \\
\Leftrightarrow & \exists a, b, G, \sigma : G, \sigma \not\models P_{ab} \to K_a^P P_{ab} \\
\Leftrightarrow & \exists a, b, G, \sigma, \tau : G, \sigma \models P_{ab},\ G, \tau \not\models P_{ab} \text{ and } (G, \sigma) \sim_a^P (G, \tau) \\
\Leftrightarrow & \exists a, b, G, \sigma, \tau : (\sigma; ab) \in Q(G),\ (\tau; ab) \notin Q(G) \text{ and } (G, \sigma) \sim_a^P (G, \tau) \\
\Leftrightarrow & \exists a, b, G, \sigma, \tau : (\sigma; ab) \in Q(G),\ (\tau; ab) \notin Q(G) \text{ and } (G, \sigma) \sim_a^Q (G, \tau) \\
\Leftrightarrow & Q \text{ is not epistemic}
\end{array}
$$

$\square$

Note that Proposition 17 does not imply that every epistemic semantic protocol is the extension of a syntactic epistemic protocol, since some semantic protocols are not the extension of any syntactic protocol.

For symmetry, the situation is slightly more complex than for being epistemic.

Proposition 18. *If a syntactic protocol P is symmetric, then its extension is symmetric.*

Proof. Let Q be the extension of P. Fix any permutation J and any initial gossip graph G. To show is that $Q(J(G)) = J(Q(G))$ (where J is extended to gossip graphs in the natural way). We show by induction that for every call sequence σ, we have $\sigma \in Q(J(G)) \Leftrightarrow \sigma \in J(Q(G))$.

As base case, note that $\epsilon \in Q(J(G))$ and $\epsilon \in J(Q(G))$. Now, as induction hypothesis, assume that for every call sequence τ that is shorter than σ, we have $\tau \in Q(J(G)) \Leftrightarrow \tau \in J(Q(G))$. Let ab be the final call in σ, so $\sigma = (\tau; ab)$. Then we have the following sequence of equivalences:

$$\begin{aligned}(\tau; ab) \in Q(J(G)) &\Leftrightarrow J(G), \tau \models P_{ab} \\ &\Leftrightarrow G, J^{-1}(\tau) \models J^{-1}(P_{ab}) \\ &\Leftrightarrow G, J^{-1}(\tau) \models P_{J^{-1}(ab)} \\ &\Leftrightarrow (J^{-1}(\tau); J^{-1}(ab)) \in Q(G) \\ &\Leftrightarrow (\tau; ab) \in J(Q(G)),\end{aligned}$$

where the equivalence on the third line is due to P being symmetric. This completes the induction step and thereby the proof. □

The converse of Proposition 18 does not hold: if P is not symmetric, it is still possible for its extension to be symmetric. The reason for this discrepancy is that symmetry for syntactic protocols has the very strong condition that $J(P_{ab}) = P_{J(ab)}$. So if P is symmetric and P' is given by (i) $P'_{cd} = P_{cd} \wedge \top$ and (ii) $P'_{ab} = P_{ab}$ for $a, b \neq c, d$, then P' is not symmetric even though P and P' have the same extension. We do, however, have the following slightly weaker statement. Recall that a gossip state (G, σ) is P-reachable iff the call sequence σ is P-permitted at G.

Proposition 19. *Let P be a syntactic protocol such that, for some P-reachable gossip state (G, σ), some permutation J and some a, b we have $G, \sigma \not\models P_{J(ab)} \leftrightarrow J(P_{ab})$. Then the extension of P is not symmetric.*

Proof. Let Q be the extension of P, and suppose towards a contradiction that Q is symmetric. Then we have the following sequence of equivalences:

$$\begin{aligned} G, \sigma \models P_{J(ab)} &\Leftrightarrow (\sigma; J(ab)) \in Q(G) \\ &\Leftrightarrow (J^{-1}(\sigma); ab) \in J^{-1}(Q(G)) \\ &\Leftrightarrow (J^{-1}(\sigma); ab) \in Q(J^{-1}(G)) \\ &\Leftrightarrow J^{-1}(G), J^{-1}(\sigma) \models P_{ab} \\ &\Leftrightarrow G, \sigma \models J(P_{ab}), \end{aligned}$$

where the equivalence on the third line is due to Q being symmetric. This contradicts $G, \sigma \not\models P_{J(ab)} \leftrightarrow J(P_{ab})$, from which it follows that Q is not symmetric. □

So while P may be non-symmetric and still have a symmetric extension, this can only happen if $J(P_{ab})$ is equivalent to $P_{J(ab)}$ in all reachable gossip states. We conclude that our syntactic and semantic definitions of symmetry agree up to logical equivalence.

3 Strengthening of Protocols

3.1 How can we strengthen a protocol?

In our semantics it is common knowledge among the agents that they follow a certain protocol, for example *LNS*. Can they use this information to prevent making "bad" calls that lead to an unsuccessful sequence?

If we look at the execution graph given in Figure 1, then it seems easy to fix the protocol. Agents b and c should wait and not make the first call. Agent b should not make a call before he has received a call from a. We cannot say this in our logic as we have no converse modalities to reason over past calls. In this case however, there is a different way to ensure the same result. We can ensure that b and c wait before calling by a strengthening of *LNS* that only allows a first call from i to j if j does not know the number of i. To determine that a call is not the first call, we need another property: after at least one call happened, there is an agent who knows another agent's secret.

We can define this new protocol by protocol condition $P_{ij} := LNS_{ij} \wedge (\neg N_j i \vee \bigvee_{k \neq l} S_k l)$. Observe that this new protocol is again symmetric and epistemic: agents always know whether $(\neg N_j i \vee \bigvee_{k \neq l} S_k l)$. Because of synchronicity, not only the callers but also all other agents know that there are agents k and l such that k knows the secret of l. This is an ad-hoc solution specific to this initial gossip graph. Could we also give a general definition to improve *LNS* which works on more or even all initial graphs? The answer to that is: more, yes, but all, no.

We will now discuss different ways to improve protocols by making them more restrictive. Our goal is to rule out unsuccessful sequences while keeping at least some successful ones. Doing this can be difficult because we still require the strengthened protocols to be epistemic and symmetric. Hence we are not allowed to arbitrarily rule out specific calls using the names of agents, for example. Whenever a call is removed from the protocol, we also have to remove all calls to other agents that the caller cannot distinguish: it has to be done *uniformly.* But before we discuss specific ideas for strengthening, let us define it.

Definition 20 (Strengthening). *A protocol P' is a* syntactic strengthening *of a protocol P iff $P'_{ab} \to P_{ab}$ is valid for all agents $a \neq b$. A protocol P' is a* semantic strengthening *of a protocol P iff $P' \subseteq P$.*

A syntactic strengthening procedure *is a function $\heartsuit$ that for any syntactic protocol P returns a syntactic strengthening $P^{\heartsuit}$ of P. Analogously, we define* semantic strengthening procedure*.*

We stress that strengthening is a relation between two protocols P and P' whereas strengthening procedures define a restricting transformation that given any P tells us how to obtain P'. In the case of a syntactic strengthening, P and P' are implicitly required to be syntactic protocols. Vice versa however, syntactic protocols can be semantic strengthenings. In fact, we have the following.

Proposition 21. *Every syntactic strengthening is a semantic strengthening.*

Proof. Let P' be a syntactic strengthening of a protocol P. Let a gossip graph G be given. We show by induction on the length of σ that $\sigma \in P'(G)$ implies $\sigma \in P(G)$. The base case where $\sigma = \epsilon$ is trivial.

For the induction step, consider any $\sigma = \tau; ab$. As $\tau; ab \in P'(G)$, we also have $\tau \in P'(G)$ and $G, \tau \models P'_{ab}$. From $\tau \in P'(G)$ and the inductive hypothesis, it follows that $\tau \in P(G)$. From $G, \tau \models P'_{ab}$ and the validity of $P'_{ab} \to P_{ab}$ follows $G, \tau \models P_{ab}$. Finally, by Definition 10, $\tau \in P(G)$ and $G, \tau \models P_{ab}$ imply $\tau; ab \in P(G)$. $\square$

Lemma 22. *Suppose P is a strengthening of Q. Then $K^Q_a \varphi \to K^P_a \varphi$ and $\hat{K}^P_a \varphi \to \hat{K}^Q_a \varphi$ are both valid, for any agent a.*

Proof. This follows immediately from the semantics of protocol-dependent knowledge given in Definition 8. $\square$

3.2 Syntactic Strengthening: Look-Ahead and One-Step

We will now present concrete examples of syntactic strengthening procedures.

Definition 23 (Look-Ahead and One-Step Strengthenings)**.** *We define four syntactic strengthening procedures as follows. Let P be a protocol.*

$$\begin{array}{llcl} \textit{hard look-ahead strengthening}: & P^{\blacksquare}_{ab} & := & P_{ab} \land K^P_a [ab]\langle P\rangle Ex \\ \textit{soft look-ahead strengthening}: & P^{\blacklozenge}_{ab} & := & P_{ab} \land \hat{K}^P_a [ab]\langle P\rangle Ex \\ \textit{hard one-step strengthening}: & P^{\Box}_{ab} & := & P_{ab} \land K^P_a [ab](Ex \lor \bigvee_{i,j}(Nij \land P_{ij})) \\ \textit{soft one-step strengthening}: & P^{\Diamond}_{ab} & := & P_{ab} \land \hat{K}^P_a [ab](Ex \lor \bigvee_{i,j}(Nij \land P_{ij})) \end{array}$$

The *hard* look-ahead strengthening allows agents to make a call iff the call is allowed by the original protocol and moreover they *know* that making this call yields a situation where the original protocol can still succeed.

For example, consider $LNS^{\blacksquare}$. Informally, its condition is that a is permitted to call b iff a does not have the secret of b and a knows that after making the call to b, it is still possible to follow *LNS* in such a way that all agents become experts.

