HOKKAIDO UNIVERSITY

Title	SOCREAL 2010: Proceedings of the 2nd International Workshop on Philosophy and Ethics of Social Reality [Full version]
Description	SOCREAL 2010: 2nd International Workshop on Philosophy and Ethics of Social Reality. Sapporo, Japan, 2010-03-27/28 Co-Chairs: Johan van Benthem, Tomoyuki Yamada
Issue Date	2010
Doc URL	https://hdl.handle.net/2115/43231
Type	conference proceedings
File Information	SR2010.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

SOCREAL 2010

Proceedings of the
2nd International Workshop
on
Philosophy and Ethics of
Social Reality

27–28 March 2010
Hokkaido University
Sapporo, JAPAN

Co-Chairs
Johan van Benthem
Tomoyuki Yamada

Table of Contents

Preface

In the past two decades, a number of logics and game theoretical analyses have been proposed and combined to model various aspects of social interaction among agents including individual agents, organizations, and individuals representing organizations. The aim of SOCREAL Workshop is to bring together researchers working on diverse aspects of such interaction in logic, philosophy, ethics, computer science, cognitive science and related fields in order to share issues, ideas, techniques, and results.

The first SOCREAL Workshop was held in March 2007 under the auspices of GPAE (Graduate Program in Applied Ethics, Graduate School of Letters, Hokkaido University) sponsored by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). Building upon the success of SOCREAL 2007, its second edition, SOCREAL 2010, was held in March 2010 under the auspices of CAEP (Center for Applied Ethics and Philosophy, Graduate School of Letters, Hokkaido University).

SOCREAL 2010 consisted of three keynote lectures by invited speakers and eight contributed talks presenting state-of-the-art results and followed by stimulated discussions, and a general discussion session on what role can be played by formal methods in studying social reality. It was a enormous success. The present volume contains extended abstracts, handouts, and/or slides of the keynote lectures and the contributed talks given in SOCREAL 2010. For the purpose of reference, we have given virtual page numbers to the pages of each contribution, but we refrained from printing the page numbers on the pages where the format of the page is not suitable for doing so. The page number of the first page of each contribution is given in the table of contents. In the cases where the authors have given us both their abstracts and slides, we let the abstracts precede the slides in this volume.

The contributed papers are selected out of about 30 submitted papers. We thank all the researchers who submitted their papers for their interest in SOCREAL 2010, and the members of program committee for their reviews. We also thank CAEP and Graduate School of Letters of Hokkaido University for their generous support. And lastly, we thank all the participants for their contribution to the success of SOCREAL 2010.

Johan van Benthem
Tomoyuki Yamada

SOCREAL 2010

2nd International Workshop on Philosophy and Ethics of Social Reality
27–28 March 2010, Hokkaido University, Sapporo, JAPAN
http://www.hucc.hokudai.ac.jp/ k15696/home/sr10/

The Program

DAY 1: 27 March 2010 Room W409, Humanities and Social Sciences Classroom Building
09:30 - 09:50 Registration
09:50 - 10:00 Opening
Keynote Lecture 1
10:00 - 11:00 Johan van Benthem (University of Amsterdam, Netherlands, Stanford University, USA) Joining Information and Evaluation: a dynamic logical perspective
Session 1: Obligations and Preferences
11:15 - 12:00 Allard Tamminga (University of Groningen, Netherlands) Nash Equilibria in Multi-agent Deontic Logic 12:15 - 13:00 Satoru Suzuki (Komazawa University, Japan) Measurement-Theoretic Foundation of Threshold Utility Maximiser's Preference Logic
13:00 - 15:00 Lunch Break
Session 2: Normative Systems
15:00 - 15:45 Berislav Žarnić (University of Split, Croatia) A Logical Typology of Normative Systems 16:00 - 16:45 Yasuo Nakayama (Osaka University, Japan) Logical Framework for Normative Systems
Keynote Lecture 2
17:00 - 18:00 Fenrong Liu (Tsinghua University, China) Understanding Deontics from a Preference Perspective

DAY 2: 28 March 2010
Room W409, Humanities and Social Sciences Classroom Building

Session 3: Responsibility and Collective Agency

10:00 - 10:45 Matthew Braham & Martin van Hees
(University of Groningen, Netherlands)
Responsibility in Games

11:00 - 11:45 Biswanath Swain
(Indian Institute of Technology Kanpur, India)
Is Intention sufficient to explicate Collective Agency?

12:00 - 12:45 Ryoji Fujimoto & Choi Chang-bong
(Hokkaido University, Japan)
Interpretation of Action and Sociality of Action

12:45 - 14:45 Lunch Break

Session 4: Argumentations and Speech Acts

14:45 - 15:30 Miranda del Corral
(Universidad Nacional de Educación a Distancia, Spain)
Social Commitments: Expectations, Obligations and Entitlements

15:45 - 16:30 General Discussion Chaired by Johan van Benthem

Keynote Lecture 3

16:45 - 17:45 Tomoyuki Yamada (Hokkaido University, Japan)
Scorekeeping and Dynamic Logics of Speech Acts

18:00 - 18:10 Closing

WORKSHOP CO-CHAIRS

Johan van Benthem
(University of Amsterdam, The Netherlands, and Stanford University, USA)
Tomoyuki Yamada (Hokkaido University, Japan)

PROGRAM COMMITTEE

Johan van Benthem
(University of Amsterdam, The Netherlands, and Stanford University, USA)
Jose Carmo (Universidade da Madeira, Portugal)
Fenrong Liu (Tsinghua University, China)
Jun Miyoshi (Kanto Gakuin University, Japan)
Yuko Murakami (Tohoku University, Japan)
Yasuo Nakayama (Osaka University, Japan)
Manuel Rebuschi (Nancy University, France)
Nobuyuki Takahashi (Hokkaido University, Japan)
Allard Tamminga (Rijksuniversiteit Groningen, The Netherlands)
Tomoyuki Yamada (Hokkaido University, Japan)
Berislav Žarnić (University of Split, Croatia)

LOCAL ORGANIZING COMMITTEE

Nobuo Kurata (Hokkaido University, Japan)
Shunzo Majima (Hokkaido University, Japan)
Koji Nakatogawa (Hokkaido University, Japan)
Yoshihiko Ono (Hokkaido University, Japan)
Tomoyuki Yamada (Hokkaido University)

LOGICAL DYNAMICS OF INFORMATION AND EVALUATION

Johan van Benthem, Amsterdam & Stanford, http://staff.science.uva.nl/~johan/

SOCREAL Workshop, University of Hokkaido, 27 March 2010

1 Logic and the balance of information and evaluation

Logic has traditionally been concerned with truth, information, knowledge, and belief. Evaluation and other ways of 'colouring' the world as we see it belong to another realm. But one can argue that truly rational agency involves a *harmony* between informational and evaluative attitudes: including the ability to change both as required by circumstances.

2 Logical dynamics of information-driven rational agency

Agency involves a broad range of correct information processing:

Restaurant: how to figure out who has which dish? *Inference, questions*.

Card games: planning moves using theory of mind. *Mutual knowledge*.

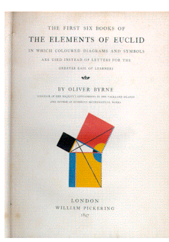
Skype Exam: secret voting on a public channel. *New social procedures*.

To be rational is to reason *intelligently*. Logic of all basic informational processes:

"Zhi: Wen, Shuo, Qin" 知 问 说 亲 (communication, inference, observation)

To be rational is to act *intelligently*. Evaluation, goals, preferences, decisions, actions.

To be rational is to interact *intelligently*. Argumentation, communication, games.





3 The research program

Stage One: charting the relevant abilities of an agent. Knowledge, belief, etc. Different sorts of information: semantic, syntactic. Awareness. Information vs. evaluation.

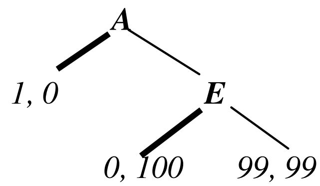
Stage Two: making the local dynamics explicit. Observation, inference, communication. Knowledge and information update, belief revision (learning, self-correction), issue management (questions, agenda), changing preferences and goals.

Stage Three: social features. Groups, collective information, attitudes, action.

Stage Four: mid-term temporal behaviour. Conversation, games and strategies.

Stage Five: long-term temporal behaviour. Dynamic logic meets dynamical systems.

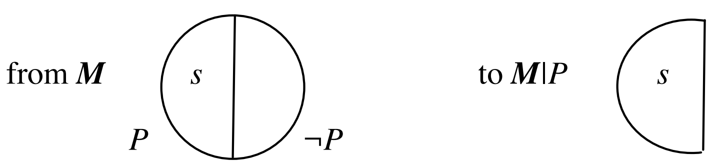
It all comes together naturally in *games*, mathematical, philosophical, computational logic:



4 Pilot system: dynamifying epistemic logic to public announcement logic

Static base logic. Language $p \mid \neg\phi \mid \phi \vee \psi \mid K_i\phi \mid C_G\phi$, models $\boldsymbol{M} = (W, \{\sim_i \mid i \in G\}, V)$, with worlds W, accessibility relations $\sim_i$, and valuation V. Truth conditions ('knowledge as semantic information'): $\boldsymbol{M}, s \models K_i\phi$ iff for all t with $s \sim_i t$: $\boldsymbol{M}, t \models \phi$, and $\boldsymbol{M}, s \models C_G\phi$ iff for all t reachable from s by some finite sequence of $\sim_i$ steps ($i \in G$): $\boldsymbol{M}, t \models \phi$.

Dynamic logics describe key update steps in observation and communication. Pilot system: *PAL*. Hard information update: learning P eliminates worlds with P false:



Language extension: $\boldsymbol{M}, s \models [!P]\,\phi$ iff *if* $\boldsymbol{M}, s \models P$, *then* $\boldsymbol{M}|P, s \models \phi$

Theorem *PAL* axiomatized completely by epistemic logic plus *recursion axioms*:

$[!P]q$	$\leftrightarrow$	$P \rightarrow q$	*for atomic facts* q
$[!P]\neg\phi$	$\leftrightarrow$	$P \rightarrow \neg[!P]\phi$	
$[!P]\phi \wedge \psi$	$\leftrightarrow$	$[!P]\phi \wedge [!P]\psi$	
$\boldsymbol{[!P]K_i\phi}$	$\leftrightarrow$	$\boldsymbol{P \rightarrow K_i(P \rightarrow [!P]\phi)}$	*key recursion axiom*

Aside on 'schematic validities': $[!P][!Q]\phi \leftrightarrow [!(P \wedge [!P]Q)]\phi$

Methodology ***Make actions explicit*** on top of static logic. Compositional analysis effects. *Hunt for right* ***recursion axioms***: also for private information, belief revision, questions. Can describe information flow under many events. Current developments: (a) eventually, no reduction: from single steps to *temporal setting with protocols*, (b) syntactic dynamics.

Similar methods developed for evaluation dynamics: ***preference, intentions, goals***.

5 Hard information, world elimination and belief change

Belief and plausibility models Conditional logic of relative plausibility:

$\boldsymbol{M}, s \models B_i\phi$ iff $\boldsymbol{M}, t \models \phi$ for all worlds t *minimal* in the ordering $\lambda xy. \leq_{i,s} xy$.

Belief change under hard facts:

$[!P]\, B_i\phi \;\leftrightarrow\; P \rightarrow B_i^{P}\, [!P]\phi$

Conditional belief helps *pre-encode* beliefs we would have if we learnt certain things:

$\boldsymbol{M}, s \models B_i^{\psi}\, \phi$ iff $\boldsymbol{M}, t \models \phi$ for all worlds t which are minimal for $\lambda xy. \leq_{i,s} xy$ in the set $\{u \mid \boldsymbol{M}, u \models \psi\}$.

Satisfies the standard principles of the minimal conditional logic.

How beliefs change under hard information (cf. changing obligations):

Theorem The logic of knowledge and conditional belief under public announcements is axiomatized completely by (a) any complete base logic with $B_i^{\psi}\, \phi$ for favorite model class, (b) *PAL* axioms, plus (c) a recursion axiom for conditional beliefs:

$[!P]\, B_i^{\psi}\, \phi \;\leftrightarrow\; P \rightarrow B_i^{P \wedge [!P]\psi}\, [!P]\phi$

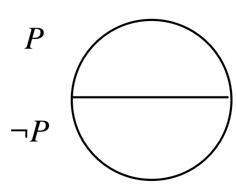
Interesting scenarios Misleading true information. Motivates new notion of 'safe belief' (truth in all more plausible worlds) between knowledge and belief: also in *AI,* game theory, philosophy. So, ***what is natural repertoire of agent attitudes?*** Same issues for evaluation.

6 Belief revision as relation change under soft information

Soft triggers Call for belief change $*p$ 'softer' than world elimination, introducing just greater 'preference' for p-worlds. *Default rule* $A \Rightarrow B$ does not say all A-worlds are B. Just makes exceptional $A \wedge \neg B$-worlds less plausible. 'Soft information' does not eliminate, but *changes the plausibility ordering* of the existing worlds. Typical case:

Lexicographic upgrade $\Uparrow P$ changes the current model $\boldsymbol{M}$ to $\boldsymbol{M}\Uparrow P$:

P-worlds now better than all $\neg P$-worlds; within zones, old order remains.



Social revolution: underclass P now becomes upper class. Other policies (Rott's '27'; or Macchiavelli's conservative advice, $\uparrow P$, just co-opt *leaders* of the underclass!). Logic:

$$\boldsymbol{M}, s \models [\Uparrow P]\phi \qquad \text{iff} \qquad \boldsymbol{M}\Uparrow P, s \models \phi.$$

Theorem The dynamic logic of lexicographic upgrade is axiomatized completely by the logic of conditional belief + compositional analysis of effects of revision:

$$[\Uparrow P]\, B^{\psi}\phi \;\leftrightarrow\; (\langle\rangle(P \wedge [\Uparrow P]\psi) \wedge B^{\,P \wedge [\Uparrow P]\psi}\, [\Uparrow P]\phi) \vee (\neg\langle\rangle(P \wedge [\Uparrow P]\psi) \wedge B^{\,[\Uparrow P]\psi}\, [\Uparrow P]\phi$$

Here E is the epistemic existential modality. Special cases: factual formulas. We can now axiomatize many policies for relation change, in various formats (cf. Baltag's lecture).

7 Preference and deontic betterness dynamics

Static models: ***betterness order*** on worlds. Deontic setting is social: betterness of the moral authority. Recurrent choice: total orders or *pre-orders*: indifference ≠ incomparability?

Modal language, complete axiomatizations in many flavours (back to Bouilier in AI).

Generic preference $P\varphi\psi$ between propositions by set lifting (Liu 2008, and long literature).

Special features of preference: ceteris paribus reasoning (van Benthem, Girard, Roy 2009).

Dynamic actions: betterness order can change under explicit suggestions, commands, etc. (Yamada, Liu). Example 'suggestion' $\#\boldsymbol{\phi}$: remove all links from $\boldsymbol{\phi}$–worlds to $\neg\boldsymbol{\phi}$–worlds:

For each model $\boldsymbol{M}, w$, the model $\boldsymbol{M}\#\varphi, w$ is $\boldsymbol{M}, w$ with the new relation $\leq' = \leq - \{(x, y) \mid \boldsymbol{M}, x \models \varphi \;\&\; \boldsymbol{M}, y \models \neg\varphi\}$.

Next, we enrich the formal language by adding action modalities interpreted as follows: [1]

$$M, w \models [\#(\varphi)]\psi \qquad \text{iff} \qquad M\#\varphi, w \models \psi$$

Theorem The dynamic logic of preference change under suggestions is axiomatized completely by the static modal logic of the underlying model class plus axioms as above, plus this key recursion axiom for betterness after suggestions:

$$\boldsymbol{[\#(\varphi)]\langle\leq\rangle\psi \quad\leftrightarrow\quad (\neg\varphi \;\&\; \langle\leq\rangle[\#(\varphi)]\psi) \vee ((\varphi \;\&\; \langle\leq\rangle(\varphi \;\&\; [\#(\varphi)]\psi).}$$

Which triggers for betterness change? Speech acts, juridical acts? Feelings?

---

[1] Here the syntax is recursive: the formula φ may itself contain dynamic modalities.

8 Priority graphs and two-level dynamics

Idea: derive betterness order from criteria (in many disciplines, even linguistics).

Linear priority sequences $\boldsymbol{P}$, de Jongh & Liu 2007, ***derived world order***:

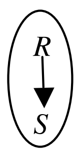
> $x < y$ iff x, y differ in at least one property in $\boldsymbol{P}$, and
> the first $P \in \boldsymbol{P}$ where this happens is one with $Py, \neg Px$.

Is lexicographic ordering, if we view each property $P \in \boldsymbol{P}$ as $x \leq^P y$ iff $(Py \rightarrow Px)$.

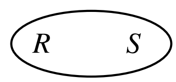
Graphs for ***relation merge*** (Andréka, Ryan & Schobbens 2002):

> Given an ordered *priority graph* $\boldsymbol{G} = (G, <)$ of indices for relations that may have multiple occurrences in the graph, the *merged group priority relation* is:
> $x \leq_G y$ iff for all indices $i \in \boldsymbol{G}$, either $x \leq_i y$, or there is some $j > i$ in $\boldsymbol{G}$ with $x <_i y$

Example



(hierarchy)



(juxtaposition)

Putting R above S, $x \leq y$ iff $x\ R\ y \wedge x\ S\ y$ or there is a difference in the S relation and $x\ R^+\ y$ with R^+ the strict form of R. Putting R alongside S is intersection $x\ R\ y \wedge x\ S\ y$.