The *soft* look-ahead strengthening allows more calls than the hard look-ahead strengthening because it only demands that *a considers it possible* that the protocol can succeed after the call. This can be interpreted as a good faith or lucky draw assumption that the previous calls between other agents have been made "in a good way". Soft look-ahead strengthening allows agents to take a risk.

The soft and the hard look-ahead strengthening include a diamond $\langle P\rangle$ labeled with the protocol P, where that protocol P by definition contains arbitrary iteration: the Kleene star $*$. To evaluate this, we need to compute the execution tree of P for the initial gossip graph G. In practice this can make it hard to check the protocol condition of the new protocol.

The *one-step* strengthenings, in contrast, only use the protocol condition P_{ij} in their formalization and not the entire protocol P. This means that they provide an easier to compute, but less reliable alternative to full look-ahead, namely by looking only one step ahead. We only demand that agent a knows (or, in the soft version, considers it possible) that after the call, everyone is an expert or the protocol can still go on for at least one more step — though it might be that all continuation sequences will eventually be unsuccessful and thus this next call would already have been excluded by both look-ahead strengthenings.

An obvious question now is, can these or other strengthenings get us from weak to strong success? Do these strengthenings only remove unsuccessful sequences, or will they also remove successful branches, and maybe even return an empty and unsuccessful protocol? In our next example everything still works fine.

Example 24. *Consider Example 12 again. It is easy to see that the soft and the hard look-ahead strengthening rule out the two unsuccessful branches in this execution tree and keep the successful ones. Protocol* $LNS^{\blacksquare}$ *only preserves alternatives that are all successful and* $LNS^{\blacklozenge}$ *only eliminates alternatives if they are all unsuccessful. In the execution tree in Figure 1, the effect is the same for* $LNS^{\blacksquare}$ *and* $LNS^{\blacklozenge}$*, because at any state the agents always know which calls lead to successful branches. This is typical for gossip scenarios with three agents: if a call happened, the agent not involved in the call might be unsure about the direction of the call, but it knows who the callers are.*

The one-step strengthenings are not enough to rule out the unsuccessful sequences. This is because the unsuccessful sequences are of length 2 *but the one-step strengthenings can only remove the last call in a sequence. In this case, the protocols* $LNS^{\Box}$ *and* $LNS^{\Diamond}$ *rule out the call* ab *after* bc *or* cb *happened.*

3.3 Semantic Strengthening: Uniform Backward Defoliation

We now present two semantic strengthening procedures. They are inspired by the notion of backward induction, a well-known solution concept in decision theory and game theory [32]. We will discuss this at greater length when defining the arbitrary iteration of these semantic strengthenings and in Section 5.

In backward induction, given a game tree or search tree, a parent node is called *bad* if all its children are loosing or bad nodes. Similarly, in trees with information sets of indistinguishable nodes, a parent node can be called bad if all its children are bad *and if also all children from indistinguishable nodes are bad.* Similar notions were considered in [7, 35]. Again, we have a soft and a hard version. We define *uniform backward defoliation* on the execution trees of dynamic gossip as follows to obtain two semantic strengthenings. We choose the name "defoliation" here because a single application of this strengthening procedure only removes leaves and not whole branches of the execution tree. The iterated versions we present later are then called *uniform backward induction.*

Definition 25 (Uniform Backward Defoliation)**.** *Suppose we have a protocol* P *and an initial gossip graph* G*. We define the* Hard Uniform Backward Defoliation (HUBD) *and* Soft Uniform Backward Defoliation (SUBD) *of* P *as follows.*

$$\begin{array}{rcll} P^{\mathsf{HUBD}}(G) & := & \{\sigma \in P(G) \mid & \sigma = \epsilon, \textit{ or } \sigma = \tau; ab \textit{ and } \forall (G, \tau') \sim_a^P (G, \tau) \\ & & & \textit{such that } \tau' \in \overline{P(G)} \textit{ implies } (G, \tau'; ab) \models Ex\} \\ P^{\mathsf{SUBD}}(G) & := & \{\sigma \in P(G) \mid & \sigma = \epsilon, \textit{ or } \sigma = \tau; ab \textit{ and } \exists (G, \tau') \sim_a^P (G, \tau) \\ & & & \textit{such that } \tau' \in \overline{P(G)} \textit{ implies } (G, \tau'; ab) \models Ex\} \end{array}$$

In this definition, $\forall (G, \tau') \sim_a^P (G, \tau)$ implicitly stands for "for all $\tau' \in P(G)$ such that $(G, \tau') \sim_a^P (G, \tau)$", because for (G, τ') to be in $\sim_a^P$ relation to another gossip state, τ' must be P-permitted; similarly for the existential quantification.

The HUBD strengthening keeps the calls which *must* lead to a non-terminal state or a state where everyone is an expert and SUBD keeps the calls which *might* do so. Equivalently, we can say that HUBD removes calls which may go wrong and SUBD removes those calls which will go wrong — where going wrong means leading to a terminal node where not everyone is an expert.

We can now prove that for any gossip protocol *Hard Uniform Backward Defoliation* is the same as *Hard One-Step Strengthening*, in the sense that their extensions are the same on any gossip graph, and that *Soft Uniform Backward Defoliation* is the same as *Soft One-Step Strengthening.*

Theorem 26. $P^{\Box} = P^{\mathsf{HUBD}}$ *and* $P^{\Diamond} = P^{\mathsf{SUBD}}$

Proof. Note that ϵ is an element of both sides of both equations. For any non-empty sequence we have the following chain of equivalences for the hard versions of UBD and one-step strengthening:

$(\sigma; ab) \in P^{\Box}(G)$

$\Updownarrow$ by Definition 10

$G, \sigma \models P_{ab}^{\Box}$

$\Updownarrow$ by Definition 23

$G, \sigma \models P_{ab} \wedge K_a^P[ab]\left(\bigvee_{i,j}(N_i j \wedge P_{ij}) \vee Ex\right)$

$\Updownarrow$ by Definition 8

$(\sigma; ab) \in P(G)$ and $(G, \sigma) \vDash K_a^P[ab]\left(\bigvee_{i,j}(N_i j \wedge P_{ij}) \vee Ex\right)$

$\Updownarrow$ by Definition 8

$(\sigma; ab) \in P(G)$ and $\forall (G, \sigma') \sim_a^P (G, \sigma) : (G, \sigma'; ab) \models \bigvee_{i,j}(N_i j \wedge P_{ij}) \vee Ex$

$\Updownarrow$ by Definition 11

$(\sigma; ab) \in P(G)$ and $\forall (G, \sigma') \sim_a^P (G, \sigma) : \sigma'; ab \notin \overline{P(G)}$ or $(G, \sigma'; ab) \models Ex$

$\Updownarrow$ by Definition 25

$(\sigma; ab) \in P^{\mathsf{HUBD}}(G)$

And we have a similar chain of equivalences for the soft versions:

$(\sigma; ab) \in P^{\Diamond}(G)$

$\Updownarrow$ by Definition 10

$G, \sigma \models P^{\Diamond}_{ab}$

$\Updownarrow$ by Definition 23

$G, \sigma \models P_{ab} \wedge \hat{K}^P_a[ab]\left(\bigvee_{i,j}(N_i j \wedge P_{ij}) \vee Ex\right)$

$\Updownarrow$ by Definition 8

$(\sigma; ab) \in P(G)$ and $(G, \sigma) \models \hat{K}^P_a[ab]\left(\bigvee_{i,j}(N_i j \wedge P_{ij}) \vee Ex\right)$

$\Updownarrow$ by Definition 8

$(\sigma; ab) \in P(G)$ and $\exists (G, \sigma') \sim^P_a (G, \sigma) : (G, \sigma'; ab) \models \bigvee_{i,j}(N_i j \wedge P_{ij}) \vee Ex$

$\Updownarrow$ by Definition 11

$(\sigma; ab) \in P(G)$ and $\exists (G, \sigma') \sim^P_a (G, \sigma) : \sigma'; ab \notin \overline{P(G)}$ or $(G, \sigma'; ab) \models Ex$

$\Updownarrow$ by Definition 25

$(\sigma; ab) \in P^{\mathsf{SUBD}}(G)$

□

Similarly to backward induction in perfect information games [4], uniform backward defoliation is *rational*, in the sense that it forces an agent to avoid calls leading to unsuccessful sequences. The strengthening SUBD avoids a call if it always leads to an unsuccessful sequence. The strengthening HUBD avoids a call if it sometimes leads to a unsuccessful sequence.

3.4 Iterated Strengthenings

The syntactic strengthenings we looked at are all defined in terms of the original protocol. In $P^{\blacksquare}_{ab} := P_{ab} \wedge K^P_a[ab]\langle P \rangle Ex$ the given P occurs in three places. Firstly, in the protocol condition P_{ab} requiring that the call is permitted according to the old protocol P — this ensures that the new protocol is a strengthening of the original P. Secondly, as a parameter to the knowledge operator, in K^P_a, which means that agent a knows that everyone followed P (and that this is common knowledge). Thirdly, in the part $\langle P \rangle$ assuming that after the considered call everyone will continue to follow protocol P in the future.

Hence we have strengthened the protocol that the agents use and thereby changed their behavior, but not their assumptions about what protocol other agents follow. For example, when $P = LNS$, all agents now act according to $LNS^{\blacksquare}$, on the assumption that all other agents act according to LNS. This does not mean that agents cannot determine what they know if $LNS^{\blacksquare}$ were common knowledge: each agent a can check that knowledge using $K_a^{LNS^{\blacksquare}}\varphi$. But this $K_a^{LNS^{\blacksquare}}$ modality is not part of the protocol $LNS^{\blacksquare}$. The agents do not use this knowledge to determine whether to make calls.

But why should our agents stop their reasoning here? It is natural to iterate strengthening procedures and determine whether we can further improve our protocols by also updating the knowledge of the agents.

For example, consider repeated hard one-step strengthening:

$$(P^{\Box})^{\Box}_{ab} = P^{\Box}_{ab} \wedge \hat{K}_a^{P^{\Box}}[ab](Ex \vee \bigvee_{i,j}(N_i j \wedge P^{\Box}_{ij}))$$

In this section we investigate iterations and combinations of strengthening procedures. In particular we investigate various combinations of hard and soft one-step and look-ahead strengthening, in order to determine how they relate to each other.

Definition 27 (Strengthening Iteration)**.** *Let P be a syntactic protocol. For any of the four syntactic strengthening procedures $\heartsuit \in \{\blacksquare, \blacklozenge, \Box, \Diamond\}$, we define its iteration by adjusting the protocol condition as follows, which implies $P^{\heartsuit 1} = P^{\heartsuit}$:*

$$\begin{array}{lcl} P^{\heartsuit 0}_{ab} & := & P_{ab} \\ P^{\heartsuit (k+1)}_{ab} & := & (P^{\heartsuit k})^{\heartsuit}_{ab} \end{array}$$

Let now P be a semantic protocol, and let $\heartsuit \in \{\mathsf{HUBD}, \mathsf{SUBD}\}$. We define their iteration, for all gossip graphs G, by:

$$\begin{array}{lcl} P^{\heartsuit 0}(G) & := & P(G) \\ P^{\heartsuit (k+1)}(G) & := & (P^{\heartsuit k})^{\heartsuit}(G) \end{array}$$

It is easy to check that Theorem 26 generalizes to the iterated strengthenings as follows.

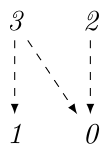
Corollary 28. *For any $k \in \mathbb{N}$, we have:*

$$P^{\Box k} = P^{\mathsf{HUBD}k} \text{ and } P^{\Diamond k} = P^{\mathsf{SUBD}k}$$

Proof. By induction using Theorem 26. $\Box$

Example 29. *We reconsider Examples 12 and 24, and we recall that $LNS^{\Box}$ and $LNS^{\Diamond}$ rule out the call ab after bc or cb happened. To eliminate bc and cb as the first call, we have to iterate one-step strengthening: $(LNS^{\Box})^{\Box}$ is strongly successful on this graph, as well as $(LNS^{\Diamond})^{\Diamond}$, $(LNS^{\Box})^{\Diamond}$ and $(LNS^{\Diamond})^{\Box}$.*

Example 30. *We consider the "N"-shaped gossip graph shown below. There are 21 LNS sequences for this graph, of which* 4 *are successful (✓) and 17 are unsuccessful ($\times$).*



$20; 30; 01; 31$	$\times$	$30; 20; 01; 31; 21$	✓	$30; 31; 20; 21; 01$	$\times$
$20; 30; 31; 01$	$\times$	$30; 20; 21; 01; 31$	✓	$31; 10; 20; 30$	$\times$
$20; 31; 10; 30$	$\times$	$30; 20; 21; 31; 01$	✓	$31; 10; 30; 20$	$\times$
$20; 31; 30; 10$	$\times$	$30; 20; 31; 01; 21$	$\times$	$31; 20; 10; 30$	$\times$
$30; 01; 20; 31$	$\times$	$30; 20; 31; 21; 01$	$\times$	$31; 20; 30; 10$	$\times$
$30; 01; 31; 20$	$\times$	$30; 31; 01; 20$	$\times$	$31; 30; 10; 20$	$\times$
$30; 20; 01; 21; 31$	✓	$30; 31; 20; 01; 21$	$\times$	$31; 30; 20; 10$	$\times$

We can show the call sequences in a more compact way if we only distinguish call sequences up to the moment when it is decided whether LNS will succeed. Formally, consider the set of minimal $\sigma \in LNS(G)$ such that for all two terminal LNS-sequences $\tau, \tau' \in \overline{LNS(G)}$ extending σ, we have $G, \tau \models Ex$ iff $G, \tau' \models Ex$. We will use this shortening convention throughout the paper.

20	$\times$
$30; 01$	$\times$
$30; 20; 01$	✓
$30; 20; 21$	✓
$30; 20; 31$	$\times$
$30; 31$	$\times$
31	$\times$

It is pretty obvious what the agents should do here: Agent 2 should not make the first call but let 3 *call* 0 *first. The soft look-ahead strengthening works well on this graph: It disallows all unsuccessful sequences and keeps all successful ones. For example, after call* 30*, agent* 2 *considers it possible that call* 30 *happened and in this case the call* 20 *can lead to success. Hence the protocol condition of $LNS^{\blacklozenge}$ is fulfilled. The strengthening $LNS^{\blacklozenge}$ is strongly successful on this graph.*

But note that 2 *does not* know *that* 20 *can lead to success, because the first call could have been* 31 *as well and for agent* 2 *this would be indistinguishable from* 30*. Therefore the hard look-ahead strengthening is too restrictive here. In fact, the only* *call which $LNS^{\blacksquare}$ still allows is* 30 *at the beginning. After that no more calls are allowed by the hard look-ahead strengthening.*

A full list showing which call sequences are allowed by which strengthenings of LNS for this example is provided in Table 2. "Full" means that we continue iterating the strengthening until $P^{\heartsuit k}(G) = P^{\heartsuit (k+1)}(G)$ for the given graph G. Such fixpoints of protocol strengthening will be formally introduced in the next section.

The hard look-ahead strengthening restricts the set of allowed calls based on a full analysis of the whole execution tree. One might thus expect, that applying hard look-ahead more than once would not make a difference. However, we have the following negative results on iterating hard look-ahead strengthening and the combination of hard look-ahead and hard one-step strengthening.

Fact 31. *Hard look-ahead strengthening is not idempotent and does not always yield a fixpoint of hard one-step strengthening:*

(i) There exist a graph G and a protocol P for which $P^{\blacksquare}(G) \neq (P^{\blacksquare})^{\blacksquare}(G)$.

(ii) There exist a graph G and a protocol P for which $(P^{\blacksquare})^{\square}(G) \neq P^{\blacksquare}(G)$.

Proof.

(i) Let G be the "N" graph from Example 30 and consider the protocol $P = LNS$. Applying hard look-ahead strengthening once only allows the first call 30 and nothing after that call. If we now apply hard look-ahead strengthening again we get the empty set: $P^{\blacksquare}(G) \neq (P^{\blacksquare})^{\blacksquare}(G) = \varnothing$. See also Table 2.

(ii) The "diamond" graph that we will present in Section 3.6 can serve as an example here. We can show that the inequality holds for this graph by exhaustive search, using our Haskell implementation described in the Appendix. Plain *LNS* has 48 successful and 44 unsuccessful sequences on this graph. Of these, $LNS^{\blacksquare}$ still includes 8 successful and 8 unsuccessful sequences. If we now apply hard one-step strengthening, we get $(LNS^{\blacksquare})^{\square}$ where 4 of the unsuccessful sequences are removed. See also Table 3 in the Appendix. We note that for $P = LNS$ there is no smaller graph to show the inequality. This can be checked by manual reasoning or with our implementation. □

Similarly, we can ask whether the soft strengthenings are related to each other, analogous to Fact 31. We do not know whether there is a protocol P for which $(P^{\blacklozenge})^{\lozenge} \neq P^{\blacklozenge}$ and leave this as an open question.

Another interesting property that strengthenings can have is *monotonicity*. Intuitively, a strengthening is monotone iff it preserves the inclusion relation between extensions of protocols. This property is useful to study the fixpoint behavior of strengthenings. We will now define monotonicity formally and then obtain some results for it.

Definition 32. *A strengthening* $\heartsuit$ *is called* monotone *iff for all protocols* Q *and* P *such that* $Q \subseteq P$*, we also have* $Q^{\heartsuit} \subseteq P^{\heartsuit}$*.*

Proposition 33 (Soft one-step strengthening is monotone)**.** *Let* P *be a protocol and* Q *be an arbitrary strengthening of* P*, i.e.* $Q \subseteq P$*. Then we also have* $Q^{\Diamond} \subseteq P^{\Diamond}$*.*

Proof. As Q is a strengthening of P, the formula $Q_{ab} \to P_{ab}$ is valid. We want to show that $Q^{\Diamond}_{ab} \to P^{\Diamond}_{ab}$. Suppose that $G, \sigma \models Q^{\Diamond}_{ab}$, i.e.:

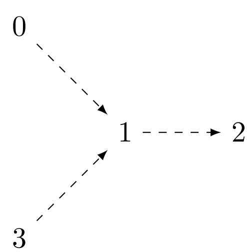
$$G, \sigma \models Q_{ab} \text{ and } G, \sigma \models \hat{K}^{Q}_{a}[ab](Ex \vee \bigvee_{i,j}(N_i j \wedge Q_{ij}))$$

From the first part and the validity of $Q_{ab} \to P_{ab}$, we get $G, \sigma \models P_{ab}$. The second part and the validity of $Q_{ij} \to P_{ij}$ give us $G, \sigma \models \hat{K}^{Q}_{a}[ab](Ex \vee \bigvee_{i,j}(N_i j \wedge P_{ij}))$. From that and Lemma 22 it follows that $G, \sigma \models \hat{K}^{P}_{a}[ab](Ex \vee \bigvee_{i,j}(N_i j \wedge P_{ij}))$. Combining these, it follows by definition of soft one-step strengthening that we have $G, \sigma \models P^{\Diamond}_{ab}$. $\square$

Proposition 34 (Both hard strengthenings are not monotone)**.** *Let* P *and* Q *be protocols. If* $Q \subseteq P$*, then (i)* $Q^{\blacksquare} \subseteq P^{\blacksquare}$ *may not hold, and also (ii)* $Q^{\square} \subseteq P^{\square}$ *may not hold.*

Proof. (i) *Hard one-step strengthening is not monotone*:

Consider the "spaceship" graph below with four agents 0, 1, 2 and 3 where 0 and 3 know 1's number, 1 knows 2's number, and 2 knows no numbers.