Dynamics at graph level Add propositions in front, behind, in middle (syntactic).

Two natural mathematical operations that change and combine priority graphs:

> *sequential composition* $G_1 ; G_2$ (putting G_1 on top of G_2, retaining the same order inside) and *parallel composition* $G_1 \,||\, G_2$ (disjoint union of graphs).

Removing items, much harder.

Two-level dynamics *Two-level structures* $(W, \leq, P, <)$ having both worlds with a betterness order $\leq$ and a set of 'important propositions' with a primitive priority order $<$:

$$\begin{array}{ccc} & \boldsymbol{P}, < & \\ \textit{lifting} \uparrow & & \downarrow \textit{deriving} \\ & W, \leq & \end{array}$$

Fact The identity $lex(\varphi ; \boldsymbol{P}) = \Uparrow\varphi\,(lex(\boldsymbol{P}))$ holds, making this diagram commute:

$$\begin{tikzcd} \boldsymbol{P}, < \arrow[r] \arrow[d, dash, "lex"'] & \varphi ; \boldsymbol{P}, < \arrow[d, dash, "lex"] \\ \boldsymbol{W}, \leq \arrow[r] & \boldsymbol{W}, \Uparrow\varphi(\leq) \end{tikzcd}$$

9 Issues from information dynamics that also play for evaluation

Evaluation Sources for its importance: individual values, decisions, games, 'belonging' and shared agency, society. (Connections to social epistemology and cognitive science.)

Concrete case study: deontics Does this line of thinking apply to deontic logic, norms, laws? Law = priority graph? (See lecture of Fenrong Liu.) Which sort of structure needed to model real deliberation and *judgment*? A few ways of extending the above framework:

Entanglement: preference, belief, obligation, action

Foundations of action: preference ~ beliefs ~ actions. Belief-entangled preference (de Jongh & Liu). Duties and having or gathering information: Pacuit & Parikh.

Information change and *betterness change*, both needed? (Curse: *re-encoding*.)

Games and social interaction

Deontics is social. Legal procedure in court. Laws as social software: Law and Economics.

Fine-grained models: syntactic plus semantic dynamics

Problem of semantic versus syntactic information: informativity of deduction.

Syntactic structure of criteria, laws, proof steps, organization of reasoning.

Moral duty to deliberate well? Duty to think?

10 Conclusion

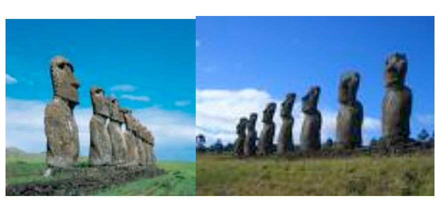
The dynamic logical tools that we have seem to apply to both information and evaluation. Now the challenge is to look at concrete issues, and see how far they take us.

References

J. van Benthem, 1996, *Exploring Logical Dynamics*, CSLI Publications, Stanford.

2009, 'For Better of for Worse: Dynamic Logics of Preference', in T. Grüne-Yanoff & S-O Hansson, eds., *Preference Change*, Springer, Dordrecht, 57–84.

2010, *Dynamic Logic of Information and Interaction*, Cambridge UP, to appear.



Understanding Deontics from a Preference Perspective

Fenrong Liu

Tsinghua University, Beijing, China

27 March 2010, Hokkaido University, Japan

Outline

This is a report on my recent joint work with
Johan van Benthem and **Davide Grossi**.

Betterness and Obligation

“ [...] to assert that a certain line of conduct is [...] absolutely right or obligatory, is obviously to assert that more good or less evil will exist in the world, if it is adopted, than if anything else be done instead." [Moore, Principia Ethica, 1903]

OBSERVATION 1: No obligation without an order on the “possible states of the world".

Example: Dyadic obligations

Dyadic obligations of the type "it is obligatory that φ under condition ψ" are interpreted by making use of an 'ideality relation' and the notion of maximality:

$$\mathcal{M}, s \models \mathbf{O}(\varphi \mid \psi) \Longleftrightarrow Max(||\psi||_{\mathcal{M}}) \subseteq ||\varphi||_{\mathcal{M}} \tag{1}$$

where $||.||_{\mathcal{M}}$ denotes the truth-set function of $\mathcal{M}$ and $\mathcal{M}$ is a model built on a Kripke frame $\mathcal{F} = (S, \preceq)$. In this frame the states in S are ordered according to the ideality relation $\preceq$.

The deontic notions of obligation, permission and prohibition can be naturally made sense of in terms of an "ideality" ordering $\preceq$ on possible worlds.

Betterness and Priority

OBSERVATION 2: The betterness relation between states is, often, derived from some kind of explicit *coding* of what is better in terms of relevant properties.

As the following quote illustrates in a lively manner:

> *"It is good for a man not to touch a woman. But if they cannot contain, let them marry: for it is better to marry than to burn." [St. Paul, Ch. 7]*

In the terminology of deontic logic, this is a typical **contrary-to-duty** structure (Prakken and Sergot, 1996) expressing what states are best, what states are best among the non-best ones, and so on, up to a finite depth.

Our Plan

In the following sections we will discuss the above type of structures in the light of notions and results developed in recent work in preference logic, and illustrate them by formalizing a classical example of CTD obligations.

Basic definitions: P-sequence

Definition

Let $\mathcal{L}(\mathbf{P})$ be a propositional language, S a non-empty set of states and $\mathcal{I} : \mathbf{P} \longrightarrow 2^S$ a valuation function. A P-sequence for $\mathcal{I}$ is a tuple $\mathcal{B}^{\mathcal{I}} = \langle B, \prec \rangle$ where:

- $B \subseteq \mathbf{P}$ and $|B| < \omega$;
- $\prec$ is a strict linear order on B(Irreflexive, transitive, antisymmetric and total);
- for all $\varphi, \psi \in B$, $\varphi \prec \psi$ iff $||\psi||_{\mathcal{I}} \subset ||\varphi||_{\mathcal{I}}$.

where $||\varphi||_{\mathcal{I}}$ denotes the truth-set of φ according to $\mathcal{I}$.

Deriving preferences from P-sequences

Definition

Let $\mathcal{B} = \langle B, \prec \rangle$ be a P-sequence, S a non-empty set of states and $\mathcal{I} : \mathbf{P} \longrightarrow 2^S$ a valuation function. The preference relation $\preceq^{IM}_{\mathcal{B}} \subseteq S^2$ is defined as follows:

$$w \preceq^{IM}_{\mathcal{B}} w' \quad := \quad \forall \varphi \in B : w \in ||\varphi|| \Rightarrow w' \in ||\varphi||. \qquad (2)$$

where IM is just a mnemonics for 'implication'.
Given a P-sequence $\mathcal{B}$ for a valuation $\mathcal{I}$, Formula 2 generates also a Kripke model $\mathcal{M}^{IM}_{\mathcal{B}} = (S, \preceq^{IM}_{\mathcal{B}}, \mathcal{I})$.

There are various definitions of deviation in the literature.

P-sequence and CTDs

J. Forrester. Gentle murder, or the adverbial samaritan. *Journal of Philosophy*, 81:193-197, 1984.

Example (Gentle murder)

"Here is the problem: Let us suppose a legal system which forbids all kinds of murder, but which considers murdering violently to be a worse crime than murdering gently. [. . .] The system then captures its views about murder by means of a number of rules, including these two:

- It is obligatory under the law that Smith not murder Jones.
- It is obligatory that, if Smith murders Jones, Smith murders Jones gently."

Assuming Anderson's reductionistic perspective, the scenario partitions all possible states in three classes:

- class I_1, in which Smith does not murder Jones.
- class I_2, in which Smith murders Jones gently. Note that it is contained in $\neg I_1 = V_1$)
- class $\neg I_2$, in which Smith murders Jones and he does not do it gently. Note that it is also contained in $\neg I_1 = V_1$).

We thus have the P-sequence $\mathcal{B}$ such that $I_2 \prec I_1$. Such P-sequence is sufficient to order the states.

Anderson's Reduction

The idea is to reduce deontic logic formulae, such as $\mathbf{O}\varphi$, to alethic modal logic $\Box$-formulae containing a designated violation or ideality constant, that is:

$$\mathbf{O}\varphi := \Box(\neg\varphi \to \mathsf{V}) \tag{3}$$

$$\mathbf{O}\varphi := \Box(\mathsf{I} \to \varphi) \tag{4}$$

To sum up, the intuition behind a P-sequence $p_1, \ldots, p_n$ for a given interpretation function is that each atom p_i gives rise to a bipartition $\{\mathcal{I}(p_i), -\mathcal{I}(p_i)\}$ of the domain of discourse S, and the more we move towards the right-hand side (i.e., the bottom) of the sequence the more atoms p_i are falsified.

A similar analysis can be applied to other puzzles in deontic logic.

Chisholm's paradox

1. It ought to be that Smith refrains from robing Jones.
2. Smith robs Jones.
3. If Smith robs Jones, he ought to be punished for robbery.
4. It ought to be that if Smith refrains from robbing Jones he is not punished for robbery.

A natural and consistent interpretation of the Chisholm's scenario in terms of classification and preference goes as follows:

1. It is most ideal that Smith refrains to rob Jones;
2. Smith robs Jones;
3. The most ideal states under the assumption that Smith robs Jones are states in which Smith is punished;
4. It is most ideal that Smith is not punished.

Modal Preference Logic: Language

Language $\mathcal{L}(\forall, \preceq)$ is built from a countable set **P** of atoms according to the following BNF:

$$\mathcal{L}(\forall, \preceq) : \varphi ::= p \mid \top \mid \neg\varphi \mid \varphi \wedge \varphi \mid [\preceq]\varphi \mid [\forall]\,\varphi$$

where $p \in$ **P**. Modal duals and Boolean operators are defined as usual. Intuitively, $[\preceq]$ quantifies over all states which are at least as good as the current one, and $[\forall]$ over all states (universal modality).

It was proposed first in [Boutilier 1993, 1994], later used in [Halpern 1997], recently developed in [Girard 2008], [Liu 2008] and [Roy, 2008].

Models

Definition (Models)

A model for $\mathcal{L}(\forall, \preceq)$ on the set of atoms **P** is a tuple $\mathcal{M} = \langle S, \preceq, \mathcal{I} \rangle$ where:

- S is a non-empty set of states;
- $\preceq$ is a conversely well-founded total preorder over S; (reflexive, transitive, connected, no infinite ascending chain.)
- $\mathcal{I} : \mathbf{P} \longrightarrow 2^S$.

As usual, we define $s \prec s'$ as $s \preceq s'$ and $s' \not\preceq s$.

Truth Definition

Definition

Let $\mathcal{M} \in \mathbb{M}$. The satisfaction of a formula $\varphi \in \mathcal{L}(\forall, \preceq)$ by a pointed model $(\mathcal{M}, s)$ is inductively defined as follows:

$$\begin{array}{rcl} \mathcal{M}, s \models p & \iff & w \in \mathcal{I}(p) \\ \mathcal{M}, s \models [\preceq]\varphi & \iff & \forall s' \in S \text{ s.t. } s \preceq s' : \mathcal{M}, s' \models \varphi \\ \mathcal{M}, s \models [\forall]\,\varphi & \iff & \forall s' \in S : \mathcal{M}, s' \models \varphi \end{array}$$

The standard Boolean clauses are omitted.

Axiomatization

The logic is axiomatized as follows, where $i \in \{\preceq, \forall\}$:

(Prop)	propositional tautologies
(K)	$[i](\varphi_1 \to \varphi_2) \to ([i]\varphi_1 \to [i]\varphi_2)$
(T)	$[i]\varphi \to \varphi$
(4)	$[i]\varphi \to [i][i]\varphi$
(5)	$\neg[\forall]\varphi \to [\forall]\neg[\forall]\varphi$
(.3)	$((\langle\preceq\rangle\varphi \wedge \langle\preceq\rangle\psi) \to ((\langle\preceq\rangle(\varphi \wedge \langle\preceq\rangle\psi) \vee \langle\preceq\rangle(\varphi \vee \psi) \vee \langle\preceq\rangle(\psi \wedge$
(Incl)	$[\forall]\varphi \to [\preceq]\varphi$
(Dual)	$\langle i\rangle\varphi \leftrightarrow \neg[i]\neg\varphi$

The logic is an extension of **S4.3** with the universal modality.

Completeness

Theorem (Strong completeness)

The logic above is sound and strongly complete with respect to the class of total pre-orders.

Defining 'best'

Our logic is quite expressive. The very first semantics for dyadic deontic logic ([Hansson, 1969]) interpreted formulae **O**$(\varphi \mid \psi)$ as "all the best ψ-states are φ" . Within our logic, a maximality operator can be defined as follows:

$$[\mathrm{Best}(\psi)]\,\varphi \quad := \quad [\forall]\,(\psi \to \langle\preceq\rangle(\psi \wedge [\preceq](\psi \to \varphi)) \tag{5}$$

That is, the best ψ-states are φ if and only if, for all states, either they are not ψ or there is a better ψ-state such that all states above it are either not ψ or φ.

Example (Gentle murder (continued))

Consider the P-sequence for valuation $\mathcal{I}$ of the Gentle murder introduced above, and let $\preceq_{\mathcal{B}}^{IM}$ be the total pre-order generated by that sequence. We have that, for any state s in the model $\mathcal{M}_{\mathcal{B}}^{IM} = (S, \preceq_{\mathcal{B}}^{IM}, \mathcal{I})$:

$$\begin{array}{rcl} \mathcal{M}_{\mathcal{B}}^{IM}, s & \models & [\mathrm{Best}(\top)]\ \neg \mathrm{V}_1 \\ \mathcal{M}_{\mathcal{B}}^{IM}, s & \models & [\mathrm{Best}(\mathrm{V}_1)]\ \mathrm{I}_2 \\ \mathcal{M}_{\mathcal{B}}^{IM}, s & \models & [\mathrm{Best}(\mathrm{V}_2)]\ \mathrm{V}_2. \end{array}$$

From Static to Dynamics

Various Kinds of Dynamics

- Changing information, conditional obligation, betterness order stays the same.
- Changing evaluation of worlds locally, maybe by a command(see Yamada 2006, 2007, etc.), at betterness change level.
- Changing the whole norm system: at level of changing the priority structure.

Two Level Dynamics

In the current framework, we can handle dynamical changes that are located both at the level of P-sequences or on the level of possible worlds:

- at the level of P-sequence, we can think of dynamic actions of *adding* a priority, *deleting* a priority, etc.
- on the level of possible worlds, there are various operations which change the ordering over possible worlds.

The operations that were considered in [Liu, 2008] for those two kinds of dynamics apply naturally here.

For Instance: Upgrade on the level of possible worlds

Definition (Upgrade)

The *(radical) upgrade* operation $\Uparrow\varphi$ is defined as:

$$(?\varphi; \preceq; ?\varphi) \cup (?\neg\varphi; \preceq; ?\neg\varphi) \cup (?\neg\varphi; \top; ?\varphi).$$

where $?$ and $;$ are the standard relational operations of test and sequencing, and $\top$ denotes the universal relation.

After the new proposition φ has been incorporated, the upgrade places all φ-worlds on top of all $\neg\varphi$-worlds, keeping all other comparisons the same. Here, besides cutting links between the φ-worlds and $\neg\varphi$-worlds, new betterness links may be added by the disjunct $(?\neg\varphi; \top; ?\varphi)$.

Connecting the Two Levels

Definition

Let F: $(\mathcal{B}, \varphi) \rightarrow \mathcal{B}'$, with $\mathcal{B}$, $\mathcal{B}'$ two P-sequences, and φ a new formula not occurring in $\mathcal{B}$. Let σ: $(\preceq, \varphi) \rightarrow \preceq'$, where $\preceq$ and $\preceq'$ are betterness relations over possible worlds. We say that F ***induces the map*** σ, given a definition for deriving betterness relations from P-sequence (e.g., Definition 2), if, for any P-sequence $\mathcal{B}$ and new formula φ, $\sigma(\preceq_{\mathcal{B}}, \varphi) = \preceq_{F(\mathcal{B},\varphi)}$.

Real Example

Example (Obama and global climate change)

With the issue of global climate change, there were two alternative approaches the Obama administration could take to regulate greenhouse gases, like carbon dioxide emissions from coal plants. One was to craft regulations under existing legal authority, say, by the Clean Air Act. The other was to work with Congress on the enactment of legislation to address climate change. The Obama administration faced great difficulties when working with Congress, so they shifted to the first alternative. On April 17, 2009, the Environmental Protection Agency(EPA) officially designated carbon dioxide and five other heat-trapping gases to be dangerous pollutants.
——Recomposed from *The New York Times* (April 17, 2009)

Analysis

The two alternative approaches reflects two options of making changes:

- Enacting a new legislation. It is tantamount to adding a new proposition to the top of the existing system of norms (the given P-sequence).
- Making use of the existing legal authority. This can be thought of one way of changing the underlying betterness ordering.