On this graph the *LNS* sequences up to decision point are:

$01; 02$	$\times$	$01; 31; 12$	$\checkmark$	$31; 01; 02$	$\checkmark$	$31; 12$	$\times$
$01; 12$	$\times$	$01; 31; 32$	$\checkmark$	$31; 01; 12$	$\checkmark$	$31; 32$	$\times$
$01; 31; 02$	$\times$	12	$\times$	$31; 01; 32$	$\times$		

Note that

$$LNS^{\blacklozenge}(G) = \left\{ \begin{array}{l} (01;31;12;02;32),(01;31;12;32;02),(01;31;32;02;12), \\ (01;31;32;12;02),(31;01;02;12;32),(31;01;02;32;12), \\ (31;01;12;02;32),(31;01;12;32;02) \end{array} \right\}$$

is strongly successful and therefore hard one-step strengthening does not change it — we have $(LNS^{\blacklozenge})^{\Box}(G) = LNS^{\blacklozenge}(G)$. On the other hand, consider

$$LNS^{\Box}(G) = \left\{ \begin{array}{l} (01;02;12),(01;12;02),(01;31;02;12),(01;31;02;32), \\ (01;31;12;32;02),(01;31;32;12;02),(12;01),(12;31), \\ (31;01;02;12;32),(31;01;12;02;32),(31;01;32;02), \\ (31;01;32;12),(31;12;32),(31;32;12) \end{array} \right\}$$

and note that this is not a superset of $(LNS^{\blacklozenge})^{\Box}(G) = LNS^{\blacklozenge}(G)$, because we have $(01;31;12;02;32) \in (LNS^{\blacklozenge})^{\Box}(G) = LNS^{\blacklozenge}(G)$ but $(01;31;12;02;32) \notin LNS^{\Box}(G)$.

Together, we have $LNS^{\blacklozenge}(G) \subseteq LNS(G)$ but $(LNS^{\blacklozenge})^{\Box}(G) \not\subseteq LNS^{\Box}(G)$.

Hence $Q = LNS^{\blacklozenge} \subseteq LNS = P$ is a counterexample and $\Box$ is not monotone.

(ii) *Hard look-ahead strengthening is not monotone*:

For hard look-ahead strengthening we can use the same example. Because $LNS^{\blacklozenge}$ is strongly successful, hard look-ahead strengthening does not change it: $(LNS^{\blacklozenge})^{\blacksquare}(G) = LNS^{\blacklozenge}(G)$.

Moreover, $LNS^{\blacksquare}(G) = \{(01),(31)\}$ is not a superset of $(LNS^{\blacklozenge})^{\blacksquare}(G) = LNS^{\blacklozenge}(G)$.

Together we have $LNS^{\blacklozenge}(G) \subseteq LNS(G)$ but $(LNS^{\blacklozenge})^{\blacksquare}(G) \not\subseteq LNS^{\blacksquare}(G)$, hence hard look-ahead strengthening is not monotone either. $\Box$

This result is relevant for our pursuit to pin down how rational agents can employ common knowledge of a protocol to improve upon it. It shows that hard look-ahead strengthening is not rational, as follows.

We consider again the "spaceship" graph in the proof of Proposition 34. Let us define a *bad call* as a call after which no successful continuation is possible. Correspondingly, a *good call* is one after which success is still possible. The initial call could be 12, but that is a bad call. All successful *LNS* sequences on this graph start with $01;31$ or $31;01$.

Let us place ourselves in the position of agent 3 after the call 01 has been made. As far as 3 can tell (if the only background common knowledge is that everyone follows *LNS*), the first call may have been 12, at which point no agent can make a good call because no continuation is successful. In particular, the second call 31 is then bad. So 3 will not call 1, because it is possible that the call 31 is bad, and we are following hard look-ahead.

Symmetrically, the same reasoning is made by agent 0: even if the first call is 31, it could also have been 12, after which any continuation is unsuccessful, and therefore 0 will not call 1, which again seems irrational.

So nobody will make a call. The extension of $LNS^{\blacksquare}$ on this graph is empty.

But as all agents know that 12 is bad, agent 1 knows this in particular, and as agent 1 is rational herself, she would therefore not have made that call. And agents 3 and 0 can draw that conclusion too. It therefore seems after all irrational for 3 not to call 1, or for 0 not to call 1.

This shows that hard look-ahead strengthening is not rational. In particular, it ignores the rationality of other agents.

3.5 Limits and Fixpoints of Strengthenings

Given the iteration of strengthenings we discussed in the previous section, it is natural to consider limits and fixpoints of strengthening procedures. In this subsection we discuss them and give some small results. A detailed investigation is deferred to future research.

Note that the protocol conditions of all four basic syntactic strengthenings are conjunctions with the original protocol condition as a conjunct. Therefore, all these four strengthenings are *non-increasing*: For all $\heartsuit \in \{\blacksquare, \blacklozenge, \square, \lozenge\}$ and all protocols P, we have $P^{\heartsuit} \subseteq P$. The same holds, by definition, for semantic strengthenings. This implies that if, on any gossip graph, we start with a protocol that only allows finite call sequences, such as LNS, then applying strengthening repeatedly will eventually lead to a fixpoint. This fixpoint might be the empty set, or a non-empty set and thereby provide a new protocol.

For other protocols that allow infinite call sequences, such as ANY, we do not know if this procedure leads to a unique fixpoint and whether fixpoints are always reached. We therefore distinguish fixpoints from limits.

Definition 35 (Strengthening Limit; Fixpoint)**.** *Consider any strengthening $\heartsuit$. The* $\heartsuit$-limit *of a given protocol P is the semantic protocol $P^{\heartsuit *}$ defined as $\bigcap_{k \in \mathbb{N}} P^{\heartsuit k}$. A given protocol P is a* fixpoint *of a strengthening $\heartsuit$ iff $P = P^{\heartsuit}$.*

Note that limit protocols $P^{\heartsuit *}$ are *not* in the logical language, unlike their constituents $P^{\heartsuit k}$. We now define $P^{\square *}$ as *Hard Uniform Backward Induction*, and $P^{\lozenge *}$ as *Soft Uniform Backward Induction.* Again using induction on Theorem 26, it follows that Uniform Backward Induction is the same as arbitrarily often iterated Uniform Backward Defoliation.

Corollary 36.

$$P^{\Box *} = P^{\mathsf{HUBD}*} \text{ and } P^{\Diamond *} = P^{\mathsf{SUBD}*}.$$

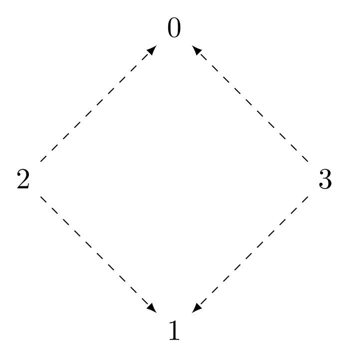
Example 37. *Consider $P = LNS$. The number of LNS calls between n agents is bounded by $\binom{n}{2} = n(n-1)/2$. The limit $LNS^{\heartsuit *}$ is therefore reached after a finite number of iterations, and expressible in the gossip protocol language: $LNS^{\heartsuit n(n-1)/2} = LNS^{\heartsuit *}$.*

As a further observation, the look-ahead strengthenings are not always the limits of one-step strengthenings. In other words, we do *not* have for all G that $P^{\Box *}(G) = P^{\blacksquare}(G)$ or that $P^{\Diamond *}(G) = P^{\blacklozenge}(G)$. Counterexamples are the "N" graph from Example 30 and the extension of various strengthenings relating to the example in the upcoming Section 3.6, as shown in Table 3 in the Appendix.

However, we know by the Knaster-Tarski theorem [37] that on any gossip graph soft one-step strengthening $\Diamond$ has a unique greatest fixpoint, because $\Diamond$ is monotone and the lattice we are working in is the powerset of the set of all call sequences and thereby complete.

3.6 Detailed Example: the Diamond Gossip Graph

Consider the initial "diamond" gossip graph below.



There are 92 different terminating sequences of *LNS* calls for this initial graph of which 48 are successful and 44 are unsuccessful. Also below we give an overview of all sequences. For brevity we only list them in the compact way, up to the call after which success has been decided.

$20;01$	$\times$	$21;10$	$\times$	$30;01$	$\times$	$31;10$	$\times$
$20;21$	$\times$	$21;20$	$\times$	$30;20;01$	$\checkmark$	$31;20$	$\checkmark$
$20;30;01$	$\checkmark$	$21;30$	$\checkmark$	$30;20;21$	$\checkmark$	$31;21;10$	$\checkmark$
$20;30;21$	$\times$	$21;31;10$	$\checkmark$	$30;20;31$	$\times$	$31;21;20$	$\checkmark$
$20;30;31$	$\checkmark$	$21;31;20$	$\times$	$30;21$	$\checkmark$	$31;21;30$	$\times$
$20;31$	$\checkmark$	$21;31;30$	$\checkmark$	$30;31$	$\times$	$31;30$	$\times$

Table 1 shows how many sequences are permitted by the different strengthenings. Both soft strengthenings rule out no successful sequences and rule out some unsuccessful sequences. The hard look-ahead strengthening removes some successful sequences and rules out the same number of unsuccessful sequences as the soft lookahead strengthening, but interestingly enough this is a different set.

This demonstrates that Table 1 may be misleading: the same number of sequences does not imply the same set of sequences. Table 3 in the Appendix is more detailed and lists sequences. If a further iteration of a strengthening does not change the number and also not the set of sequences, it has the same extension, and is therefore a fixpoint. For example, Table 3 shows that $LNS^{\Diamond 2}$ and $LNS^{\Diamond 3}$ both have 48 successful and 32 unsuccessful sequences on the diamond graph. They also have the same extension, hence $LNS^{\Diamond 2}$ is a fixpoint of $\Diamond$ on this graph.

Recall that one-step strengthening is uniform backward defoliation (Theorem 26) and that the limit of one-step strengthening is uniform backward induction (Corollary 36). Table 1 shows the difference between the look-ahead strengthenings and the one-step/defoliation strengthenings. Although on this "diamond" graph, the hard strengthenings $LNS^{\blacksquare k}$ and $LNS^{\Box k}$ have the same fixpoint, namely the empty extension for all $k \geq 4$, the soft strengthenings $LNS^{\blacklozenge k}$ and $LNS^{\Diamond k}$ have different fixpoints. Both are reached when $k = 2$.

We now present two strengthenings that are strongly successful on this graph (only successfully terminating call sequences remain).

Firstly, consider the protocol $(LNS^{\Diamond})^{\Box 3}$. Its extension is as follows, see also Tables 1 and 3.

$20;30;01;31;21$	$21;30;01;31;20$	$30;20;01;21;31$	$31;20;01;21;30$
$20;30;31;01;21$	$21;30;31;01;20$	$30;20;21;01;31$	$31;20;21;01;30$
$20;31;10;30;21$	$21;31;10;30;20$	$30;21;10;20;31$	$31;21;10;20;30$
$20;31;30;10;21$	$21;31;30;10;20$	$30;21;20;10;31$	$31;21;20;10;30$

Its extension has no sequences with only four calls. There are sequences with redundant second-to-last calls, for example 10 in $20;31;30;10;21$.

Protocol	# successful	# unsuccessful
LNS	48	44
$LNS^{\blacksquare}$	8	8
$LNS^{\blacksquare 2}$	0	4
$LNS^{\blacksquare 3}$	0	0
$LNS^{\blacklozenge}$	48	8
$LNS^{\blacklozenge 2}$	48	8
$LNS^{\blacklozenge 3}$	48	8
$LNS^{\Box}$	24	36
$LNS^{\Box 2}$	8	16
$LNS^{\Box 3}$	8	4
$LNS^{\Box 4}$	0	4
$LNS^{\Box 5}$	0	0
$LNS^{\Diamond}$	48	36
$LNS^{\Diamond 2}$	48	32
$LNS^{\Diamond 3}$	48	32
$(LNS^{\Diamond})^{\Box 3}$	16	0
$((LNS^{\Diamond})^{\Box})^{\blacksquare}$	16	0

Table 1: Statistics for the diamond example.

Secondly, we present a protocol that is strongly successful on this graph and that has no redundant calls. Its description is far more involved than the previous protocol, but the effort seems worthwhile as is shows that: (i) for some initial gossip graphs we can strengthen LNS up to finding strongly successful as well as optimal extensions; (ii) the hard and soft strengthening procedures described so far merely touch the surface and are not all that goes around, because one can easily show that the following protocol does not correspond to any of those or their iterations.

We first describe it as a semantic protocol, liberally referring to call histories in our description (which cannot be done in our logical language) and only then give a formalization using the syntax of our protocol logic. Consider the following semantic protocol:

(1) agent 2 or agent 3 makes a call to either 0 or 1.

(2) the agent among 2 and 3 that did not make a call in step (1) calls either 0 or 1.

(3) the agent x that made the call in step (2) now makes a second call; if

x called agent 1 before then x now calls 0 and vice versa.

(4) the agent y that made the call in step (1) now makes a second call; if y called agent 1 before then y now calls 0 and vice versa.

(5) if the agent z that was called in step (2) is not yet an expert, then z calls the last remaining agent whose secret z does not know.

Now let us explain why this protocol is strongly successful on the "diamond" graph, and why it is a strengthening of *LNS*. There are four possibilities for the first call: 2 may call 0, 2 may call 1, 3 may call 0 or 3 may call 1. These four cases are symmetrical, so let us assume that the first call is 20. The next call will then be made by agent 3, and there are two possibilities: either 3 also calls agent 0, or 3 calls agent 1. The call sequences, and the secrets known by the agents after each call has been made, are shown in the following two tables.

First case: 2 and 3 call the same agent

Stage	Call	0	1	2	3
(1)	20	$\{0,2\}$	$\{1\}$	$\{0,2\}$	$\{3\}$
(2)	30	$\{0,2,3\}$	$\{1\}$	$\{0,2\}$	$\{0,2,3\}$
(3)	31	$\{0,2,3\}$	$\{0,1,2,3\}$	$\{0,2\}$	$\{0,1,2,3\}$
(4)	21	$\{0,2,3\}$	$\{0,1,2,3\}$	$\{0,1,2,3\}$	$\{0,1,2,3\}$
(5)	01	$\{0,1,2,3\}$	$\{0,1,2,3\}$	$\{0,1,2,3\}$	$\{0,1,2,3\}$

Second case: 2 and 3 call different agents

Stage	Call	0	1	2	3
(1)	20	$\{0,2\}$	$\{1\}$	$\{0,2\}$	$\{3\}$
(2)	31	$\{0,2\}$	$\{1,3\}$	$\{0,2\}$	$\{1,3\}$
(3)	30	$\{0,1,2,3\}$	$\{1,3\}$	$\{0,2\}$	$\{0,1,2,3\}$
(4)	21	$\{0,1,2,3\}$	$\{0,1,2,3\}$	$\{0,1,2,3\}$	$\{0,1,2,3\}$

Note that all of these calls are possible, in the sense that all callers know the number of the agent they are calling. Agents 2 and 3 start out knowing the numbers of 0 and 1, so the calls 20, 30, 21 and 31 are possible from the start. Furthermore, agent 0 learns the number of agent 1 from agent 2 in the first call, so after the call 20 the call 01 is also possible.

In the second case there is no fifth call, since the agent that received the call in step (2) is already an expert after step (4). As a result, there are no redundant calls in either possible call sequence. Furthermore, in either case, all agents become experts. Finally, every call is to an agent whose secret is unknown to the caller before the call. So, the described protocol is a strongly successful strengthening of *LNS*.

The two call sequences shown above are possible if the first call is 20. There are six other call sequences corresponding to the other three options for the first call. Overall, the protocol allows the following 8 sequences.

$20; 30; 31; 21; 01$	$21; 31; 30; 20; 10$	$30; 20; 21; 31; 01$	$31; 21; 20; 30; 10$
$20; 31; 30; 21$	$21; 30; 31; 20$	$30; 21; 20; 31$	$31; 20; 21; 30$

We can also define a syntactic protocol that has the above semantic protocol as its extension. This syntactic protocol is not particularly elegant, but it illustrates how the logical language can be used to express more complex conditions. The call condition P_{ij} of this syntactic protocol is of the form $P_{ij} = K_i\psi_{ij}$ (where K_i abbreviates K_i^{ANY}, as defined in Section 2.2). This guarantees that the protocol is epistemic, because Lemma 22 implies that $K_i\psi_{ij} \to K_i^P K_i\psi_{ij}$ is valid. The formula ψ_{ij} is a disjunction with the following five disjuncts, one for each of the clauses (1) – (5) of the protocol as described above.

The formula $\varphi_0 := \bigwedge_k \bigwedge_{l\neq k} \neg S_k l$ holds if and only if no calls have taken place yet. Since agents 2 and 3 are the only ones that know the number of another agent, if φ_0 is true then any agent who can make a call is allowed to make that call. So φ_0 is the first disjunct of ψ_{ij}, enabling the call in stage (1).

Defining "exactly one call has been made" is a bit harder, but we can do it: after the first call, there will be two agents that know two secrets, while everyone else only knows one secret. So $\varphi_1 := \bigvee_{k\neq l}(S_k l \wedge S_l k \wedge \bigwedge_{m\notin\{k,l\}} \bigwedge_{n\neq m} \neg S_m n)$ holds if and only if exactly one call has been made. In that case, any agent that is capable of making calls and only knows their own secret is allowed to make a call, so $\varphi_1 \wedge \bigwedge_{k\neq i} \neg S_i k$ is the second disjunct of ψ_{ij}, enabling the call in stage (2).

In stage (3), the second caller is supposed to make another call. We make a case distinction based on whether the first two calls were to the same agent or to different agents. If they were to the same agent, then the second caller now knows three different secrets: $\bigvee_{k\neq i} \bigvee_{l\notin\{i,k\}} S_i kl$. But that holds not only for the agent who made the second call, but also for the agent that received the second call. The difference between them is that the secret of the receiver of this call is now known by three agents, while the secret of the caller is known by only two: $\bigwedge_{k\neq i}(S_k i \to \bigwedge_{l\notin\{i,k\}} \neg S_l i)$.

If the first two calls were to different agents, the second caller knows that every agent now knows exactly two secrets: $K_i \bigwedge_k \bigvee_{l\neq k}(S_k l \wedge \bigwedge_{m\notin\{k,l\}} \neg S_k m)$. This holds for the receiver of the second call as well, but the difference between them is that the number of the receiver is known to an agent who does not know their secret, while the number of the caller is not: $\bigwedge_k(N_k i \to S_k i)$.

In either case, the target of the call should be the unique agent whose number the caller knows but whose secret the caller does not know. Since calls are always to an agent whose number is known, we only have to stipulate that the target's secret is not known. So the third disjunct of ψ_{ij} is

$$\neg S_i j \wedge ((\bigvee_{k \neq i} \bigvee_{l \notin \{i,k\}} S_i kl \wedge \bigwedge_{k \neq i} (S_k i \to \bigwedge_{l \notin \{i,k\}} \neg S_l i)) \vee$$
$$(K_i \bigwedge_{k} \bigvee_{l \neq k} (S_k l \wedge \bigwedge_{m \notin \{k,l\}} \neg S_k m) \wedge \bigwedge_{k} (N_k i \to S_k i))),$$

enabling the call in stage (3).

It is relatively easy to express when the call in stage (4) should happen: before the third call, all agents know that there is no expert yet, while after the third call all agents consider it possible that there is at least one expert. This can be expressed as $\hat{K}_i \bigvee_k Ex_k$. It is slightly more difficult to identify the agent who should make the call. The agent who should make the call, the one who made the call in stage (1), is the only agent who only knows two secrets, and whose number is only known by agents that also know their secret. So $\neg \bigvee_{k \neq i} \bigvee_{l \notin \{i,j\}} S_i kl \wedge \bigwedge_k (N_k i \to S_k i)$. Finally, the person who should be called in this stage is the unique agent of whom the caller knows the number but not the secret. The fourth disjunct is therefore $\neg S_i j \wedge \hat{K}_i \bigvee_k Ex_k \wedge \neg \bigvee_{k \neq i} \bigvee_{l \notin \{i,j\}} S_i kl \wedge \bigwedge_k (N_k i \to S_k i)$.