Theorem (Correspondence of two-level dynamics)

Given a betterness relation $\preceq$ derived from $\mathcal{B} = (B, \prec)$ by Definition 2, action $\Uparrow\varphi$ on $\preceq$ is induced by prefixing the P-sequence $\mathcal{B}$ with a new formula φ. More precisely, the following diagram commutes:

$$
\begin{tikzcd}
\mathcal{B} \arrow[r, "\varphi;\mathcal{B}"] \arrow[d, "IM"'] & \varphi;\mathcal{B} \arrow[d, "IM"] \\
(S, \preceq^{IM}_{\mathcal{B}}) \arrow[r, "\Uparrow\varphi"] & (S, \Uparrow\varphi(\preceq^{IM}_{\mathcal{B}}))
\end{tikzcd}
$$

Proof

Proof.

We need to prove the following equivalence:
$s \preceq^{IM}_{A;\mathcal{B}} s'$ iff $s, \Uparrow A(\preceq^{IM}_{\mathcal{B}})\, s'$. [$\Leftarrow$]. After $\Uparrow A$ the relation between s and s' becomes this:
$\Uparrow \varphi(\preceq^{IM}_{\mathcal{B}}) := (?\varphi; \preceq^{IM}_{\mathcal{B}}; ?\varphi) \cup (?\neg\varphi; \preceq^{IM}_{\mathcal{B}}; ?\neg\varphi) \cup (?\neg\varphi; \top; ?\varphi)$. In terms of a relation between arbitrary worlds s' and s, the above three cases give the implication $s \in ||\varphi|| \to s' \in ||\varphi||$. By $s \preceq^{IM}_{\mathcal{B}} s'$, we also have that $\forall\psi\in\mathcal{B}$: $s\in||\psi||\to w'\in||\psi||$. Hence $\forall\psi\in\varphi;\mathcal{B}$: $s\in||\psi||\to s'\in||\psi||$: i.e., $s\preceq^{IM}_{\varphi;\mathcal{B}} s'$. [$\Rightarrow$]. Assume that $s\preceq^{IM}_{\varphi;\mathcal{B}} s'$, i.e., $\forall\psi\in\varphi;\mathcal{B}$: $s\in||B\psi|| \to s'\in||\psi||$. In particular, it cannot the case that $s\in||\varphi|| \wedge s'\notin||\varphi||$. Thus, out of all pairs in the given relation R, those satisfying $(?\varphi; \top; ?\neg\varphi)$ can no longer occur. This is precisely how we defined the relation $s \Uparrow\varphi(\preceq^{IM}_{\mathcal{B}})\, s'$. $\square$

Connection to Norm Change

The dynamic aspects of norms—The *norm change* problem—have recently gained much attention from researchers in deontic logic, legal theory and multi-agent systems.

Two main approaches:

- In syntactic approaches—inspired by legal practice—norm change is an operation performed directly on the explicit provisions in the “code" of the normative system ([Governatori and Rotolo, 2008],[Boella, Pigozzi and van der Torre, 2009]).
- In semantic approaches, norm change follows the dynamic logic update paradigm (e.g.[Aucher, etc., 2009]).

Connection to Norm Change

The above Theorem can be viewed as establishing a precise match between changes at the level of **models** with changes at the level of **syntax** of a normative code, i.e., the P-sequences.

Future Work

- Our analysis has focused on linearly ordered P-sequences and the induced total pre-orders—fitting for CTDs. We would like to look at the case of pre-orders, generalizing the logical machinery presented here to that case.
- Future work will look at more elaborate representations of the syntactic side of our approach, including graph structures of criteria and laws, maintaining a normative syntax-semantics correspondence along the lines of the correspondence Theorem.
- Obligations are often entangled with beliefs in reality. We would like to extend our basic setting with beliefs, and explore the belief dynamics and norm change all together.

References

- G.Aucher, D.Grossi, A. Herzig and E.Lorini. Dynamic context logic. In X.He, J. Horty and E.Pacuit, editors, *Proceedings of LORI 2009*, volume 5834 of LNAI. Springer, 2009.
- G. Boella, G. Pigozzi, and L. van der Torre. Normative framework for normative system change. In Decker P., Sichman J., Sierra C., and Castelfranchi C., editors, *Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2009), 2009.
- C. Boutilier. Conditional logics of normality as modal systems. In *Proceedings of AAAI'90*, pages 594-599, 1990.
- C. Boutilier. Conditional logics of normality: A modal approach. *Artificial Intelligence*, 68:87-154, 1994.

References

- J. Forrester. Gentle murder, or the adverbial samaritan. *Journal of Philosophy*, 81:193-197, 1984.
- P. Girard. *Modal Logic for Belief and Preference Change*. PhD thesis, University of Amsterdam, 2008.
- G. Governatori and A. Rotolo. Changing legal systems: Abrogation and annulment. part 1: Revision and defeasible theories. In Ron van der Meyden and Leendert van der Torre, editors, *Proceedings of the 9th International Conference on Deontic Logic in Computer Science* (DEON 2008), Luxembourg, Luxembourg, July 15-18, 2008, volume 5076 of LNAI, pages 3-18. Springer, 2008.

References

- B. Hansson. An analysis of some deontic logics. *Nous*, 3:373-398, 1969.
- F. Liu. *Changing for the Better. Preferemce dynamics and Agent Diversity*. PhD thesis, University of Amsterdam, 2008.
- G. E. Moore. *Principia Ethica*. Cambridge University Press, 1903.
- H. Prakken and M. Sergot. Contrary-to-duty obligations. *Studia Logica*, 57:91-115, 1996.
- O.Roy. *Thinking before Acting: Intentions, Logic and Rational Choice*,ILLC, University of Amsterdam, 2008.
- T. Yamada. Acts of Commands and Changing Obligations, Inoue,K. and Satoh, K. and Toni, F. editors, *Proceedings of the 7th Workshop on Computational Logic in Multi-Agent Systems* (CLIMA VII)", 2006. Revised version appeared in LNAI 4371, pages 1-19, Springer-Verlag, 2007

Scorekeeping and Dynamic Logics of Speech Acts

Tomoyuki Yamada

Research Group of Philosophy
Hokkaido University

28 March 2010, Hokkaido University, Sapporo

北海道大学 HOKKAIDO UNIVERSITY

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

Outline

北海道大学 HOKKAIDO UNIVERSITY

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

The gap

Van Benthem & Liu (2007) on commanding

For instance, intuitively, a command

“See to it that φ!”

makes worlds where φ holds preferred over those where it does not - at least, if we accept the preference induced by the issuer of the command.

The need they felt for the proviso here reflects an important logical gap between what an illocutionary act of commanding involves and perlocutionary effects it may have upon our preferences.

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

Austin’s Distinction (1955, pp.101-3.)

Locutionary Act

He said to me “Shoot her!” meaning by ‘shoot’ shoot and referring by ‘her’ to her.

Illocutionary Act

He urged (advised, ordered, etc.) me to shoot her.

Perlocutionary Act

(a) He persuaded me to shoot her.
(b) He got me to shoot her.

北海道大学 HOKKAIDO UNIVERSITY

Speech acs as acts

- If the notion of speech act is to be taken seriously, it must be possible to treat speech acts as acts.
- If we succeed in characterizing speech acts in terms of dynamic changes they bring about, it becomes possible to treat them within a general theory of action.
- But how can we do that?

Perlocutinary acts as acts

Perlocutionary Act

(a) He persuaded me to shoot her.
(b) He got me to shoot her.

Austin on perlocutionary acts (1955, p.103)

According to Austin, perlocutionary acts are acts that really produce “real effects” upon the feelings, thoughts, or actions of addressees, or of speakers, or of other people.

They are recognized only when their effects are recognized.

Illocutinary acts as acts

Illocutionary Act

He urged (advised, ordered, etc.) me to shoot her.

The Problem

What effects do they have?
What role do they play in our social life?

Austin, Strawson, and Searle

Austin on illocutionary acts (1955, p.103)

Austin considered illocutionary acts as acts whose effects are “what we regard as mere conventional consequences”

After Strawson (1964) and Seale (1969)

Austin’s conception of illocutionary acts as acts whose effects are conventional has been disregarded both by those who follow Strawson and those who follow Searle.

Strawson (1964) on Austin

- Strawson (1964) observed that the kind of conventional effects involved in the examples used by Austin are dependent on special extralinguistic conventions.
- He then argued that there are many other illocutionary acts that do not seem to be dependent on any such special extralinguistic conventions.
- Thus, according to Strawson, Austin made an unwarranted overgeneralization when he attributed conventional effects to illocutionary acts in general.

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

Conventional effects vs. utterers' intentions

- Strawson and his followers tried to characterize uses of sentences not in terms of conventional effects, but in terms of utterers' intentions to produce various effects in addressees along the lines initiated by Grice (1957).
- Utterers' intentions, however, usually go beyond illocutionary acts by involving reference to perlocutionary effects, while illocutionary acts can be effective even if they failed to produce intended perlocutionary effects.

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

Searle (1969) on Austin and Strawson

- Searle criticized Grice (and Strawson) for treating meaning as "a matter of intending to perform a perlocutionary acts",
- but agreed with Strawson in seeing Austin's notion of conventional effect as an unwarranted overgeneralization.
- Searle sees conventionality of illocutionary acts as a matter of meaning, and denied the distiction between locutionary acts and illocutionary acts.
- He identified what he called "the illocutionary effect" with "the hearer understanding the utterance of the speaker" (p.46-47).

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

Beyond the securing of uptake

- Austin considered the sequring of uptake of this kind as necessary condition for illocutionary acts, but didn't considered it to be sufficient.
- Indeed, even typical illocutionary acts such as acts of promising, which both Strawson and Searle see not conventional in what they take to be Austin's sense, seem to involve more than the mere securing of uptake.
- The social or institutional consequences they have, such as generation of obligations, can be said to be "conventional" in Austin's sense.
- They are institutional in Searle's sense.

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

What Austin's Earlier Answer Enables us to See

Perlocutionary acts

Since perlocutionary acts are acts that really produce real effects, they cannot be completed without really producing them.

Illocutionary acts

Illocutionary acts are completed when the "mere conventional" effects are produced.

Austin 1955, pp.103-4.

Thus Austin says, "we can say 'I argue that' or 'I warn you that' but we cannot say 'I convince you that' or 'I alarm you that" .

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

The problem

- Is it possible to develop this conception of illocutionary acts into a general theory of illocutionary acts?
- In order to do so, we have to
  - specify conventional effects of a sufficiently rich variety of illocutionary acts, and
  - develop a theory in which these illocutionary acts are shown to be fully characterized in terms of those conventional effects.

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

The plan

- The recent development of Dynamic Epistemic Logics suggests a recipe for developing logics that can capture effects of various speech acts.
- We have developed dynamic logics that can deal with acts of commanding, promising, asserting, conceding, and withdrawing according to this recipe (Yamada 07a, 07b, 08a, 08b, unfinished draft).
- We will review these developmets .
- We will then show how the results obtained can be incorporated into a more comprehensive picture of social interaction with the help of the notion of scorekeeping for language games.

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

The developments of dynamic epistemic logics

$[\varphi!]K_i\psi$

Dynamic Epistemic Logics DEL

adding dynamic modalities ↑ ↓ translation along reduction axioms

Multi-agent Epistemic Logics EL

$K_i\varphi$

Cf. Plaza (1989), Gerbrandy & Groeneveld (1997), Gerbrandy (1999), Baltag, Moss, & Solecki (1999), Kooi & van Benthem (2004), van Ditmarsch, Kooi, and van der Hoek (2007)

北海道大学 HOKKAIDO UNIVERSITY

Two points to be noted

The formulas of the form $\varphi \to [\varphi!]K_i\varphi$ are shown to be valid for any $i \in I$ if no operators of the form K_i occur in φ.

- This is too strong for interpreting natural language public announcements.
- A gap similar to the one we have seen is also present here.

The method used in developing *DEL* can be used to develop logics that deal with a much wider variety of speech acts.

北海道大学 HOKKAIDO UNIVERSITY

The recipe (Yamada, to appear)

1. Carefully identify the aspect affected by the kind of speech acts you want to study
2. find the modal logic that characterizes this aspect
3. add dynamic modalities that represent types of those speech acts
4. define model updating operation that interprets the speech acts under study as what update the very aspect
5. (if possible) find a complete set of reduction axioms for the resulting dynamic logic.

北海道大学 HOKKAIDO UNIVERSITY

This recipe works for acts of commanding (Yamada, 2007a)

$[!_i\varphi]O_i\psi$

Eliminative Command Logic ECL

adding dynamic modalities ↑ ↓ translation along reduction axioms

Multi-agent Deontic Logic MDL$^+$

$O_i\varphi$

北海道大学 HOKKAIDO UNIVERSITY

1 & 2. Identifying the relevant aspect and its logic

$[!_i\varphi]O_i\psi$

Eliminative Command Logic ECL

adding dynamic modalities ↑ ↓ translation along reduction axioms

Multi-agent Deontic Logic MDL$^+$

$O_i\varphi$

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

The language of multi-agent deontic logic

Definition

Take a countably infinite set Aprop of proposition letters and a finite set I of agents, with p ranging over Aprop and i over I. The multi-agent monadic deontic language $\mathcal{L}_{\mathrm{MDL}^+}$ is given by:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi\wedge\psi \mid \Box\varphi \mid O_i\varphi$$

$O_a\varphi$ It is obligatory upon an agent a to see to it that φ.

$P_a\varphi$ $\neg O_a\neg\varphi$.

$F_a\varphi$ $O_a\neg\varphi$.

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

$\mathcal{L}_{\mathrm{MDL}^+}$-models

Definition

By an $\mathcal{L}_{\mathrm{MDL}^+}$-model, we mean a tuple $M = \langle W^M, \Rightarrow^M, \{\smile_i^M \mid i\in I\}, V^M\rangle$ where:

(i) W^M is a non-empty set (heuristically, of 'possible worlds'),

(ii) $\Rightarrow^M \subseteq W^M \times W^M$,

(iii) $\smile_i^M \subseteq \Rightarrow^M$ for each $i \in I$,

(iv) V^M is a function that assigns a subset $V^M(p)$ of W^M to each proposition letter $p \in Aprop$.

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

Truth definition for $\mathcal{L}_{\mathrm{MDL}^+}$

Definition

Let M be an $\mathcal{L}_{\mathrm{MDL}^+}$-model and w a point in M. If $p \in Aprop$, and $i \in I$, then:

(a) $M, w \models_{\mathrm{MDL}^+} p$ iff $w \in V^M(p)$

(b) $M, w \models_{\mathrm{MDL}^+} \top$

(c) $M, w \models_{\mathrm{MDL}^+} \neg\varphi$ iff it is not the case that $M, w \models_{\mathrm{MDL}^+} \varphi$

(d) $M, w \models_{\mathrm{MDL}^+} (\varphi\wedge\psi)$ iff $M, w \models_{\mathrm{MDL}^+} \varphi$ and $M, w \models_{\mathrm{MDL}^+} \psi$

(to be continued)

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Truth definition for $\mathcal{L}_{\text{MDL}^+}$ (continued)

$$\text{(e)}\quad M, w \models_{\text{MDL}^+} \Box\varphi \quad \text{iff} \quad \text{for every } v \text{ such that } \langle w, v\rangle \in \Rightarrow^M,\ M, v \models_{\text{MDL}^+} \varphi$$

$$\text{(f)}\quad M, w \models_{\text{MDL}^+} O_i\varphi \quad \text{iff} \quad \text{for every } v \text{ such that } \langle w, v\rangle \in \smile_i^M,\ M, v \models_{\text{MDL}^+} \varphi$$

A formula φ is true in an $\mathcal{L}_{\text{MDL}^+}$-model M at a point w of M if $M, w \models_{\text{MDL}^+} \varphi$. The semantic consequence relation and the notion of validity can also be defined in the standard way.

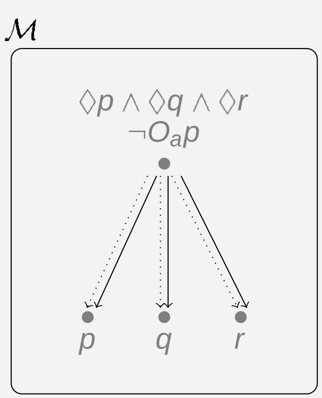
The proof system for MDL$^+$

Definition

The proof system for MDL$^+$ includes (i) all instantiations of propositional tautologies over the present language, (ii) K-axioms for alethic modality and O_i-modality for each $i \in I$, (iii) modus ponens, and (iv) necessitation rules for alethic modality and O_i-modality for each $i \in I$, in addition to the axiom of the following form for each $i \in I$:

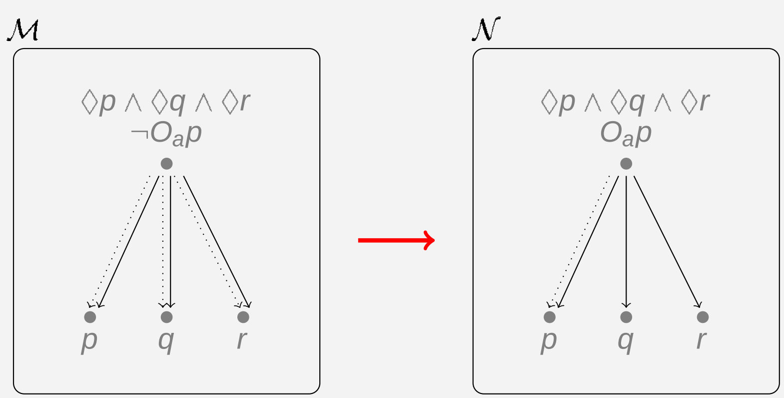
$$\text{(Mix)}\quad P_i\varphi \to \Diamond\varphi$$

Example 1: on a hot day in a shared office



- p The window is open.
- q The air conditioner is running.
- r The temperature is rising.

Your boss's act of commanding in MDL$^+$



3 & 4. Dynamic Extension

$[!_i\varphi]O_i\psi$

Eliminative Command Logic ECL

adding dynamic modalities ↑ ↓ translation along reduction axioms

Multi-agent Deontic Logic MDL$^+$

$O_i\varphi$

The language of command logic

Definition

Take the same countably infinite set *Aprop* of proposition letters and the same finite set *I* of agents as before, with *p* ranging over *Aprop*, and *i* over *I*. The language $\mathcal{L}_{\mathrm{ECL}}$ of eliminative command logic ECL is given by:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi \mid O_i\varphi \mid [\pi]\varphi$$
$$\pi ::= !_i\varphi$$

$[!_a\psi]O_a\varphi$ After every effective act of commanding an agent *a* to see to it that ψ, it is obligatory upon *a* to see to it that φ.

The truth definition for $\mathcal{L}_{\mathrm{ECL}}$

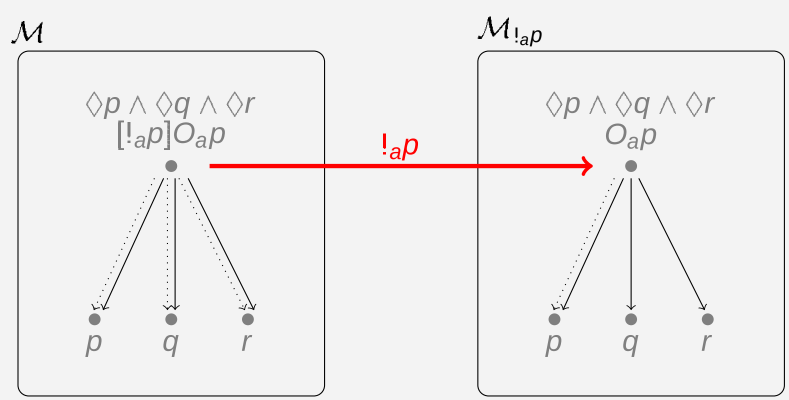
Definition

Let *M* be an $\mathcal{L}_{\mathrm{MDL^+}}$-model and *w* a point in *M*. If $p \in Aprop$, and $i \in I$, then the truth definition for $\mathcal{L}_{\mathrm{ECL}}$ is given by expanding that of $\mathcal{L}_{\mathrm{MDL^+}}$ mutatis mutandis with the following new clause:

$$\text{(g)} \quad M, w \models_{\mathrm{ECL}} [!_i\chi]\varphi \text{ iff } M_{!_i\chi}, w \models_{\mathrm{ECL}} \varphi \ ,$$

where $M_{!_i\chi}$ is the $\mathcal{L}_{\mathrm{MDL^+}}$-model obtained from *M* by replacing $\smile_i^M$ with $\{\langle x, y\rangle \in \smile_i^M \mid M, y \models_{\mathrm{ECL}} \chi\}$.

Your boss's act of commanding in ECL



Some interesting principles

CUGO Ptrinciple

If φ is a formula of $\mathcal{L}_{\mathrm{MDL}^+}$ and is free of occurrences of modal formulas of the form O_i, then $[!_i\varphi]O_i\varphi$ is valid.

Dead End Principles

$[!_i(\varphi \wedge \neg\varphi)]O_i\psi$ is valid.

Restricted Sequential Conjunction

If φ and ψ are formulas of $\mathcal{L}_{\mathrm{MDL}^+}$ and are free of occurrences of modal formulas of the form O_i, then $[!_i\varphi][!_i\psi]\chi \leftrightarrow [!_i(\varphi \wedge \psi)]\chi$ is valid.

5. Finding reduction axioms

$[!_i\varphi]O_i\psi$

Eliminative Command Logic ECL

adding dynamic modalities ↑

↓ translation along reduction axioms

Multi-agent Deontic Logic MDL$^+$

$O_i\varphi$

The proof system for ECL

Definition

The proof system for ECL includes all the axioms and all the rules of the proof system for MDL$^+$, and in addition, the following rule and axioms:

(!-nec) $\dfrac{\psi}{[!_i\varphi]\psi}$ (for each $i \in I$)

(To be continued)

The proof system for ECL (continued)

Continued

(!1)	$[!_i\varphi]p$	$\leftrightarrow$	p	
(!2)	$[!_i\varphi]\top$	$\leftrightarrow$	$\top$	
(!3)	$[!_i\varphi]\neg\psi$	$\leftrightarrow$	$\neg[!_i\varphi]\psi$	
(!4)	$[!_i\varphi](\psi \wedge \chi)$	$\leftrightarrow$	$[!_i\varphi]\psi \wedge [!_i\varphi]\chi$	
(!5)	$[!_i\varphi]\Box\psi$	$\leftrightarrow$	$\Box[!_i\varphi]\psi$	
(!6)	$[!_i\varphi]O_j\psi$	$\leftrightarrow$	$O_j[!_i\varphi]\psi$	$(i \neq j)$
(!7)	$[!_i\varphi]O_i\psi$	$\leftrightarrow$	$O_i(\varphi \rightarrow [!_i\varphi]\psi)$	

Translation from $\mathcal{L}_{\text{ECL}}$ to $\mathcal{L}_{\text{MDL}^+}$

Definition

$$
\begin{array}{rlrl}
t(p) = & p & t([!_i\varphi]p) = & p \\
t(\top) = & \top & t([!_i\varphi]\top) = & \top \\
t(\neg\varphi) = & \neg t(\varphi) & t([!_i\varphi]\neg\psi) = & \neg t([!_i\varphi]\psi) \\
t(\varphi \wedge \psi) = & t(\varphi) \wedge t(\psi) & t([!_i\varphi](\psi \wedge \chi)) = & t([!_i\varphi]\psi) \wedge t([!_i\varphi]\chi) \\
t(\Box\varphi) = & \Box t(\varphi) & t([!_i\varphi]\Box\psi) = & \Box t([!_i\varphi]\psi) \\
t(O_i\varphi) = & O_i t(\varphi) & t([!_i\varphi]O_j\psi) = & O_j t([!_i\varphi]\psi) \quad (i \neq j) \\
 & & t([!_i\varphi]O_i\psi) = & O_i(t(\varphi) \to t([!_i\varphi]\psi)) \\
 & & t([!_i\varphi][!_j\psi]\chi) = & t([!_i\varphi]t([!_j\psi]\chi)) \\
 & & & (\text{for any } j \in I)
\end{array}
$$

Completeness of ECL

Corollary

If $\xi \in \mathcal{L}_{\text{MDL}^+}$, then $M, w \models_{\text{MDL}^+} \xi$ iff $M, w \models_{\text{ECL}} \xi$.

Lemmas

For any $\eta \in \mathcal{L}_{\text{ECL}}$, $t(\eta) \in \mathcal{L}_{\text{MDL}^+}$.
For any $\eta \in \mathcal{L}_{\text{ECL}}$, $M, w \models_{\text{ECL}} \eta$ iff $M, w \models_{\text{ECL}} t(\eta)$.
For any $\eta \in \mathcal{L}_{\text{ECL}}$, $\vdash_{\text{ECL}} \eta \leftrightarrow t(\eta)$.

Theorem

ECL *is strongly complete with respect to* MDL$^+$*-models.*

Proof of the weak completeness of ECL

completeness of ECL

$$
\begin{array}{llcll}
 & \models_{\text{ECL}} \eta & & \vdash_{\text{ECL}} \eta & \\
\text{lemma} & \quad\downarrow & & \quad\uparrow & \vdash_{\text{ECL}} t(\eta) \leftrightarrow \eta \\
 & \models_{\text{ECL}} t(\eta) & & \vdash_{\text{ECL}} t(\eta) & \\
\text{corollary} & \quad\downarrow & & \quad\uparrow & \text{MDL}^+ \subseteq \text{ECL} \\
 & \models_{\text{MDL}^+} t(\eta) & \longrightarrow & \vdash_{\text{MDL}^+} t(\eta) &
\end{array}
$$

completeness of MDL$^+$

A refinement (Yamada 2007b)

$$\begin{tikzcd} \text{ECL} & \text{ECLII} \\ \text{MDL}^{+} \arrow[u, "\text{dynamification}"] \arrow[r, "\text{refinement}"'] & \text{MDL}^{+}\text{II} \arrow[u, "\text{dynamification}"'] \end{tikzcd}$$

$O_{(i,j)}\varphi$ It is obligatory upon an agent i with respect to an authority j to see to it that φ.

$[!_{(i,j)}\psi]\chi$ After an authority j commands an agent i to see to it that ψ, χ holds.

Contradictory commands from two distinct authorities

A dilemma

$[!_{(a,b)}p][!_{(a,c)}\neg p](O_{(a,b)}p \wedge O_{(a,c)}\neg p)$.

Note that this does not lead to deontic explosion.

Example 2: Conflicting commands from your boss and your guru

A contingent dilemma

$[!_{(a,b)}p][!_{(a,c)}q](O_{(a,b)}p \wedge O_{(a,c)}q) \wedge \neg(p \wedge q)$.

p You will attend the conference in São Paulo on 11 June 2010.

q You will join the demonstration in Sapporo on 11 June 2010.

Some results (Yamada, 2007b)

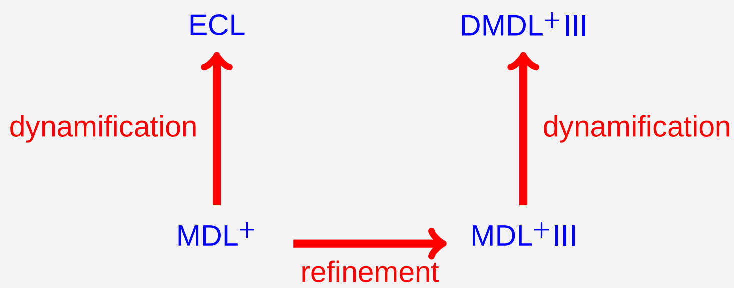
CUGO Principle

If φ is a formula of MDL^{+}II and is free of modal operators of the form $O_{(i,j)}$, $[!_{(i,j)}\varphi]O_{(i,j)}\varphi$ is valid.

Theorem

There is a complete axiomatization of ECLII.

A further refinement and extension (Yamada 2008a)



$O_{(i,j,k)}\varphi$ It is obligatory upon an agent i with respect to an obligee j in the name of k to see to it that φ.

$Com_{(i,j)}\varphi$ Act of commanding.

$Prom_{(i,j)}\varphi$ Act of promising.

Example 3: a command and a promise can lead to a dilemma

A contingent dilemma

$[Prom_{(a,b)}p][Com_{(c,a)}q](O_{(a,b,a)}p \wedge O_{(a,c,c)}q) \wedge \neg(p \wedge q)$.

p You will attend the conference in São Paulo on 11 June 2010.

q You will join the demonstration in Sapporo on 11 June 2010.

Some results (Yamada, 2008a)

CUGO Principle

If φ is a formula of MDL$^+$III and is free of modal operators of the form $O_{(j,i,i)}$, $[Com_{(i,j)}\varphi]O_{(j,i,i)}\varphi$ is valid.

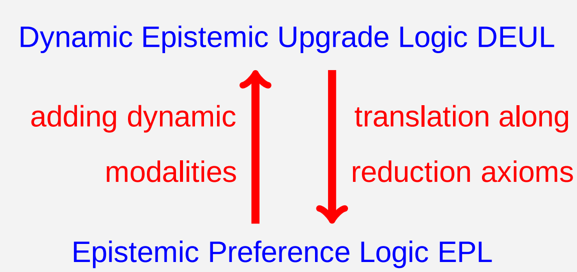
PUGO Principle

If φ is a formula of MDL$^+$III and is free of modal operators of the form $O_{(i,j,i)}$, $[Prom_{(i,j)}\varphi]O_{(i,j,i)}\varphi$ is valid.

Theorem

There is a complete axiomatization of DMDL$^+$III.

The same strategy works for changing preferences (van Benthem and Liu, 2007) (Liu, 2008)



The language of EPL

Definition

Take a set *Aprop* of proposition letters, and a set I of agents, with p ranging over *Aprop* and i over I. The epistemic preference language is given by:

$$\varphi ::= \bot \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid U\varphi \mid K_i\varphi \mid [pref]_i\varphi$$

Intuitively, $[pref]_i\varphi$ means that all worlds i considers at least as good as the current one satisfy φ.
U is the so-called “universal modality”, and $U\varphi$ means that φ holds at every world.

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

The language of DEUL

Definition

Take the same set *Aprop* of proposition letters, and the same set I of agents as before, with p ranging over *Aprop* and i over I. The dynamic epistemic preference language is given by:

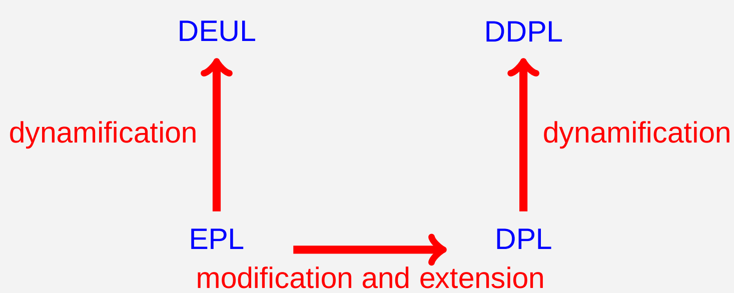
$$\begin{array}{lll} \varphi & ::= & \bot \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid U\varphi \mid K_i\varphi \mid [pref]_i\varphi \mid [\pi]\varphi \\ \pi & ::= & \varphi! \mid \sharp\varphi \end{array}$$

$\varphi!$: the type of acts of publicly announcing that φ,
$\sharp\varphi$: the type of acts of publicly suggesting φ.

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

Combining preference upgrades and deontic updates (Yamada 2008b)



Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

The language of DPL

Definition

Take a set *Aprop* of proposition letters, and a set I of agents, with p ranging over *Aprop* and i, j over I. The deontic preference language is given by:

$$\varphi ::= \bot \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid U\varphi \mid [pref]_i\varphi \mid O_{(i,j)}\varphi$$

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

The language of DDPL

Definition

Take a set *Aprop* of proposition letters, and a set I of agents, with p ranging over *Aprop* and i, j over I. The dynamic deontic preference language is given by:

$$\varphi \quad ::= \quad \bot \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid U\varphi \mid [pref]_i\varphi \mid O_{(i,j)}\varphi \mid [\pi]\varphi$$

$$\pi \quad ::= \quad \sharp_i\varphi \mid !_{(i,j)}\varphi$$

Some results (Yamada, 2008b)

Theorem

There is a complete axiomatization of DDPL.

The following formulas are satisfiable.

$O_{(i,j)}p \wedge U(p \to \langle pref\rangle_i\neg p)$.
$[!_{(i,j)}p]U(p \to \langle pref\rangle_i\neg p)$.

$\langle pref\rangle_i\varphi$ is an abbreviation of $\neg[pref]_i\neg\varphi$.

The same recipe works for acts of asserting and conceding (Yamada, Unfinished Draft)

Dynamified Multiagent Propositional Commitment Logic
DMPCL

adding dynamic modalities ↑ ↓ translation along reduction axioms

MPCL
Multi-agent Propositional Commitment Logic

Walton & Krabbe (1995)

Three Kinds of propositional commitments

- commitments incurred by making concessions
- commitments called assertions
- participant's dark-side commitments

Since dark-side commitments are hidden commitments and supposed to be fixed, we will ignore them.

We call the remaining two kinds of commitments c-commitments and a-commitments respectively.

A-commitments and c-commitments

According to Walton and Krabbe (1995, p.186)

Propositional commitments constitute a special case of commitments to a course of action.

- an agent who has an a-commitment to the proposition p is obliged to defend it if the other party in the dialogue require her to justify it
- an agent who has a c-commitments to p is only obliged to allow the other party to use it in the arguments.

As anyone who asserts that p will be obliged to allow the other party to use it in the arguments, a-commitments imply c-commitments.

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

The language of MPCL

Definition

Take a countably infinite set *Aprop* of proposition letters, and a finite set I of agents, with p ranging over *Aprop*, and i over I. The language $\mathcal{L}_{\text{MPCL}}$ of the multi-agent propositional commitment logic MPCL is given by:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid [a\text{-}cmt]_i\varphi \mid [c\text{-}cmt]_i\varphi$$

$[\text{a-cmt}]_i\varphi$: an agent i has an a-commitment to the proposition φ,
$[\text{c-cmt}]_i\varphi$: an agent i has a c-commitment to the proposition φ.

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

P-commitments are different from knowledge

The following formulas are not valid.

$[\text{a-cmt}]_i\varphi \rightarrow \varphi$

$[\text{c-cmt}]_i\varphi \rightarrow \varphi$

Cf. $K_i\varphi \rightarrow \varphi$

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

P-commitments are different from belief

The following formulas are not valid.

$\neg[\text{a-cmt}]_i\bot$

$\neg[\text{c-cmt}]_i\bot$

Cf. $\neg B_i\bot$

$\mathcal{L}_{\text{MPCL}}$-models

Definition

By an $\mathcal{L}_{\text{MPCL}}$-model, we mean a tuple $M = \langle W^M, \{\triangleright_i^M \mid i \in I\}, \{\blacktriangleright_i^M \mid i \in I\}, V^M\rangle$ where:

(i) W^M is a non-empty set (heuristically, of 'possible worlds'),

(ii) $\triangleright_i^M \subseteq W^M \times W^M$ for each $i \in I$,

(iii) $\blacktriangleright_i^M \subseteq \triangleright_i^M$ for each $i \in I$,

(iv) V^M is a function that assigns a subset $V^M(p)$ of W^M to each proposition letter $p \in \mathrm{Aprop}$.