Finally, the call in stage (5) should only happen if there remains a non-expert agent. This non-expert considers it possible that all other agents are experts, so the final disjunct of ψ_{ij} is $\neg S_i j \wedge \hat{K}_i \bigwedge_{k \neq i} Ex_k$.

On the "diamond" graph the extension of the syntactic protocol with call condition P_{ij} is the semantic protocol defined above. Clearly, this protocol is symmetric. We already showed that the protocol is epistemic as well.

All in all, this gives us the protocol that we were looking for. Manually verifying the extension of the protocol is somewhat tedious, so we have also checked the extension using the model checking tool described in the Appendix.

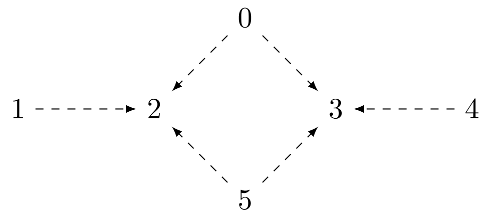
4 An Impossibility Result on Strengthening LNS

4.1 An Impossibility Result

In this section we will show that there are graphs where (i) *LNS* is weakly successful and (ii) no epistemic symmetric strengthening of *LNS* is strongly successful. Recall that we assume that the system is synchronous and that the initial gossip graph is common knowledge. Without such assumptions it is even easier to obtain such an impossibility result, a matter that we will address in the final section.

Theorem 38. *There is no epistemic symmetric protocol that is a strongly successful strengthening of LNS on all graphs.*

Proof. Consider the following "candy" graph G:



LNS is weakly successful on G, but there is no epistemic symmetric protocol P that is a strengthening of *LNS* and that is strongly successful on G.

In [16], it was shown that *LNS* is weakly successful on any graph that is neither a "bush" nor a "double bush". Since this graph G is neither a bush nor a double bush, *LNS* is weakly successful on it. For example, the sequence

$$02; 12; 53; 43; 13; 03; 23; 52; 42$$

is a successful *LNS* sequence which makes everyone an expert. LNS is not strongly successful on this graph, however. For example,

$$02; 12; 53; 43; 13; 03; 52; 42$$

is an unsuccessful *LNS* sequence, because 5 learns neither the number nor the secret of 4 and no further calls are allowed.

Now, suppose towards a contradiction that P is an epistemic symmetric strengthening of *LNS*, and that P is strongly successful on G.

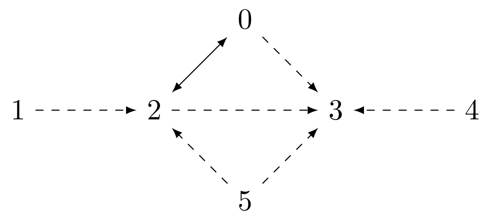
Before we look at specific calls made by P, we consider a general fact. Recall that knowing a *pure number* means knowing the number of an agent without knowing their secret. For any gossip graph and any agent a, if no one has a's pure number, then no call sequence will result in anyone learning a's pure number. After all, in order to learn a's number, one would have to call or be called by someone who already knows that number, but in such a call one would also learn a's secret.

In *LNS*, you are only allowed to call an agent if you have the number but not the secret of that agent, i.e., if you have their pure number. It follows that if, in a given gossip graph, no one has a's pure number, then no *LNS* sequence on that graph will contain any calls where a is the receiver.

In the gossip graph G under consideration, agents 0, 1, 4 and 5 are in the situation that no one else knows their number. So in particular, no one knows the pure number of any of these agents. It follows that 2 and 3 are the only possible targets for *LNS* calls in this graph.

Now, let us consider the first call according to P. This call must target 2 or 3. The calls 12 and 43 are bad calls, since they would result in 1 (resp. 4) being unable to make calls or be called, while still not being an expert.

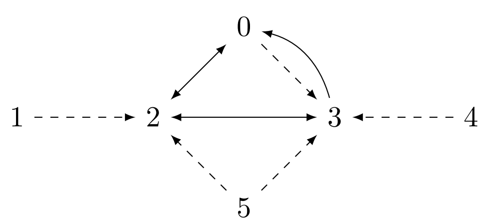
This means that either 0 or 5 must make the first call. By symmetry, we can assume without loss of generality that the first call is 02. This yields the following situation.



Now, let us look at the next call.

- The sequence $02; 43$ is bad, because that would make it impossible for 4 to ever become an expert.
- Because of the symmetry of P, the initial call could have been 03 instead of 02. The sequence $03; 12$ is bad, since 1 cannot become an expert, so $03; 12$ is not allowed by the strongly successful protocol P.

  But agent 1 cannot tell the difference between 03 and 02, so from the fact that $03; 12$ is disallowed and that P is epistemic it follows that $02; 12$ is also disallowed.
- The sequence $02; 03$ is bad, since 0 will not be able to make any call afterwards. Because 0 can also never be called, this implies that 0 will never become an expert.
- Consider then the sequence $02; 23$. This results in the following diagram.



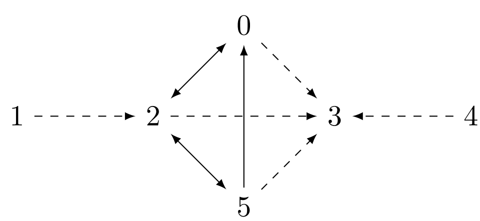
This graph has the following property: it is impossible (in any *LNS* sequence) for any agent to get to learn a new pure number. That is, nobody can learn a new number without also getting to know the secret of that agent: agents 1, 0, and 4 each know only one pure number, so they cannot teach anyone a new number, and agent 5 knows two pure numbers (2 and 3), but those agents already know each other's secrets.

As a result, any call that will become allowed by *LNS* in the future is already allowed now. There are 5 such calls that are currently allowed, namely 12, 52, 53, 03 and 43. Furthermore, of those calls 52 and 53 are mutually exclusive, since calling 2 will teach 5 the secret of 3, and calling 3 will teach 5 the secret of 2.

So any continuation of $02; 23$ allowed by *LNS* can only contain (in any order) 12, 03, 43 and either 52 or 53. Since P is a strengthening of *LNS*, the same holds for P. But using only those calls, there is no way to teach 3 the secret of 1: secret 1 can reach agent 2 using the call 12, but in order for the secret to travel any further we need the call 52. After that call only 03 and 43 are still allowed (in particular, 53 is ruled out), so the knowledge of secret 1 remains limited to agents 1, 2 and 5.

Since 02;13 cannot be extended to a successful *LNS* sequence, 02;13 must be disallowed.

- Consider the call sequence $02; 52$. This gives the following diagram.



Note that in this situation, it is impossible for agents 3 and 4 to learn any new number without also learning the secrets corresponding to those numbers: there is no agent that knows the number of agent 3 and that also knows another pure number, and this will remain the case whatever other calls happen.

This means that agent 3 cannot make any calls, and that agent 4 can make exactly one call, to agent 3.

Suppose now that $02; 52$ is extended to a successful *LNS* sequence. This sequence has to contain the call 43 at some point. This will be the only call by agent 4, so in order for the sequence to be successful, agent 3 already has to know secret 1 by the time 43 takes place.

In particular, this means that the call 12 has already happened, and that either agent 1 or agent 2 has then called agent 3 to transmit this secret. Whichever agent among 1 and 2 makes this call, afterwards they are unable to make any more calls. Furthermore, this takes place before the call 43, so whatever agent $x \in \{1, 2\}$ informs 3 of secret 1 does not learn secret 4. Since this agent x can neither make another call nor be called, it follows that x does not become an expert.

So $02; 52$ is not allowed by P which we assumed to be strongly successful.

- Finally, consider the call sequence $02; 53$. By symmetry, 03 could have been the first call as opposed to 02. Furthermore, the same reasoning that showed 02;52 to be unsuccessful above can, with an appropriate permutation of agents, be used to show that 03;53 is unsuccessful.

  Agent 5 cannot distinguish between the first call 02 and 03 before making the call 53, so if $03; 53$ is disallowed then so is $02; 53$ because P is epistemic.

Remember that 02 is, without loss of generality, the only initial call that can lead to success. We have shown that all of the *LNS*-permitted calls following the initial call 02 (namely, the calls 43, 12, 03, 23, 52 and 53) are disallowed by P. This contradicts P being a strongly successful strengthening of *LNS*. □

4.2 Backward Induction and Look-Ahead applied to Candy

Given this impossibility result, it is natural to wonder what would happen if we use the syntactic strengthenings from Definition 23, or their iterations, on the "candy" graph G.

All second calls are eliminated by $LNS^{\blacksquare}$, because for any two agents a and b we have $G, 02 \models \neg K_a^{LNS}[ab]\langle LNS\rangle Ex$. By symmetry this also holds for the three other possible first calls, hence $LNS^{\blacksquare}$ is unsuccessful on G. However, the first calls *are* still allowed according to $LNS^{\blacksquare}$.

There are 9468 *LNS*-sequences on this graph of which 840 are successful. Using the implementation discussed in the Appendix we found out that $LNS^{\blacklozenge}$, the soft look-ahead strengthening of *LNS*, is weakly successful on this graph and allows 840 successful and 112 unsuccessful sequences.

5 Conclusions, Comparison, and Further Research

Conclusions We modeled common knowledge of protocols in the setting of distributed dynamic gossip. A crucial role is played by the novel notion of protocol-dependent knowledge. This knowledge is interpreted using an epistemic relation over states in the execution tree of a gossip protocol in a given gossip graph. As the execution tree consists of gossip states resulting from calls permitted by the protocol, this requires a careful semantic framework.