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

Truth definition for $\mathcal{L}_{\text{MPCL}}$ (crucial part)

In addition to the standard clauses for proposition letters and Boolean operations,

(e) $M, w \models_{\text{MPCL}} [a\text{-}cmt]_i\varphi$ iff for every v such that $\langle w, v\rangle \in \triangleright_i^M$, $M, v \models_{\text{MPCL}} \varphi$

(f) $M, w \models_{\text{MPCL}} [c\text{-}cmt]_i\varphi$ iff for every v such that $\langle w, v\rangle \in \blacktriangleright_i^M$, $M, v \models_{\text{MPCL}} \varphi$

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

The Proof system for MPCL

Definition

The proof system for MPCL includes (i) all instantiations of propositional tautologies over the present language, (ii) K-axioms for $[\text{a-cmt}]_i$-modality and $[\text{c-cmt}]_i$-modality for each $i \in I$, (iii) modus ponens, and (iv) necessitation rules for $[\text{a-cmt}]_i$-modality and $[\text{c-cmt}]_i$-modality for each $i \in I$, in addition to the axiom of the following form for each $i \in I$:

$$\text{(Mix)} \quad [a\text{-}cmt]_i\varphi \to [c\text{-}cmt]_i\varphi$$

Theorem (Completeness of MPCL)

MPCL *is strongly complete with respect to* $\mathcal{L}_{\text{MPCL}}$*-models.*

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

Closure

Propositional commitments are closed with respect to the logical consequence.

$$([a\text{-}cmt]_i\varphi \wedge [a\text{-}cmt]_i(\varphi \to \psi)) \to [a\text{-}cmt]_i\psi$$

$$([c\text{-}cmt]_i\varphi \wedge [c\text{-}cmt]_i(\varphi \to \psi)) \to [c\text{-}cmt]_i\psi$$

Rational agents should withdraw at least one of their assertions or concessions if some unwanted consequences are derived from what they have explicitly asserted or conceded.

They are taken to be responsible for the logical consequences of what they have said at least to this extent.

The language of DMPCL

Definition

Take the same countably infinite set *Aprop* of proposition letters and the same finite set I of agents as before, with p ranging over *Aprop*, and i over I. The language $\mathcal{L}_{\mathrm{DMPCL}}$ of dynamified multi-agent propositional commitment logic DMPCL is given by:

$$\begin{array}{lll}\varphi & ::= & \top \mid p \mid \neg\varphi \mid \varphi\wedge\psi \mid [\textit{a-cmt}]_i\varphi \mid [\textit{c-cmt}]_i\varphi \mid [\pi]\varphi \\ \pi & ::= & \textit{assert}_i\varphi \mid \textit{concede}_i\varphi\end{array}$$

The truth definition for $\mathcal{L}_{\mathrm{DMPCL}}$

Definition

Let M be an $\mathcal{L}_{\mathrm{MPCL}}$-model and w a point in M. If $p \in Aprop$, and $i \in I$, then the truth definition for $\mathcal{L}_{\mathrm{DMPCL}}$ is given by expanding that of $\mathcal{L}_{\mathrm{MPCL}}$ mutatis mutandis with the following new clause:

(g) $M, w \models_{\mathrm{DMPCL}} [\textit{assert}_i\chi]\varphi$ iff $M_{\mathrm{assert}_i\chi}, w \models_{\mathrm{DMPCL}} \varphi$
(h) $M, w \models_{\mathrm{DMPCL}} [\textit{concede}_i\chi]\varphi$ iff $M_{\mathrm{concede}_i\chi}, w \models_{\mathrm{DMPCL}} \varphi$,

where $M_{\mathrm{assert}_i\chi}$ is the $\mathcal{L}_{\mathrm{MPCL}}$-model obtained from M by replacing $\rhd_i^M$ with $\{\langle x, y\rangle \in \rhd_i^M \mid M, y \models_{\mathrm{DMPCL}} \chi\}$ and $\blacktriangleright_i^M$ with $\{\langle x, y\rangle \in \blacktriangleright_i^M \mid M, y \models_{\mathrm{DMPCL}} \chi\}$,
and $M_{\mathrm{concede}_i\chi}$ is the $\mathcal{L}_{\mathrm{MPCL}}$-model obtained from M by replacing $\blacktriangleright_i^M$ with $\{\langle x, y\rangle \in \blacktriangleright_i^M \mid M, y \models_{\mathrm{DMPCL}} \chi\}$.

The proof system for $\mathcal{L}_{\mathrm{DMPCL}}$

Definition

The proof system for DMPCL includes all the axioms and all the rules of the proof system for MPCL, and in addition, necessitation rules for assertion modality and concession modality for each $i \in I$, and the following axioms:

$$\begin{array}{llcll}
(A1) & [\mathrm{assert}_i\varphi]p & \leftrightarrow & p & \\
(A2) & [\mathrm{assert}_i\varphi]\top & \leftrightarrow & \top & \\
(A3) & [\mathrm{assert}_i\varphi]\neg\psi & \leftrightarrow & \neg[\mathrm{assert}_i\varphi]\psi & \\
(A4) & [\mathrm{assert}_i\varphi](\psi\wedge\chi) & \leftrightarrow & [\mathrm{assert}_i\varphi]\psi\wedge[\mathrm{assert}_i\varphi]\chi & \\
(A5) & [\mathrm{assert}_i\varphi][\textit{a-cmt}]_j\psi & \leftrightarrow & [\textit{a-cmt}]_j[\mathrm{assert}_i\varphi]\psi & (i\neq j) \\
(A6) & [\mathrm{assert}_i\varphi][\textit{a-cmt}]_i\psi & \leftrightarrow & [\textit{a-cmt}]_i(\varphi\to[\mathrm{assert}_i\varphi]\psi) & \\
(A7) & [\mathrm{assert}_i\varphi][\textit{c-cmt}]_j\psi & \leftrightarrow & [\textit{c-cmt}]_j[\mathrm{assert}_i\varphi]\psi & (i\neq j) \\
(A8) & [\mathrm{assert}_i\varphi][\textit{c-cmt}]_i\psi & \leftrightarrow & [\textit{c-cmt}]_i(\varphi\to[\mathrm{assert}_i\varphi]\psi) & \\
(C1) & [\mathrm{concede}_i\varphi]p & \leftrightarrow & p & \\
(C2) & [\mathrm{concede}_i\varphi]\top & \leftrightarrow & \top & \\
(C3) & [\mathrm{concede}_i\varphi]\neg\psi & \leftrightarrow & \neg[\mathrm{concede}_i\varphi]\psi & \\
(C4) & [\mathrm{concede}_i\varphi](\psi\wedge\chi) & \leftrightarrow & [\mathrm{concede}_i\varphi]\psi\wedge[\mathrm{concede}_i\varphi]\chi & \\
(C5) & [\mathrm{concede}_i\varphi][\textit{a-cmt}]_j\psi & \leftrightarrow & [\textit{a-cmt}]_j[\mathrm{concede}_i\varphi]\psi & (\text{for any} j) \\
(C6) & [\mathrm{concede}_i\varphi][\textit{c-cmt}]_j\psi & \leftrightarrow & [\textit{c-cmt}]_j[\mathrm{concede}_i\varphi]\psi & (i\neq j) \\
(C7) & [\mathrm{concede}_i\varphi][\textit{c-cmt}]_i\psi & \leftrightarrow & [\textit{c-cmt}]_i(\varphi\to[\mathrm{concede}_i\varphi]\psi) &
\end{array}$$

Translation from $\mathcal{L}_{\mathrm{DMPCL}}$ to $\mathcal{L}_{\mathrm{MPCL}}$

Definition

The translation function that takes a formula from $\mathcal{L}_{\mathrm{DMPCL}}$ and yields a formula in $\mathcal{L}_{\mathrm{MPCL}}$ is defined as follows:

$$\begin{array}{llll}
t(p) & =p & t([\mathrm{assert}_i\varphi]p) & =p \\
 & & t([\mathrm{concede}_i\varphi]p) & =p \\
t(\top) & =\top & t([\mathrm{assert}_i\varphi]\top) & =\top \\
 & & t([\mathrm{concede}_i\varphi]\top) & =\top \\
t(\neg\varphi) & =\neg t(\varphi) & t([\mathrm{assert}_i\varphi]\neg\psi) & =\neg t([\mathrm{assert}_i\varphi]\psi) \\
 & & t([\mathrm{concede}_i\varphi]\neg\psi) & =\neg t([\mathrm{concede}_i\varphi]\psi) \\
t(\varphi\wedge\psi) & =t(\varphi)\wedge t(\psi) & t([\mathrm{assert}_i\varphi](\psi\wedge\chi)) & =t([\mathrm{assert}_i\varphi]\psi)\wedge t([\mathrm{assert}_i\varphi]\chi) \\
 & & t([\mathrm{concede}_i\varphi](\psi\wedge\chi)) & =t([\mathrm{concede}_i\varphi]\psi)\wedge t([\mathrm{concede}_i\varphi]\chi) \\
t([\textit{a-cmt}]_i\varphi) & =[\textit{a-cmt}]_i t(\varphi) & t([\mathrm{assert}_i\varphi][\textit{a-cmt}]_j\psi) & =[\textit{a-cmt}]_j t([\mathrm{assert}_i\varphi]\psi)\quad (i\neq j) \\
 & & t([\mathrm{assert}_i\varphi][\textit{a-cmt}]_i\psi) & =[\textit{a-cmt}]_i t(\varphi\to[\mathrm{assert}_i\varphi]\psi) \\
 & & t([\mathrm{concede}_i\varphi][\textit{a-cmt}]_j\psi) & =[\textit{a-cmt}]_j t([\mathrm{concede}_i\varphi]\psi) \\
t([\textit{c-cmt}]_i\varphi) & =[\textit{c-cmt}]_i t(\varphi) & t([\mathrm{assert}_i\varphi][\textit{c-cmt}]_j\psi) & =[\textit{c-cmt}]_j t([\mathrm{assert}_i\varphi]\psi)\quad (i\neq j) \\
 & & t([\mathrm{assert}_i\varphi][\textit{c-cmt}]_i\psi) & =[\textit{c-cmt}]_i t(\varphi\to[\mathrm{assert}_i\varphi]\psi) \\
 & & t([\mathrm{concede}_i\varphi][\textit{c-cmt}]_j\psi) & =[\textit{c-cmt}]_j t([\mathrm{concede}_i\varphi]\psi)\quad (i\neq j) \\
 & & t([\mathrm{concede}_i\varphi][\textit{c-cmt}]_i\psi) & =[\textit{c-cmt}]_i t(\varphi\to[\mathrm{concede}_i\varphi]\psi) \\
 & & t([\mathrm{assert}_i\varphi][\mathrm{assert}_j\psi]\chi) & =t([\mathrm{assert}_i\varphi]t([\mathrm{assert}_j\psi]\chi)) \\
 & & t([\mathrm{assert}_i\varphi][\mathrm{concede}_j\psi]\chi) & =t([\mathrm{assert}_i\varphi]t([\mathrm{concede}_j\psi]\chi)) \\
 & & t([\mathrm{concede}_i\varphi][\mathrm{assert}_j\psi]\chi) & =t([\mathrm{concede}_i\varphi]t([\mathrm{assert}_j\psi]\chi)) \\
 & & t([\mathrm{concede}_i\varphi][\mathrm{concede}_j\psi]\chi) & =t([\mathrm{concede}_i\varphi]t([\mathrm{concede}_j\psi]\chi))
\end{array}$$

Some results

Proposition

If $\varphi \in \mathcal{L}_{\text{MPCL}}$ is free of modalities indexed by i, the following formulas are valid:

$$[\text{assert}_i\varphi][\textit{a-cmt}]_i\varphi$$
$$[\text{assert}_i\varphi][\textit{c-cmt}]_i\varphi$$
$$[\text{concede}_i\varphi][\textit{c-cmt}]_i\varphi \ .$$

Theorem

There is a complete axiomatization of DMPCL.

Does the same strategy work for acts of asserting and conceding combined with acts of withdrawing ?

Dynamified Multiagent Propositional Commitment Logic with withdrawals DMPCL$^+$

adding dynamic modalities ↑ ↓ translation available ?

Multi-agent Propositional Commitment Logic MPCL

The language of DMPCL$^+$

Definition

Take the same countably infinite set *Aprop* of proposition letters and the same finite set *I* of agents as before, with *p* ranging over *Aprop*, and *i* over *I*. The language $\mathcal{L}_{\text{DPCMT}^+}$ of dynamified multi-agent propositional commitment logic with withdrawals DMPCL$^+$ is given by:

$$\varphi \ ::= \ \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\textit{a-cmt}]_i\varphi \mid [\textit{c-cmt}]_i\varphi \mid [\pi]\varphi$$
$$\pi \ ::= \ \text{assert}_i\varphi \mid \text{concede}_i\varphi \mid \circlearrowleft\text{assert}_i\varphi \mid \circlearrowleft\text{concede}_i\varphi$$

An update by withdrawing?

A sequence of acts: $\ldots, \text{assert}_i\chi, \text{assert}_j\xi, \text{assert}_i\eta, \ldots$
$\Downarrow \quad \circlearrowleft\text{assert}_j\xi$
A reduced sequence: $\ldots, \text{assert}_i\chi, \text{assert}_i\eta, \ldots$

The set of propositional commitments agents bear after *j*'s act of withdrawing of the form $\circlearrowleft\text{assert}_j\xi$ will be, other things being equal, the same as the set of propositional commitments they would bear if *j* had not asserted that ξ.

A positive commitment act sequence

If σ is a sequence of moves in an argumentation, it may involve not only acts of asserting and conceding but also acts of withdrawing.

For the sake of simplicity, we will only consider a special kind of sequences, namely, a sequence $\sigma = \langle \pi_1, \pi_2, \cdots, \pi_n \rangle$ of speech acts π_j $(1 \leq j \leq n)$ such that each π_j is either of the form $\mathrm{assert}_i\varphi$ for some $i \in I$ or of the form $\mathrm{concede}_i\varphi$ for some $i \in I$. We call such a sequence a positive commitment act sequence , or a pca-sequence for short.

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

Reduced positive commitmment act sequence

Definition

Let σ be a (possibly empty) positive commitment act sequence $\langle \pi_1, \cdots, \pi_n \rangle$ such that each π_j $(1 \leq j \leq n)$ is of the form $\mathrm{assert}_i\varphi$ or of the form $\mathrm{concede}_i\varphi$ for some $i \in I$. We define the reduced sequence $\sigma \upharpoonright \circlearrowleft \mathrm{assert}_i\varphi$ ($\sigma \upharpoonright \circlearrowleft \mathrm{concede}_i\varphi$) obtained by withdrawing every occurrence of an act of type $\mathrm{assert}_i\varphi$ ($\mathrm{concede}_i\varphi$) from σ as follows:

(To be continued)

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

Reduced pca-sequence (continued)

(Ai) if σ is empty, $\sigma \upharpoonright \circlearrowleft \mathrm{assert}_i\varphi = \sigma$

(Aii) if $\sigma = \langle \pi_1, \cdots, \pi_n \rangle$, and $\pi_n = \mathrm{assert}_i\varphi$,
$\sigma \upharpoonright \circlearrowleft \mathrm{assert}_i\varphi = \langle \pi_1, \cdots, \pi_{n-1} \rangle \upharpoonright \circlearrowleft \mathrm{assert}_i\varphi$

(Aiii) if $\sigma = \langle \pi_1, \cdots, \pi_n \rangle$, and $\pi_n \neq \mathrm{assert}_i\varphi$,
$\sigma \upharpoonright \circlearrowleft \mathrm{assert}_i\varphi = \langle \langle \pi_1, \cdots, \pi_{n-1} \rangle \upharpoonright \circlearrowleft \mathrm{assert}_i\varphi, \pi_n \rangle$

(Ci) if σ is empty, $\sigma \upharpoonright \circlearrowleft \mathrm{concede}_i\varphi = \sigma$

(Cii) if $\sigma = \langle \pi_1, \cdots, \pi_n \rangle$, and $\pi_n = \mathrm{concede}_i\varphi$,
$\sigma \upharpoonright \circlearrowleft \mathrm{concede}_i\varphi = \langle \pi_1, \cdots, \pi_{n-1} \rangle \upharpoonright \circlearrowleft \mathrm{concede}_i\varphi$

(Ciii) if $\sigma = \langle \pi_1, \cdots, \pi_n \rangle$, and $\pi_n \neq \mathrm{concede}_i\varphi$,
$\sigma \upharpoonright \circlearrowleft \mathrm{concede}_i\varphi = \langle \langle \pi_1, \cdots, \pi_{n-1} \rangle \upharpoonright \circlearrowleft \mathrm{concede}_i\varphi, \pi_n \rangle$.

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion
Conflicting commands
Acts of commanding and promising
Obligations and preferences
Assertions, concessions and their withdrawals

The Problem of Notation

Given a pca-sequence $\sigma = \langle \pi_1, \cdots, \pi_n \rangle$, the model obtained by updating M with σ is denoted by $(\ldots(M_{\pi_1})\ldots)_{\pi_n}$ in the notation of the truth definition for $\mathcal{L}_{\mathrm{DMPCL}}$.

This notation leads to a paradox when we deal with withdrawals. Let abbreviate $(\ldots(M_{\pi_1})\ldots)_{\pi_n}$ as M_σ. Now there may be another model N and a pcs-sequence τ such that $N_\tau = M$. Then we may have

$(N_\tau)_\sigma = M_\sigma$ but $((N_\tau)_\sigma)_{\circlearrowleft \mathrm{concede}_i\varphi} \neq (M_\sigma)_{\circlearrowleft \mathrm{concede}_i\varphi}$.

Truth Definition 1/5

Definition

Let M be an $\mathcal{L}_{\text{MPCL}}$-model, σ a positive commitment act sequence, and w a point in M. If $p \in Aprop$, and $i \in I$, then:

(a) $M, \sigma, w \models_{\text{DMPCL+}} p$ iff $w \in V^M(p)$

(b) $M, \sigma, w \models_{\text{DMPCL+}} \top$

(c) $M, \sigma, w \models_{\text{DMPCL+}} \neg\varphi$ iff it is not the case that $M, \sigma, w \models_{\text{DMPCL+}} \varphi$

(d) $M, \sigma, w \models_{\text{DMPCL+}} (\varphi \wedge \psi)$ iff $M, \sigma, w \models_{\text{DMPCL+}} \varphi$ and $M, \sigma, w \models_{\text{DMPCL+}} \psi$

Truth Definition 2/5

(e) $M, \sigma, w \models_{\text{DMPCL+}} [a\text{-}cmt]_i\varphi$ iff for all v s. t. $\langle w, v\rangle \in \rhd_i^M \upharpoonright \sigma$, $M, \sigma, v \models_{\text{DMPCL+}} \varphi$

(f) $M, \sigma, w \models_{\text{DMPCL+}} [c\text{-}cmt]_i\varphi$ iff for all v s. t. $\langle w, v\rangle \in \blacktriangleright_i^M \upharpoonright \sigma$, $M, \sigma, v \models_{\text{DMPCL+}} \varphi$

(g) $M, \sigma, w \models_{\text{DMPCL+}} [\mathrm{assert}_i\chi]\varphi$ iff $M, \langle\sigma, \mathrm{assert}_i\chi\rangle, w \models_{\text{DMPCL+}} \varphi$

(h) $M, \sigma, w \models_{\text{DMPCL+}} [\mathrm{concede}_i\chi]\varphi$ iff $M, \langle\sigma, \mathrm{concede}_i\chi\rangle, w \models_{\text{DMPCL+}} \varphi$

Truth Definition 3/5

(i) $M, \sigma, w \models_{\text{DMPCL+}} [\circlearrowleft\mathrm{assert}_i\chi]\varphi$ iff $M, \sigma \upharpoonright\circlearrowleft \mathrm{assert}_i\chi$, $w \models_{\text{DMPCL+}} \varphi$

(j) $M, \sigma, w \models_{\text{DMPCL+}} [\circlearrowleft\mathrm{concede}_i\chi]\varphi$ iff $M, \sigma \upharpoonright\circlearrowleft \mathrm{concede}_i\chi$, $w \models_{\text{DMPCL+}} \varphi$,

where $\rhd_i^M \upharpoonright \sigma$ and $\blacktriangleright_i^M \upharpoonright \sigma$ are

(To continue)

Truth Definition 4/5

where $\rhd_i^M \upharpoonright \sigma =$

- $\rhd_i^M$ *if σ is empty,*
- $\{\langle x, y\rangle \in \rhd_i^M \upharpoonright \langle\pi_1, \ldots, \pi_{n-1}\rangle \mid M, \langle\pi_1, \ldots, \pi_{n-1}\rangle, y \models_{\text{DMPCL+}} \psi\}$ *if* $\sigma = \langle\pi_1, \ldots, \pi_n\rangle$ *and* $\pi_n = \mathrm{assert}_i\psi$,
- $\rhd_i^M \upharpoonright \langle\pi_1, \ldots, \pi_{n-1}\rangle$ *if* $\sigma = \langle\pi_1, \ldots, \pi_n\rangle$ *and* $\pi_n \neq \mathrm{assert}_i\psi$,

and

(To continue)

Truth Definition 5/5

$\blacktriangleright_i^M \upharpoonright \sigma =$

- $\blacktriangleright_i^M$ *if* σ *is empty,*
- $\{\langle x, y\rangle \in \blacktriangleright_i^M \upharpoonright \langle \pi_1, \ldots, \pi_{n-1}\rangle \mid M, \langle \pi_1, \ldots, \pi_{n-1}\rangle, y \models_{\text{DMPCL+}} \psi\}$ *if* $\sigma = \langle \pi_1, \ldots, \pi_n\rangle$ *and* *either* $\pi_n = \text{assert}_i\psi$ *or* $\pi_n = \text{concede}_i\psi$*,*
- $\blacktriangleright_i^M \upharpoonright \langle \pi_1, \ldots, \pi_{n-1}\rangle$ *if* $\sigma = \langle \pi_1, \ldots, \pi_n\rangle$*,* $\pi_n \neq \text{assert}_i\psi$*, and* $\pi_n \neq \text{concede}_i\psi$*.*

A result and an open problem

A result

Acts of withdrawing behave slightly differently from contraction studied in belief revision. Let $\mathcal{B}$ be a set of beliefs of an agent, say a. Then in the AGM approach, contraction $\ominus$ is supposed to satisfy the postulate that $\varphi \notin \mathcal{B} \ominus \varphi$ if $\nvdash \varphi$, but we have $M, \sigma \upharpoonright\circlearrowleft \text{assert}_a p, w \models_{\text{DMPCL+}} [\text{a-cmt}]_a p$ if σ include $\text{assert}_a q$ and $\text{assert}_a(q \rightarrow p)$.

An open problem

The completeness problem of DMPCL^+ is still open.

Scorekeeping

- The notion of scorekeeping is introduced into the discussion of language by Lewis (1979) and utilized by Brandom (1994) in developing his theory of meaning based on Wittgenstein's notion of meaning as use.
- In Brandom's version, each agent is considered as a deontic scorekeeper, and "the significance of an assertion of *p*" is considered as "a mapping that associates with one social deontic score—characterising the stage before that speech act is performed, according to some scorekeeper—the set of scores for conversational stage that results from the assertion, according to the same scorekeeper"(*ibid.*, 190).

Scorekeeping for argumentation games

- We will only consider "the official score" kept by an idealised scorekeeper, and examine how DMPCL$^+$ can be applied to such official scorekeeping in an argumentation game.
- In order to do so, we need a special model that can represent the initial stage of the game.

Definition

Given a countably infinite set Aprop of proposition letters, and the set $I = \{a, b\}$, with p ranging over Aprop and i over I. Then, the initial stage model is the tuple . . .

Continued

$M^0 = \langle W^0, \{\rhd_i^0 \mid i \in I\}, \{\blacktriangleright_i^0 \mid i \in I\}, V^0\rangle$ where:

(i) W^0 is the power set $\mathscr{P}(\mathrm{Aprop})$ of Aprop,
(ii) $\rhd_i^0 = W^0 \times W^0$ for each $i \in I$,
(iii) $\blacktriangleright_i^0 = W^0 \times W^0$ for each $i \in I$,
(iv) V^0 is the function that assigns a subset $V^0(p) = \{w \in W^0 \mid p \in w\}$ of W^0 to each proposition letter $p \in \mathrm{Aprop}$.

What we have at the initial stage

For any agent i, for any proposition letter p, and for any point $w \in W^0$, if σ is empty, we have

$$M^0, \sigma, w \not\models_{\mathrm{DMPCL}^+} [\text{a-cmt}]_i p$$
$$M^0, \sigma, w \not\models_{\mathrm{DMPCL}^+} [\text{a-cmt}]_i \neg p$$
$$M^0, \sigma, w \not\models_{\mathrm{DMPCL}^+} [\text{c-cmt}]_i p$$
$$M^0, \sigma, w \not\models_{\mathrm{DMPCL}^+} [\text{c-cmt}]_i \neg p$$

Thus each agent has no substantial propositional commitments at the initial stage.

What DMPCL$^+$ enables us to do

- Then we can reason about what propositional commitments agents will bear at each stage after each of their acts of asserting, conceding or withdrawing.
- This doesn't gives us the whole score of each play of an argumentation game. The official scorekeeper may have to record other factors such as penalties for withdrawing, the relation between the moves made by the participants, etc.
- But DMPCL$^+$ can be said to capture the evolution of the score at least partially.

Scorekeeping for language games

- The notion of, or the metaphor of, scorekeeping can be extended to more complex language games where not only acts of asserting, conceding, and withdrawing but also acts of commanding, promising, etc. are involved along with non-verbal actions.
- Then the dynamic logics of speech acts can be used to reason about the changes and non-changes brought about by speech acts.

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

Scorekeeping for language games

- The dynamic logics of speech acts can thus partially characterize the scorekeeping function for a language game, where sppeech acts they deal with are involved as moves made by participants at various stages.
- The conventional or institutional effects of an illocutionary act can then be seen in the score that charcterizes the stage after it is performed.
- In contrast, the real effects of a perlocutionary act may be seen in the reponses agents make or be concealded in their feelings, thoughts, etc.

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Introduction
DEL and A dynamic logic of acts of commanding
Refinements and Variations
Scorekeeping and dynamic logics of speech acts
Conclusion

Summary

- The effects of illocutionary acts of asserting and conceding as well as illocutionary acts of commanding and promising can be captured in logical terms.
- Their effects are public and involve more than mere securing of the upatke.
- The workings of acts of withdrawing can also be modeled, but the axiomatizability is still open.
- The dynamic logics can be seen as patially characterizing the scorekeeping function for a complex social interaction.

Tomoyuki Yamada Scorekeeping and Dynamic Logics of Speech Acts

Nash Equilibria in Multi-agent Deontic Logic

Allard Tamminga

Department of Theoretical Philosophy
University of Groningen
The Netherlands

SOCREAL 2010
Hokkaido University
Sapporo, 27 March 2009

Outline

- Extrinsic Preferences
- Act Consequentialism
- Language of Multi-agent Deontic Logic
- Consequentialist Models
- Absolute and Conditional $\mathcal{F}$-Dominance
- Semantics of Multi-agent Deontic Logic
- Strategic Games and Nash Equilibria
- From Strategic Games to Consequentialist Frames
- A Formal Characterization of Nash Equilibria

Extrinsic Preferences

How to bring deontic logic and game theory together conceptually?

- It is perfectly possible that I have the obligation to do X, but at the same time prefer not to do X.
- We make a distinction between *extrinsic* preferences (which are the result of a previous judgment of betterness on the basis of reasons) and *intrinsic* preferences (which reflect the unreasoned subjective likings of the agents concerned).
- It makes perfect sense that I have the obligation to do X and at the same time *intrinsically* prefer not to do X (I just don't feel like it). Some intellectual effort is needed, however, to imagine a situation where I have the obligation to do X and at the same time *extrinsically* prefer not to do X.
- Hence, we assume that the preferences that figure in our deontic logic, guiding agents in evaluating the moral rightness of their actions, are extrinsic.

Act Consequentialism

- It still might be that I have the obligation to do X (I promised to do so) and at the same time extrinsically prefer not to do X.
- This possibility is minimized once we adopt an evaluative version of *act consequentialism* as our moral theory. In evaluative act consequentialism, the moral rightness of an action only depends on the value of its consequences.
- In evaluative act consequentialism using extrinsic preferences it is plausible that an action is obligatory if and only if that action is extrinsically preferred.

Language of Multi-agent Deontic Logic

Within a formal framework inspired by evaluative act consequentialism using extrinsic preferences, we study the logical behavior of two types of action permissions:

- *Absolute action permissions*: "In group $\mathcal{F}$'s interest, group $\mathcal{G}$ may perform action $\alpha_{\mathcal{G}}$" (abbreviated as $\mathsf{P}^{\mathcal{F}}_{\mathcal{G}}\alpha_{\mathcal{G}}$)
- *Conditional action permissions*: "If group $\mathcal{H}$ were to perform action $\alpha_{\mathcal{H}}$, then, in group $\mathcal{F}$'s interest, group $\mathcal{G}$ may perform action $\alpha_{\mathcal{G}}$" (abbreviated as $\mathsf{P}^{\mathcal{F}}_{\mathcal{G}}(\alpha_{\mathcal{G}}/\alpha_{\mathcal{H}})$).

Language of Multi-agent Deontic Logic

$\mathfrak{L}$ is a propositional modal language built from a countable set $\mathfrak{A} = \{\alpha^n_{\mathcal{G}} : \mathcal{G} \subseteq \mathcal{N} \text{ and } n \in \mathbb{N}\}$ of atomic propositions.

$\mathfrak{L}$ is the smallest set satisfying the following conditions:

- $\mathfrak{A} \subseteq \mathfrak{L}$
- If $\varphi \in \mathfrak{L}$, then $\neg\varphi \in \mathfrak{L}$
- If $\varphi, \psi \in \mathfrak{L}$, then $\varphi \wedge \psi \in \mathfrak{L}$
- If $\alpha_{\mathcal{G}} \in \mathfrak{A}$ and $\mathcal{F} \subseteq \mathcal{N}$, then $\mathsf{P}^{\mathcal{F}}_{\mathcal{G}}\alpha_{\mathcal{G}} \in \mathfrak{L}$
- If $\alpha_{\mathcal{G}}, \alpha_{\mathcal{H}} \in \mathfrak{A}$ and $\mathcal{F} \subseteq \mathcal{N}$ and $\mathcal{H} \subseteq \mathcal{N} - \mathcal{G}$, then $\mathsf{P}^{\mathcal{F}}_{\mathcal{G}}(\alpha_{\mathcal{G}}/\alpha_{\mathcal{H}}) \in \mathfrak{L}$.

Language of Multi-agent Deontic Logic

Two examples:

- $\mathsf{P}^i_i \alpha_i$ means "In his own interest, agent i may perform action α_i".
- $\mathsf{P}^{\mathcal{N}}_{\mathcal{G}}(\alpha_{\mathcal{G}}/\alpha_i)$ means "If agent i were to perform action α_i, then, in the grand coalition's interest, group $\mathcal{G}$ may perform action $\alpha_{\mathcal{G}}$".

Consequentialist Frames 1

A *consequentialist frame* $\mathfrak{F}$ is a quadruple $\langle \mathcal{W}, \mathcal{N}, Choice, (\succeq_{\mathcal{F}}) \rangle$, where $\mathcal{W}$ is a non-empty set of possible worlds, $\mathcal{N}$ is a finite set of agents, *Choice* is a choice function, and $\succeq_{\mathcal{F}}$ is a reflexive, transitive, and complete relation on $\mathcal{W}$ for each $\mathcal{F} \subseteq \mathcal{N}$.

Choice sets of *individual agents* are given by a function $Choice : \mathcal{N} \to \mathscr{P}(\mathscr{P}(\mathcal{W}))$, such that

1. for each agent $i \in \mathcal{N}$ it holds that $Choice(i)$ is a partition of $\mathcal{W}$,
2. for each selection function s assigning to each agent $i \in \mathcal{N}$ a set of possible worlds $s(i)$ such that $s(i) \in Choice(i)$ it holds that $\bigcap_{i \in \mathcal{N}} s(i)$ is non-empty (*independence of agents*).

Consequentialist Frames 2

This choice function for individual agents is extended to a function $\mathit{Choice} : \mathscr{P}(\mathcal{N}) \to \mathscr{P}(\mathscr{P}(\mathcal{W}))$ for groups of agents. Let *Select* be the set of all selection functions s assigning to each individual agent $i \in \mathcal{N}$ an option $s(i) \in \mathit{Choice}(i)$. Then

$$\mathit{Choice}(\mathcal{G}) = \{\bigcap_{i \in \mathcal{G}} s(i) : s \in \mathit{Select}\},$$

if $\mathcal{G}$ is non-empty. Otherwise, $\mathit{Choice}(\mathcal{G}) = \{\mathcal{W}\}$.

Consequentialist Models

A *consequentialist model* $\mathfrak{M}$ is an ordered pair $\langle \mathfrak{F}, v \rangle$, where $\mathfrak{F}$ is a consequentialist frame and v an interpretation function that assigns to each atomic proposition $\alpha_{\mathcal{G}} \in \mathfrak{A}$ an action $v(\alpha_{\mathcal{G}}) \in Choice(\mathcal{G})$.

In a consequentialist model, each possible action of each group of agents is assumed to have a name, that is, for each $\mathcal{G} \subseteq \mathcal{N}$ and each $K \in Choice(\mathcal{G})$ there is an atomic proposition $\alpha_{\mathcal{G}} \in \mathfrak{A}$ such that $v(\alpha_{\mathcal{G}}) = K$.