We described various syntactically or semantically definable strengthenings of gossip protocols, and investigated the combination and iteration of such strengthenings, in view of strengthening a weakly successful protocol into one that is strongly successful on all graphs. In the setting of gossip, a novel notion we used in such strengthenings is that of uniform backward induction, as a variation on backward induction in search trees and game trees.

Finally, we proved that for the *LNS* protocol, in which agents are only allowed to call other agents if they do not know their secrets, it is impossible to define a strengthening that is strongly successful on all graphs.

Comparison As already described at length in the introductory section, our work builds upon prior work on dynamic distributed gossip [16, 15], which itself has a prior history in the networks community [23, 29, 20] and in the logic community [3, 1]. Many aspects of gossip may or may not be common knowledge among agents: how many agents there are, the time of a global clock, the gossip graph, etc. The point of our result is that even under the strongest such assumptions, one can still not guarantee that a gossip protocol always terminates successfully. How common knowledge of agents is affected by gossip protocol execution is investigated in [2]: for example, the authors demonstrate how sender-receiver subgroup common knowledge is obtained (and lost) during calls. However, they do not study common knowledge of gossip protocols. We do not know of other work on that topic. Outside the area of gossip, protocol knowledge has been well investigated in the epistemic logic community [26, 39, 12].

While the concept of backward induction is well-known in game theory (see for example [4]), it is only used in perfect-information settings, where all agents know what the real world or the actual state is. Our definition of *uniform* backward induction is a generalization of backward induction to the dynamic gossip setting, where only partial observability is assumed. A concept akin to uniform backward induction has been proposed in [35] (rooted in [8]), under the name of *common belief in future rationality*, with an accompanying recursive elimination procedure called *backward dominance.*[3] As in our approach, this models a decision rule faced with uncertainty over indistinguishable moves. In [35], the players are utility maximizers with probabilistic beliefs, which in our setting would correspond to *randomizing* over all indistinguishable moves/calls. As a decision rule this is also known as the *insufficient reason* (or *Laplace*) criterion: all outcomes are considered equiprobable. Seeing uniform backward induction as the combination of backward induction and a decision rule immediately clarifies the picture. Soft uniform backward induction applies the *minimax regret* criterion for the decision whom to call, minimizing the maximum utility loss. In contrast, hard uniform backward induction applies the *maximin utility* criterion, maximizing the minimum utility (also known as risk-averse, pessimistic, or Wald criterion).

In the gossip scenario, the unique minimum value is unsuccessful termination, and the unique maximum value is successful termination. Minimax prescribes that as long as the agent considers it possible that a call leads to successful termination, the agent is allowed to make the call (as long as the minimum of the maximum is success, go for it): the soft version. Maximin prescribes that, as long as the agent considers it possible that a call lead to unsuccessful termination, the agent should not make the call (as long as the maximum of the minimum is failure, avoid it): the hard version. Such decision criteria over uncertainty also crop up in areas overlapping with social software and social choice, e.g. [7, 11, 33, 31]. In [7] a somewhat similar concept has been called "common knowledge of stable belief in rationality". However, there it applies to a weaker epistemic notion, namely belief.

Further Research The impossibility result for *LNS* is for dynamic gossip where agents exchange both secrets and numbers, and where the network expands. Also in the non-dynamic setting we can quite easily find a graph where static *LNS* is weakly successful but cannot be strengthened to an epistemic symmetric strongly successful protocol. Consider again the "diamond" graph of Section 3.6, for which we described various strongly successful strengthenings. Also in "static" gossip *LNS* is weakly successful on this graph, since $01; 30; 20; 31$ is successful. All four possible first calls are symmetric. After 21, the remaining possible calls are 20, 31 and 30. But 20 is bad, since 2 will never learn secret 3 that way. Also 31 is bad, since agent 1 will never learn the secret of 0. The call 30 is safe and in fact guarantees success, but by epistemic symmetry it cannot be allowed while 31 is disallowed. Therefore, in the static setting it is impossible to strengthen *LNS* on "diamond" such that it becomes strongly successful. We expect a completely different picture for strengthening "static" gossip protocols in similar fashion as we did here, for dynamic gossip.

[3]We kindly thank Andrés Perea for his interactions.

We assumed synchronicity (a global clock) and common knowledge of the initial gossip graph. These strong assumptions were made on purpose, because without them agents will have even less information available and will therefore not be able to coordinate any better. Such and other parameters for gossip problems are discussed in [13]. It is unclear what results still can be obtained under fully distributed conditions, where agents only know their own history of calls and their neighbors.

We wish to determine the logic of protocol-dependent knowledge K_a^P, and also on fully distributed gossip protocols, without a global clock, and to further generalize this beyond the setting of gossip.

Appendix: A Model Checker for Dynamic Gossip

Analyzing examples of gossip graphs and their execution trees by hand is tedious. To help us find and check the examples in this paper we wrote a Haskell program which is available at `https://github.com/m4lvin/gossip`. Our program can show and randomly generate gossip graphs, execute the protocols we discussed and draw the resulting execution trees with epistemic edges. The program also includes an epistemic model checker for the formal language we introduced, similar to DEMO [17], but tailor-made for dynamic gossip. For further details, see also [19, Section 6.6].

Figure 2 is an example output of the implementation, showing the execution tree for Example 30 up to two calls, together with the epistemic edges for agent 2, here called c. Note that we use a more compact way to denote gossip graphs: lower case stands for a pure number and capital letters for knowing the number and secret.

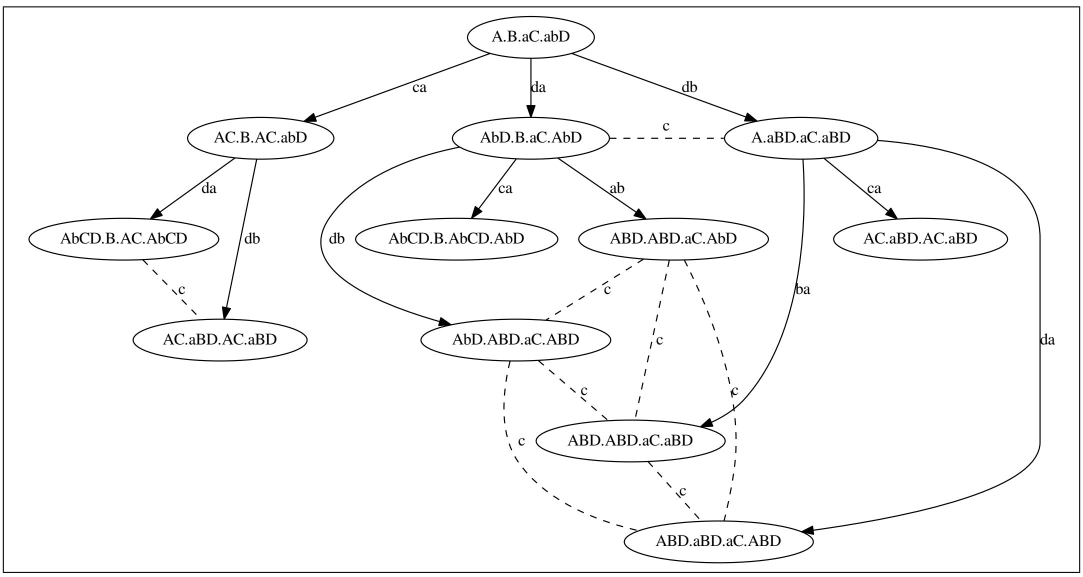


Figure 2: Two levels of the execution tree for Example 30, with epistemic edges for c.

Our implementation can run different protocols on a given graph and output a LaTeX table showing and comparing the extension of those protocols. Tables 2 and 3 have been generated in this way. They provide details how various strengthenings behave on the gossip graphs from Example 30 and Section 3.6.

	LNS	$.\blacksquare$	$.\blacklozenge$	$.\Box$	$.\Box 2$	$.\Box 3$	$.\Box 4$	$.\Diamond$	$.\Diamond 2$	$.\Diamond 3$	$.\Diamond 4$	$.\Diamond 5$
ϵ							×					
20						×				×		
20;30					×				×			
20;30;01				×				×				
20;30;01;31	×											
20;30;31				×				×				
20;30;31;01	×											
20;31					×				×			
20;31;10				×				×				
20;31;10;30	×											
20;31;30				×				×				
20;31;30;10	×											
30		×				×						
30;01					×							
30;01;20				×								
30;01;20;31	×							×	×	×	×	×
30;01;31				×				×	×	×	×	×
30;01;31;20	×											
30;20;01					×							
30;20;01;21;31	✓		✓	✓				✓	✓	✓	✓	✓
30;20;01;31;21	✓		✓					✓	✓	✓	✓	✓
30;20;21					×							
30;20;21;01;31	✓		✓	✓				✓	✓	✓	✓	✓
30;20;21;31;01	✓		✓					✓	✓	✓	✓	✓
30;20;31;01				×								
30;20;31;01;21	×							×	×	×	×	×
30;20;31;21				×								
30;20;31;21;01	×							×	×	×	×	×
30;31					×							
30;31;01				×				×				
30;31;01;20	×											
30;31;20									×	×	×	×
30;31;20;01				×				×				
30;31;20;01;21	×											
30;31;20;21				×				×				
30;31;20;21;01	×											
31						×						
31;10					×							
31;10;20				×				×	×	×	×	×
31;10;20;30	×											
31;10;30				×				×				
31;10;30;20	×											
31;20					×				×	×	×	×
31;20;10				×				×				
31;20;10;30	×											
31;20;30				×				×				
31;20;30;10	×											
31;30					×							
31;30;10				×				×				
31;30;10;20	×											
31;30;20				×				×	×	×	×	×
31;30;20;10	×											

Table 2: N Example 30: Extensions of strengthenings.

	LNS	$.^{\blacksquare}$	$(.^{\blacksquare})^{\Box}$	$.^{\blacklozenge}$	$.^{\Box}$	$.^{\Box 2}$	$.^{\Box 3}$	$.^{\Box 4}$	$.^{\Diamond}$	$.^{\Diamond 2}$	$.^{\Diamond 3}$	$(.^{\Diamond})^{\Box 3}$
ϵ								×				
01						×						
01;21					×				×	×	×	
01;21;30	×											
01;21;31	×											
01;30					×							
01;30;21	×								×	×	×	
01;31					×							
01;31;21	×								×	×	×	
21						×						
21;01					×				×	×	×	
21;01;30	×											
21;01;31	×											
21;30										×	×	
21;30;01					×				×			
21;30;01;31	×											
21;30;31					×				×			
21;30;31;01	×											
21;31					×							
21;31;01	×								×	×	×	
30			×				×					
30;01		×				×						
30;01;21;31	✓			✓					✓	✓	✓	
30;01;31;21	✓			✓	✓				✓	✓	✓	✓
30;21;01					×							
30;21;01;31	×			×					×	×	×	
30;21;31					×							
30;21;31;01	×			×					×	×	×	
30;31		×				×						
30;31;01;21	✓			✓	✓				✓	✓	✓	✓
30;31;21;01	✓			✓					✓	✓	✓	
31;01;21;30	✓			✓					✓	✓	✓	
31;01;30;21	✓			✓	✓				✓	✓	✓	
31;10;21;30	✓			✓					✓	✓	✓	
31;10;30;21	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
31;21;01;30	✓			✓					✓	✓	✓	
31;21;30	✓			✓	✓				✓	✓	✓	
31;30;10;21	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
31;30;21	✓			✓					✓	✓	✓	

Table 3: Diamond Example of Section 3.6: Extensions of strengthenings, after 20.

References

[1] Krzysztof R. Apt, Davide Grossi, and Wiebe van der Hoek. Epistemic protocols for distributed gossiping. In Ramanujam, editor, *Proceedings of TARK 2015*, 2015. https://doi.org/10/cm72.

[2] Krzysztof R. Apt and Dominik Wojtczak. Common knowledge in a logic of gossips. In Jérôme Lang, editor, *Proceedings of TARK 2017*, 2017. https://doi.org/10/gctp2b.

[3] Maduka Attamah, Hans van Ditmarsch, Davide Grossi, and Wiebe van der Hoek. Knowledge and gossip. In *Proceedings of the Twenty-first European Conference on Artificial Intelligence*, Frontiers in Artificial Intelligence and Applications, pages 21–26, 2014. https://doi.org/10/cm7w.

[4] Robert J. Aumann. Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8(1):6–19, 1995. https://doi.org/10/bxrwkf.

[5] Leemon Baird. The swirlds hashgraph consensus algorithm: Fair, fast, byzantine fault tolerance, 2017. https://www.swirlds.com/downloads/SWIRLDS-TR-2016-01.pdf.

[6] Brenda Baker and Robert Shostak. Gossips and telephones. *Discrete Mathematics*, 2(3):191–193, 1972. https://doi.org/10/bddz44.

[7] Alexandru Baltag, Sonja Smets, and Jonathan Alexander Zvesper. Keep 'hoping' for rationality: a solution to the backward induction paradox. *Synthese*, 169(2):301–333, 2009. https://doi.org/10/drvr2k.

[8] Pierpaolo Battigalli and Marciano Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106(2):356–391, 2002. https://doi.org/10/c4z66x.

[9] Johan van Benthem, Jelle Gerbrandy, Tomohiro Hoshi, and Eric Pacuit. Merging frameworks for interaction. *Journal of Philosophical Logic*, 38:491–526, 2009.

[10] Kai Brünnler, Dandolo Flumini, and Thomas Studer. A logic of blockchain updates. In Sergei Artemov and Anil Nerode, editors, *Logical Foundations of Computer Science*, pages 107–119, 2017. https://doi.org/10/cvpp.

[11] Vincent Conitzer, Toby Walsh, and Lirong Xia. Dominating manipulations in voting with partial information. In *Proc. of AAAI*, 2011.

[12] Hans van Ditmarsch, Sujata Ghosh, Rineke Verbrugge, and Yanjing Wang. Hidden protocols: Modifying our expectations in an evolving world. *Artificial Intelligence*, 208:18–40, 2014.

[13] Hans van Ditmarsch, Davide Grossi, Andreas Herzig, Wiebe van der Hoek, and Louwe B. Kuijer. Parameters for epistemic gossip problems. In *Proceedings of LOFT 2016*, 2016. https://sites.google.com/site/lbkuijer/LOFT_Gossip_Revised.pdf.

[14] Hans van Ditmarsch, Ioannis Kokkinis, and Anders Stockmarr. Reachability and expectation in gossiping. In *Proceedings of PRIMA 2017*, 2017. https://sites.google.com/site/ykokkinis/prima17.pdf.

[15] Hans van Ditmarsch, Jan van Eijck, Pere Pardo, Rahim Ramezanian, and François Schwarzentruber. Epistemic protocols for dynamic gossip. *Journal of Applied Logic*, 20:1–31, 2017. https://doi.org/10/f9p6c3.

[16] Hans van Ditmarsch, Jan van Eijck, Pere Pardo, Rahim Ramezanian, and François Schwarzentruber. Dynamic gossip. *Bulletin of the Iranian Mathematical Society*, 2018. https://doi.org/10/cvpm.

[17] Jan van Eijck. DEMO — a demo of epistemic modelling. In *Interactive Logic. Selected Papers from the 7th Augustus de Morgan Workshop, London*, volume 1, pages 303–362, 2007. https://homepages.cwi.nl/~jve/papers/07/pdfs/DEMO_IL.pdf.

[18] Patrick Th. Eugster, Rachid Guerraoui, Anne-Marie Kermarrec, and Laurent Massoulié. Epidemic information dissemination in distributed systems. *IEEE Computer*, 37(5):60–67, 2004. https://doi.org/10/d7rvbq.

[19] Malvin Gattinger. *New Directions in Model Checking Dynamic Epistemic Logic*. PhD thesis, University of Amsterdam, 2018. https://malv.in/phdthesis.

[20] Bernhard Haeupler. Simple, fast and deterministic gossip and rumor spreading. *Journal of the ACM*, 62(6), 2015. https://doi.org/10/f73wb4.

[21] Bernhard Haeupler, Gopal Pandurangan, David Peleg, Rajmohan Rajaraman, and Zhifeng Sun. Discovery through gossip. *Random Structures & Algorithms*, 48(3):565–587, 2016. https://doi.org/10/f8gkgm.

[22] Joseph Y. Halpern and Rafael Pass. A knowledge-based analysis of the blockchain. In Jérôme Lang, editor, *Proceedings of TARK 2017*, 2017. https://doi.org/10/gctp2b.

[23] Mor Harchol-Balter, Frank Thomson Leighton, and Daniel Lewin. Resource discovery in distributed networks. In *Proceedings of the Eighteenth Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 229–237, 1999. https://doi.org/10/dzgsmz.

[24] Sandra M. Hedetniemi, Stephen T. Hedetniemi, and Arthur L. Liestman. A survey of gossiping and broadcasting in communication networks. *Networks*, 18(4):319–349, 1988. https://doi.org/10/dnzk4d.

[25] Andreas Herzig and Faustine Maffre. How to share knowledge by gossiping. *AI Communications*, 30(1):1–17, 2017. https://doi.org/10/f94qxh.

[26] Tomohiro Hoshi. *Epistemic Dynamics and Protocol Information*. PhD thesis, Amsterdam University, 2009. https://www.illc.uva.nl/cms/Research/Publications/Dissertations/DS-2009-08.text.pdf.

[27] Cor A. J. Hurkens. Spreading gossip efficiently. *Nieuw Archief voor Wiskunde*, 5(1):208–210, 2000. http://www.nieuwarchief.nl/serie5/pdf/naw5-2000-01-2-208.pdf.

[28] Mix Irving. Gossiping securely is the new email. https://is.gd/IrvGossip.

[29] Richard M. Karp, Christian Schindelhauer, Scott Shenker, and Berthold Vöcking. Randomized rumor spreading. In *41st Annual Symposium on Foundations of Computer Science, FOCS*, pages 565–574, 2000. https://doi.org/10/fgpb7t.

[30] Rana Klein. The logical dynamics of gossip: an analysis in dynamic epistemic logic. Master's thesis, University of Amsterdam, 2017. https://eprints.illc.uva.nl/1567/.

[31] Reshef Meir. Plurality voting under uncertainty. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pages 2103–2109, 2015. https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9784.

[32] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. The MIT Press, 1994. Electronic edition, freely accessible. https://mitpress.mit.edu/books/course-game-theory.

[33] Rohit Parikh, Çağil Taşdemir, and Andreas Witzel. The power of knowledge in games. *IGTR*, 15(4), 2013. https://doi.org/10/cnch.

[34] Rohit Parikh and Ramaswamy Ramanujam. A knowledge based semantics of messages. *Journal of Logic, Language and Information*, 12:453–467, 2003.

[35] Andrés Perea. Belief in the opponents' future rationality. *Games and Economic Behavior*, 83(Supplement C):231–254, 2014. https://doi.org/10/f5t9zj.

[36] Yonatan Sompolinsky, Yoad Lewenberg, and Aviv Zohar. SPECTRE: A fast and scalable cryptocurrency protocol. Cryptology ePrint Archive, Report 2016/1159, 2016. https://eprint.iacr.org/2016/1159.

[37] Alfred Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285–309, 6 1955. https://doi.org/10/cvpn.

[38] Robert Tijdeman. On a telephone problem. *Nieuw Archief voor Wiskunde*, 3(19):188–192, 1971.

[39] Yanjing Wang. *Epistemic Modelling and Protocol Dynamics*. PhD thesis, Amsterdam University, 2010. https://www.illc.uva.nl/cms/Research/Publications/Dissertations/DS-2010-06.text.pdf.

 Received 20 April 2018