Absolute and Conditional $\mathcal{F}$-Dominance

Let $\mathfrak{F}$ be a consequentialist frame. Let $\mathcal{F}, \mathcal{G} \subseteq \mathcal{N}$ and $\mathcal{H} \subseteq \mathcal{N} - \mathcal{G}$. Let $K, K' \in \mathit{Choice}(\mathcal{G})$ and $L \in \mathit{Choice}(\mathcal{H})$. Then

$K \geq^{\mathcal{F}}_{\mathcal{G}} K'$	iff	for all $S \in \mathit{Choice}(\mathcal{N} - \mathcal{G})$ and for all $w, w' \in \mathcal{W}$ it holds that if $w \in K \cap S$ and $w' \in K' \cap S$, then $w \succeq_{\mathcal{F}} w'$.
$K \geq^{\mathcal{F}}_{(\mathcal{G}/\mathcal{H}, L)} K'$	iff	for all $S \in \mathit{Choice}((\mathcal{N} - \mathcal{G}) - \mathcal{H})$ and for all $w, w' \in \mathcal{W}$ it holds that if $w \in K \cap L \cap S$ and $w' \in K' \cap L \cap S$, then $w \succeq_{\mathcal{F}} w'$.

Semantics of Multi-agent Deontic Logic

Let $\mathfrak{M} = \langle \mathfrak{F}, v \rangle$ be a consequentialist model. Let $\mathcal{F}, \mathcal{G} \subseteq \mathcal{N}$ and let $\mathcal{H} \subseteq \mathcal{N} - \mathcal{G}$. Let $w \in \mathcal{W}$ and let $\alpha_{\mathcal{G}}, \alpha_{\mathcal{H}} \in \mathfrak{A}$ and $\varphi, \psi \in \mathfrak{L}$. Then

$\mathfrak{M}, w \models \alpha_{\mathcal{G}}$	iff	$w \in v(\alpha_{\mathcal{G}})$
$\mathfrak{M}, w \models \neg\varphi$	iff	$\mathfrak{M}, w \not\models \varphi$
$\mathfrak{M}, w \models \varphi \wedge \psi$	iff	$\mathfrak{M}, w \models \varphi$ and $\mathfrak{M}, w \models \psi$
$\mathfrak{M}, w \models \mathsf{P}^{\mathcal{F}}_{\mathcal{G}} \alpha_{\mathcal{G}}$	iff	for all K' in $Choice(\mathcal{G})$ with $K' \neq v(\alpha_{\mathcal{G}})$ it holds that $v(\alpha_{\mathcal{G}}) \geq^{\mathcal{F}}_{\mathcal{G}} K'$
$\mathfrak{M}, w \models \mathsf{P}^{\mathcal{F}}_{\mathcal{G}} (\alpha_{\mathcal{G}} / \alpha_{\mathcal{H}})$	iff	for all K' in $Choice(\mathcal{G})$ with $K' \neq v(\alpha_{\mathcal{G}})$ it holds that $v(\alpha_{\mathcal{G}}) \geq^{\mathcal{F}}_{(\mathcal{G}/\mathcal{H}, v(\alpha_{\mathcal{H}}))} K'$.

Semantics

Theorem

Let $\mathfrak{M} = \langle \mathfrak{F}, v \rangle$ *be a consequentialist model. Let* $\mathcal{F}, \mathcal{G} \subseteq \mathcal{N}$ *and let* $\mathcal{H} \subseteq \mathcal{N} - \mathcal{G}$*. Let* $\alpha_{\mathcal{G}} \in \mathfrak{A}$*. Then the following statements are equivalent:*

- $\mathfrak{M} \models \mathsf{P}^{\mathcal{F}}_{\mathcal{G}} \alpha_{\mathcal{G}}$
- $\mathfrak{M} \models \mathsf{P}^{\mathcal{F}}_{\mathcal{G}} (\alpha_{\mathcal{G}} / \alpha_{\mathcal{H}})$ *for all* $\alpha_{\mathcal{H}} \in \mathfrak{A}$

Strategic Games and Nash Equilibria

A *strategic game* is a triple $G = \langle N, (A_i), (\succsim_i) \rangle$, where N is a finite set of players, for each player $i \in N$ it holds that A_i is a non-empty set of actions available to player i, and for each player $i \in N$ it holds that $\succsim_i$ is a preference relation on the set of outcomes $A = \times_{i \in N} A_i$.

Preference relations $\succsim_i$ are assumed to be reflexive, transitive, and complete.

An outcome $a^* \in A$ is a *Nash equilibrium* of a strategic game $G = \langle N, (A_i), (\succsim_i) \rangle$ if and only if for each player $i \in N$ it holds that

$$(a^*_{-i}, a^*_i) \succsim_i (a^*_{-i}, a_i) \text{ for all } a_i \in A_i.$$

From Strategic Games to Consequentialist Frames

Let $G = \langle N, (A_i), (\succsim_i)\rangle$ be a strategic game. The quadruple $\mathfrak{T}(G) = \langle \mathcal{W}, \mathcal{N}, Choice, (\succeq_{\mathcal{F}})\rangle$ is defined as follows:

- $\mathcal{W} = A$
- $\mathcal{N} = N$
- $Choice(\mathcal{G}) =$
$$\begin{cases} \{\{(a_{\mathcal{G}}, a_{-\mathcal{G}}) \in A : a_{-\mathcal{G}} \in A_{-\mathcal{G}}\} : a_{\mathcal{G}} \in A_{\mathcal{G}}\}, & \text{if } \mathcal{G} \neq \emptyset \\ \{\mathcal{W}\}, & \text{otherwise} \end{cases}$$
- $\succeq_{\mathcal{F}} = \begin{cases} \succsim_i, & \text{if } \mathcal{F} = \{i\} \\ \mathcal{W} \times \mathcal{W}, & \text{otherwise} \end{cases}$

Theorem

Let G be a strategic game. Then $\mathfrak{T}(G)$ is a consequentialist frame.

A Formal Characterization of Nash Equilibria

Finally, we define a suitable valuation function v_f that keeps track of which atomic proposition $\alpha_{\mathcal{G}}$ in $\mathfrak{A}_{\mathcal{G}}$ is validated by the performance of which action $a_{\mathcal{G}}$ in $A_{\mathcal{G}}$.

- Let f be an injective map that for each $\mathcal{G} \subseteq \mathcal{N}$ assigns to each action $a_{\mathcal{G}}$ in each $A_{\mathcal{G}}$ an atomic proposition $\alpha_{\mathcal{G}}$ in $\mathfrak{A}_{\mathcal{G}}$.
- If there is an action $a_{\mathcal{G}}$ in $A_{\mathcal{G}}$ such that $f(a_{\mathcal{G}}) = \alpha_{\mathcal{G}}$, we define $v_f(\alpha_{\mathcal{G}}) = \{(a_{\mathcal{G}}, a_{-\mathcal{G}}) \in A : a_{-\mathcal{G}} \in A_{-\mathcal{G}}\}$. If there is no action $a_{\mathcal{G}}$ in $A_{\mathcal{G}}$ such that $f(a_{\mathcal{G}}) = \alpha_{\mathcal{G}}$, then we simply put $v_f(\alpha_{\mathcal{G}}) = K$ for some unique designated $K \in \mathit{Choice}(\mathcal{G})$.

Theorem

Let G be a strategic game. Then

$$a^* \text{ is a Nash equilibrium of } G$$
$$\text{iff}$$
$$\langle \mathfrak{T}(G), v_f \rangle \models \bigwedge_{i \in \mathcal{N}} \mathsf{P}^i_i(f(a^*_i)/f(a^*_{-i})).$$

Measurement-Theoretic Foundation of Threshold Utility Maximiser's Preference Logic (Extended Abstract)

Satoru Suzuki

Faculty of Arts and Sciences, Komazawa University
1-23-1, Komazawa, Setagaya-ku, Tokyo, 154-8525 Japan
bxs05253@nifty.com

Abstract. The first problem of this paper is as follows: what kind of preference logic can formalise inferences in which indifference is not transitive? (Nontransitivity Problem) The aim of this paper is to propose a new version of complete and decidable extrinsic preference logic–threshold utility maximiser's preference logic (TUMPL) that can solve the Nontransitivity Problem. Generally, preference logics are in danger of inviting the following problem: almost every principle which has been proposed as fundamental to one preference logic has been rejected by another one. (Fundamental Problem of Intrinsic Preference) A corollary of the Scott-Suppes theorem in measurement theory enables TUMPL to avoid the Fundamental Problem of Intrinsic Preference.

Key Words: preference logic, semiorder, Sorites Paradox, threshold utility maximisation, bounded rationality, measurement theory, representation theorem.

1 Introduction

The economist Armstrong ([1]) was one of the first to argue that *indifference* is not always *transitive*. Luce ([[5]: 179]) gave the following counterexample to the transitivity of indifference:

Example 1 (Avoidance of Sorites Paradox in Preference). If indifference were transitive, then he would be unable to detect any weight differences, however great, which is patently false.... Find a subject who prefers a cup of coffee with one cube of sugar to one with five cubes.... Now prepare 401 cups of coffee with $(1 + \frac{i}{100})x$ grams of sugar, $i = 0, 1, \ldots, 400$, where x is the weight of one cube of sugar. It is evident that he will be indifferent between cup i and cup $i + 1$, for any i, but by choice he is not indifferent between $i = 0$ and $i = 400$. ■

This example shows a situation where we would face the *Sorites Paradox*[1] in preference if indifference were transitive. The first problem now arises:

[1] [4] gives a comprehensive survey of topics of vagueness.

Problem 1 (Nontransitivity Problem). What kind of preference logic can formalise inferences in which indifference is not transitive? ■

We call it the *Nontransitivity Problem.* The aim of this paper is to propose a new version of complete and decidable *extrinsic* preference logic–*threshold utility maximiser's preference logic* (TUMPL) that can solve the Nontransitivity Problem.

Generally, preference-based logics are in danger of inviting the following problem. Von Wright ([14]) divided preferences into two categories: *extrinsic* and *intrinsic* preference. An agent is said to prefer φ_1 extrinsically to φ_2 if φ_1 is better than φ_2 in some explicit respect. So we can explain extrinsic preference from some explicit point of view. If we cannot explain preference from any explicit point of view, we call it intrinsic. Most preference logics that have been proposed are intrinsic but little attention has been paid to extrinsic preference. Von Wright ([15]) posed the following fundamental problem intrinsic preference logics faced.

Problem 2 (Fundamental Problem of Intrinsic Preference). The development of a satisfactory logic of preference has turned out to be unexpectedly problematic. The evidence for this lies in the fact that almost every principle which has been proposed as fundamental to one preference logic has been rejected by another one. ■

We call it the *Fundamental Problem of Intrinsic Preference.* For example, the status of such logical properties as Transitivity, Contraposition, Conjunctive expansion, Disjunctive Distribution and Conjunctive Distribution is as follows:

Example 2 (Variety of Preferences).

	von Wright ([14])	Martin ([6])	Chisholm and Sosa ([3])
Transitivity	+	+	+
Contraposition	−	+	−
Conjunctive Expansion	+	−	−
Disjunctive Distribution	−	−	−
Conjunctive Distribution	+	−	−

'+' denotes the property in question being provable in the logic in question. '−' denotes the property in question not being provable in the logic in question. Conjunctive Expansion says that an agent does not prefer φ_1 to φ_2 iff he does not prefer $\varphi_1\&\neg\varphi_2$ to $\varphi_2\&\neg\varphi_1$. Disjunctive Distribution says that if he does not prefer $\varphi_1 \vee \varphi_2$ to φ_3, then he does not prefer φ_1 to φ_3 or does not prefer φ_2 to φ_3. Conjunctive Distribution says that if he does not prefer φ_1 to φ_2 and does not prefer φ_3 to φ_2, then he does not prefer $\varphi_1 \vee \varphi_3$ to φ_2. ■

According to Mullen ([7]), we can analyse its cause as follows. The adequacy criteria for intrinsic preference principles considered by preference logicians have been whether the principles are consistent with our *intuitions* of reasonableness. But each intuition often disagrees even on the fundamental properties. Different theories, such as ethics, welfare economics, consumer demand theory, game theory and decision theory make different demands upon the fundamental properties of preference. So if we would like to propose preference logic that can avoid the Fundamental Problem of Intrinsic Preference, it should be constructed not from intuition but from a *theory* or a *rule in a theory*, that is, it should be *extrinsic*. In order to avoid the Fundamental Problem of Intrinsic Preference, we resort to *measurement theory*.[2] There are two main problems with measurement theory:

1. the representation problem–justifying the assignment of numbers to objects or propositions,
2. the uniqueness problem–specifying the transformation up to which this assignment is unique.

A solution to the former can be furnished by a *representation theorem*, which establishes that the chosen numerical system preserves the relations of the relational system. When we provide TUMPL with a model based on semiorders, by virtue of a *corollary* of the *Scott-Suppes representation theorem*, we adopt *threshold utility maximisation* as a rule in decision theory that makes demands upon the fundamental properties of preference, which enables TUMPL to avoid the Fundamental Problem of Intrinsic Preference.

The structure of this paper is as follows. In Section 2, we state weak orders as global rationality and semiorders as bounded rationality. In Section 3, we define the language $\mathcal{L}_{\mathsf{TUMPL}}$ of TUMPL, and define a structured Kripke model $\mathcal{M}$ for preference, and provide TUMPL with a truth definition, and provide TUMPL with a proof system, and sketch the proof of the soundness, completeness and decidability of TUMPL.

2 Weak Orders as Global Rationality and Semiorders as Bounded Rationality

The standard model of economics is based on *global rationality* that requires an *optimising behavior*. *Utility maximisation* is a typical example of an optimising behavior. Cantor ([2]) proved the representation theorem for utility maximisation.

Theorem 1 (Representation for Utility Maximisation, Cantor ([2])). *Suppose* $\mathbf{A}$ *is a countable set and* $\succeq$ *is a binary relation on* $\mathbf{A}$*. Then* $\succeq$ *is a weak order (transitive and connected) iff there is a function* $u : \mathbf{A} \to \mathbb{R}$ *such that for any* $x, y \in \mathbf{A}$*,*

$$x \succeq y \text{ iff } u(x) \geq u(y).$$

■

But according to Simon ([12]), cognitive and information-processing constrains on the capabilities of agents, together with the complexity of their environment, render an optimising behavior an unattainable ideal. He dismissed the idea that

[2] [8] gives a comprehensive survey of measurement theory.

agents should exhibit global rationality and suggested that they in fact exhibit *bounded rationality* that allows a *satisficing behavior*.[3] One explanation for Example 1 is that the nontransitivity of indifference results from the fact that we cannot generally discriminate very close quantities. The concept of a *semiorder* was introduced by Luce ([5]) to construct a model to interpret situations like Example 1 of nontransitive indifference with a *threshold* of discrimination. Scott and Suppes defined ([[9]: 117]) a semiorder as follows:

Definition 1 (Semiorder). *$\succ$ on $\mathbf{A}$ is called a semiorder if, for any $w, x, y, z \in \mathbf{A}$, the following conditions are satisfied:*

1. *$x \not\succ x$. (Irreflexivity),*
2. *If $w \succ x$ and $y \succ z$, then $w \succ z$ or $y \succ x$. (Intervality),*
3. *If $w \succ x$ and $x \succ y$, then $w \succ z$ or $z \succ y$. (Semitransitivity).*

■

Threshold utility maximisation is a typical example of a satisficing behavior. Scott and Suppes ([9]) proved a representation theorem for threshold utility maximisation when $\mathbf{A}$ is *finite*.

Theorem 2 (Representation for Threshold Utility Maximisation, Scott and Suppes ([9])). *Suppose that $\succ$ is a binary relation on a finite set $\mathbf{A}$ and δ is a positive number. Then $\succ$ is a semiorder iff there is a function $u : \mathbf{A} \to \mathbb{R}$ such that for any $x, y \in \mathbf{A}$,*

$$x \succ y \text{ iff } u(x) > u(y) + \delta.$$

■

Remark 1. Scott ([10]) simplified the Scott-Suppes theorem in terms of the solvability of finite system of linear inequalities. ■

3 Threshold Utility Maximiser's Preference Logic TUMPL

3.1 Language

We define the language $\mathcal{L}_{\mathsf{TUMPL}}$ of TUMPL.

Definition 2 (Language). *Let $\mathbf{S}$ denote a set of sentential variables, $\Box$ a necessity operator, $\mathbf{SPR}$ a strict preference relation symbol. The language $\mathcal{L}_{\mathsf{TUMPL}}$ of TUMPL is given by the following rule:*

$$\varphi ::= s \mid \top \mid \neg\varphi \mid \varphi_1 \& \varphi_2 \mid \Box\varphi \mid \mathbf{SPR}(\varphi_1, \varphi_2),$$

[3] We learned from van Rooij ([13]) the relation between the Sorites Paradox and bounded rationality.

where $s \in \mathbf{S}$, and nestings of **SPR** *do not occur.* $\perp, \vee, \rightarrow, \leftrightarrow$ *and* $\Diamond$ *are introduced by the standard definitions. We define an indifference relation symbol* **IND** *and a weak preference relation symbol* **WPR** *as follows:*

$$\mathbf{IND}(\varphi_1, \varphi_2) := \neg\mathbf{SPR}(\varphi_1, \varphi_2) \& \neg\mathbf{SPR}(\varphi_2, \varphi_1),$$
$$\mathbf{WPR}(\varphi_1, \varphi_2) := \mathbf{SPR}(\varphi_1, \varphi_2) \vee \mathbf{IND}(\varphi_1, \varphi_2).$$

The set of all well-formed formulae of $\mathcal{L}_{\mathsf{TUMPL}}$ *will be denoted by* $\Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$. ■

Remark 2. We introduce $\Box$ to construct a Boolean algebra of subsets of $\mathbf{W}$ that is accessible from $w \in \mathbf{W}$. ■

3.2 Semantics

We define a structured Kripke model $\mathcal{M}$ for TUMPL.

Definition 3 (Model). *$\mathcal{M}$ is a quadruple $(\mathbf{W}, R, V, \rho)$, where:*

- *$\mathbf{W}$ is a nonempty set of possible worlds,*
- *R is a binary relation on $\mathbf{W}$,*
- *V is a truth assignment to each $s \in \mathbf{S}$ for each $w \in \mathbf{W}$,*
- *ρ is a preference space assignment that assigns to each $w \in \mathbf{W}$ a preference space $(\mathcal{F}_w, \succ_w)$ such that $\mathcal{F}_w$ is a Boolean algebra of subsets of $\{w' \in \mathbf{W} : R(w, w')\}$ and $\succ_w$ on $\mathcal{F}_w$ is a semiorder.*

■

Since A is an *arbitrary* finite set, the next corollary follows directly from Theorem 2.

Corollary 1 (Representation on Finite Boolean Algebra). *Suppose that $\mathbf{W}$ is a finite set of possible worlds and $\mathcal{F}$ is a finite Boolean algebra of subsets of $\mathbf{W}$ and $\succ$ is a binary relation on $\mathcal{F}$, and δ is a positive number. Then $\succ$ is a semiorder iff there is a function $u : \mathcal{F} \rightarrow \mathbb{R}$ such that for any $\alpha, \beta \in \mathcal{F}$,*

$$\alpha \succ \beta \text{ iff } u(\alpha) > u(\beta) + \delta.$$

■

Remark 3. Corollary 1, together with Corollary 2 described later, can guarantee that $\succ_w$ on $\mathcal{F}_w$ is a threshold utility maximiser's preference relation. ■

We provide TUMPL with the following truth definition relative to $\mathcal{M}$:

Definition 4 (Truth). *The notion of $\varphi \in \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$ being true at $w \in W$ in $\mathcal{M}$, in symbols $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi$ is inductively defined as follows:*

- $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} s$ *iff* $V(w)(s) = \mathbf{true}$,
- $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \top$,
- $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi_1 \& \varphi_2$ *iff* $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi_1$ *and* $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi_2$,
- $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \neg\varphi$ *iff* $(\mathcal{M}, w) \not\models_{\mathsf{TUMPL}} \varphi$,
- $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \Box\varphi$ *iff, for any w' such that $R(w, w')$,* $(\mathcal{M}, w') \models_{\mathsf{TUMPL}} \varphi$,
- $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \mathbf{SPR}(\varphi_1, \varphi_2)$ *iff* $[\![\varphi_1]\!]^{\mathcal{M}}_w \succ_w [\![\varphi_2]\!]^{\mathcal{M}}_w$,

where $[\![\varphi]\!]^{\mathcal{M}}_{w} := \{w' \in \mathbf{W} : R(w, w')$ *and* $(\mathcal{M}, w') \models_{\mathsf{TUMPL}} \varphi\}$. *If* $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi$ *for all* $w \in \mathbf{W}$, *we write* $\mathcal{M} \models_{\mathsf{TUMPL}} \varphi$ *and say that* φ *is valid in* $\mathcal{M}$. *If* φ *is valid in all structured Kripke models for* TUMPL, *we write* $\models_{\mathsf{TUMPL}} \varphi$ *and say that* φ *is valid.* ■

We would like to provide a *counter-model* of the transitivity of indifference. We now return to Example 1. Assume that $\mathcal{U} := (\mathbf{W}, R, V, \rho)$ is given, where:

- $\mathbf{W} := \{w_0, \ldots, w_{400}\}$, where w_i is a possible world in which you try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar, for any i $(0 \leq i \leq 400)$,
- R is a binary relation on $\mathbf{W}$,
- V is a truth assignment to each $s \in \mathbf{S}$ for each $w_i \in \mathbf{W}$,
- ρ is a preference space assignment that assigns to each $w_i \in \mathbf{W}$ a preference space $(\mathcal{F}_{w_i}, \succ_{w_i})$, where:
  - $\mathcal{F}_{w_i}$ is a Boolean algebra of subsets of $\{w'_i \in \mathbf{W} : R(w_i, w'_i)\}$,
  - $\succ_{w_i}$ on $\mathcal{F}_{w_i}$ is a semiorder,
  - $\{w_j\} \sim_{w_i} \{w_{j+1}\}$, for any j $(0 \leq j \leq 400)$,
  - $\{w_0\} \nsim_{w_i} \{w_{400}\}$.

Let φ_i denote the sentence "You try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar", for any i $(0 \leq i \leq 400)$. Then we have, for any i $(0 \leq i \leq 400)$,

$$[\![\varphi_j]\!]^{\mathcal{U}}_{w_i} \sim_{w_i} [\![\varphi_{j+1}]\!]^{\mathcal{U}}_{w_i}, \text{ for any } j \ (0 \leq j \leq 400),$$
$$[\![\varphi_0]\!]^{\mathcal{U}}_{w_i} \nsim_{w_i} [\![\varphi_{400}]\!]^{\mathcal{U}}_{w_i},$$

for $[\![\varphi_j]\!]^{\mathcal{U}}_{w_i} = \{w_j\}$ holds for any j $(0 \leq j \leq 400)$. It must be noted that, for any i $(0 \leq i \leq 400)$, because $\succ_{w_i}$ on $\mathcal{F}_{w_i}$ is a semiorder, $[\![\varphi_j]\!]^{\mathcal{U}}_{w_i} \sim_{w_i} [\![\varphi_{j+1}]\!]^{\mathcal{U}}_{w_i}$ for any j $(0 \leq j \leq 400)$ does not imply $[\![\varphi_0]\!]^{\mathcal{U}}_{w_i} \sim_{w_i} [\![\varphi_{400}]\!]^{\mathcal{U}}_{w_i}$. So we have, for any i $(0 \leq i \leq 400)$,

$$(\mathcal{U}, w_i) \not\models_{\mathsf{TUMPL}} (\mathbf{IND}(\varphi_0, \varphi_1) \& \cdots \& \mathbf{IND}(\varphi_{399}, \varphi_{400})) \rightarrow \mathbf{IND}(\varphi_0, \varphi_{400}).$$

Therefore we obtain the following proposition.

Proposition 1 (Nontransitivity of Indifference).

$$\not\models_{\mathsf{TUMPL}} (\mathbf{IND}(\varphi_0, \varphi_1) \& \cdots \& \mathbf{IND}(\varphi_{399}, \varphi_{400})) \rightarrow \mathbf{IND}(\varphi_0, \varphi_{400}).$$

■

3.3 Syntax

We provide TUMPL with a proof system.

Definition 5 (Proof System). *The proof system of* TUMPL *consists of the following:*

1. *all tautologies of classical sentential logic,*
2. $\Box(\varphi_1 \rightarrow \varphi_2) \rightarrow (\Box\varphi_1 \rightarrow \Box\varphi_2)$ $\quad (K)$,

6

3. $\Box(\varphi_1 \leftrightarrow \varphi_2) \& \Box(\psi_1 \leftrightarrow \psi_2) \to (\mathbf{SPR}(\varphi_1, \psi_1) \leftrightarrow \mathbf{SPR}(\varphi_2, \psi_2))$
(Replacement of Necessary Equivalents),
4. $\neg\mathbf{SPR}(\varphi, \varphi)$
(Syntactic Counterpart of Irreflexivity),
5. $(\mathbf{SPR}(\varphi_1, \varphi_2) \wedge \mathbf{SPR}(\varphi_3, \varphi_4)) \to (\mathbf{SPR}(\varphi_1, \varphi_4) \vee \mathbf{SPR}(\varphi_3, \varphi_2))$
(Syntactic Counterpart of Intervality),
6. $(\mathbf{SPR}(\varphi_1, \varphi_2) \wedge \mathbf{SPR}(\varphi_2, \varphi_3)) \to (\mathbf{SPR}(\varphi_1, \varphi_4) \vee \mathbf{SPR}(\varphi_4, \varphi_3))$
(Syntactic Counterpart of Semitransitivity),
7. *Modus Ponens,*
8. *Necessitation.*

A proof of $\varphi \in \Phi_{\mathsf{TUMPL}}$ is a finite sequence of $\mathcal{L}_{\mathsf{TUMPL}}$-formulae having φ as the last formula such that either each formula is an instance of an axiom, or it can be obtained from formulae that appear earlier in the sequence by applying an inference rule. If there is a proof of φ, we write $\vdash_{\mathsf{TUMPL}} \varphi$. ■

3.4 Metalogic

We prove the metatheorems of TUMPL. It is easy to prove the soundness of TUMPL.

Theorem 3 (Soundness). *For any $\varphi \in \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$, if $\vdash_{\mathsf{TUMPL}} \varphi$, then $\models_{\mathsf{TUMPL}} \varphi$.* ■

We now turn to the task of proving the completeness of TUMPL. We prove it by developing the idea of Segerberg ([11]) that we modify *filtration theory* in such a way that completeness can be established by *Corollary 1 (Representation on Finite Boolean Algebra)*. We cannot go into details because of limited space. The outline of the proof is as follows. We begin by defining some new concepts.

Definition 6 (Stuffedness). *Suppose that Θ is a set of formulae such that Θ is closed under subformulae and $\bot \in \Theta$. Let*

$$\Delta := \{\varphi : \textit{for some } \psi,\ \mathbf{SPR}(\varphi, \psi) \textit{ or } \mathbf{SPR}(\psi, \varphi)\},$$

and let Δ' be the closure of Δ under Boolean compounds. If Θ satisfies also the condition that $\mathbf{SPR}(\varphi, \psi) \in \Theta$, for any $\varphi, \psi \in \Delta'$, we say that Θ is stuffed. ■

Definition 7 (Value Formula). *The formulae in Δ' are called the value formulae of Θ.* ■

Remark 4. There is no occurrence of **SPR** in value formulae. ■

Definition 8 (Base). *We say that $\Psi_0 \subseteq \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$ is a base (with respect to* TUMPL*) for $\Psi \subseteq \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$ if for any $\varphi \in \Psi$ there is some $\varphi_0 \in \Psi_0$ such that $\vdash_{\mathsf{TUMPL}} \varphi \leftrightarrow \varphi_0$.* ■

Definition 9 (Logical Finiteness). *We say that Ψ is logically finite (with respect to* TUMPL*) if there is a finite base for Ψ.* ■

Remark 5. A set has finite base if it has a logically finite one. ■

Lemma 1 (Logical Finiteness). *If $\Psi \subseteq \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$ is a finite set closed under subformulae, and if Θ is the smallest stuffed superset of Ψ, then Θ is logically finite.* ■

Definition 10 (Maximal Consistency). *A finite set $\{\varphi_1, \ldots, \varphi_n\} \subseteq \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$ is* TUMPL*-consistent iff $\nvdash_{\mathsf{TUMPL}} \neg(\varphi_1 \& \ldots \& \varphi_n)$. An infinite set of formulae is* TUMPL*-consistent iff all of its finite subsets are* TUMPL*-consistent. $\Gamma \subseteq \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$ is a* TUMPL*-maximal consistent set iff it is* TUMPL*-consistent and for any $\varphi \notin \Gamma$, $\Gamma \cup \{\varphi\}$ is* TUMPL*-inconsistent.* ■

Definition 11 (Canonical Model for Modal-Logical Part of TUMPL). *We define $\mathcal{U}^C := (\mathbf{W}^C, R^C, V^C)$ as the canonical model for the modal-logical part of* TUMPL *where*

- $\mathbf{W}^C := \{\Gamma \subseteq \Phi_{\mathcal{L}_{\mathsf{TUMPL}}} : \Gamma$ *is* TUMPL*-maximal consistent*$\}$,
- *for any $\Gamma, \Delta \in \mathbf{W}^C$, $R^C(\Gamma, \Delta)$ iff for any $\varphi \in \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$, if $\Box\varphi \in \Gamma$, then $\varphi \in \Delta$.*
- *for any $\Gamma \in \mathbf{W}^C$,*

$$V^C(\Gamma)(s) := \begin{cases} \textbf{\textit{true}} & \text{if } s \in \Gamma, \\ \textbf{\textit{false}} & \text{otherwise.} \end{cases}$$

■

Definition 12 ($\mathcal{U}$). *We define $\mathcal{U} := (\mathbf{W}, R, V)$ by means of $\mathcal{U}^C$ as follows:*

- *$\mathbf{W}$ is a fixed subset of $\mathbf{W}^C$, either $\mathbf{W}^C$ itself or else $\{\Delta : \Gamma = \Delta$ or $R^{C*}(\Gamma, \Delta)\}$ for some $\Gamma \in \mathbf{W}^C$, where R^{C*} is the ancestral of R^C.*
- *R and V are the restrictions of R^C and V^C to $\mathbf{W}$.*

■

Definition 13 (Equivalence Class). *Let Θ be a stuffed set of formulae that are logically finite with respect to* TUMPL*. We define, for $\Gamma, \Delta \in \mathbf{W}$,*

$$\Gamma \equiv \Delta \text{ iff } \Gamma \cap \Theta = \Delta \cap \Theta.$$

Then $\equiv$ is an equivalence relation on $\mathbf{W}$. We write $[\Gamma]$ for the equivalence class of Γ. ■

Definition 14 (Filtration). *We define $\mathcal{U}^{\equiv} := (\mathbf{W}^{\equiv}, R^{\equiv}, V^{\equiv})$ as a filtration of $\mathcal{U}$ where*

- $\mathbf{W}^{\equiv} := \{[\Gamma] : \Gamma \in \mathbf{W}\}$,
- *$R^{\equiv}$ is a binary relation on $\mathbf{W}^{\equiv}$ such that*
  1. *if $R(\Gamma, \Delta)$, then $R^{\equiv}([\Gamma], [\Delta])$.*
  2. *if $R^{\equiv}([\Gamma], [\Delta])$ and $\Box\varphi \in \Gamma$, then $\varphi \in \Delta$.*

3. $V^{\equiv}$ *is a function such that for any* $s \in \Theta$,

$$V^{\equiv}([\Gamma])(s) = V(\Gamma)(s).$$

■

Thus, for any $\xi \in \mathbf{W}^{\equiv}$,

$$[\![\varphi]\!]_{\xi}^{\mathcal{U}^{\equiv}} := \{\eta : R^{\equiv}(\xi, \eta) \text{ and } (\mathcal{U}^{\equiv}, \eta) \models_{\mathsf{TUMPL}} \varphi\}$$

is well-defined for any φ which does not contain **SPR**.

Lemma 2 (Lindenbaum). *Every* TUMPL*-consistent set of formulae is a subset of a* TUMPL*-maximal consistent set of formulae.* ■

Lemma 3 (Partial Truth). *If* $\varphi \in \Theta$ *and* φ *does not contain* **SPR**, *then for any* $\Gamma \in \mathbf{W}$,

$$(\mathcal{U}^{\equiv}, [\Gamma]) \models_{\mathsf{TUMPL}} \varphi \text{ iff } \varphi \in \Gamma.$$

■

We wish to supplement $\mathcal{U}^{\equiv}$ with preference space assignment $\rho^{\equiv}$ so as to obtain a structured Kripke model $\mathcal{U}_{\sharp}^{\equiv}$ for which Truth Lemma holds for all formulae in Θ. Doing this contributes to solving the completeness problem of TUMPL.

Definition 15 ($\mathcal{F}_{\xi}^{\equiv}$)**.** *For any* $\xi \in \mathbf{W}^{\equiv}$, *we define* $\mathcal{F}^{\equiv}$ *as the set of all* $\alpha \subseteq \{\eta : R^{\equiv}(\xi, \eta)\}$ *such that for some value formula* $\varphi \in \Theta$, $\alpha = [\![\varphi]\!]_{\xi}^{\mathcal{U}^{\equiv}}$. ■

Lemma 4 (Boolean Algebra). *For any* $\xi \in \mathbf{W}^{\equiv}$, $\mathcal{F}_{\xi}^{\equiv}$ *is a Boolean algebra with* $\{\eta : R^{\equiv}(\xi, \eta)\}$ *as unit element.* ■

Definition 16 ($\succ_{\xi}$)**.** *For any* $\xi \in \mathbf{W}^{\equiv}$ *we define* $\alpha \succ_{\xi} \beta$ *to hold between elements* $\alpha, \beta \in \mathcal{F}_{\xi}^{\equiv}$ *iff there are value formulae* $\varphi, \psi \in \Theta$ *such that* $\alpha = [\![\varphi]\!]_{\xi}^{\mathcal{U}^{\equiv}}$, $\beta = [\![\psi]\!]_{\xi}^{\mathcal{U}^{\equiv}}$ *and* $\mathbf{SPR}(\varphi, \psi) \in \Gamma$ *for any* $\Gamma \in \xi$. ■

Lemma 5 ($\succ_{\xi}$ **and SPR**)**.** *For any value formula* $\varphi, \psi \in \Theta$ *and any* $\xi \in \mathbf{W}^{\equiv}$, $[\![\varphi]\!]_{\xi}^{\mathcal{U}^{\equiv}} \succ_{\xi} [\![\psi]\!]_{\xi}^{\mathcal{U}^{\equiv}}$ *iff, for any* $\Gamma \in \xi$, $\mathbf{SPR}(\varphi, \psi) \in \Gamma$. ■

Lemma 6 (Irreflexivity, Intervality and Semitransitivity). *For any* $\xi \in \mathbf{W}^{\equiv}$, $\succ_{\xi}$ *on* $\mathcal{F}_{\xi}^{\equiv}$ *satisfies Irreflexivity, Intervality and Semitransitivity.* ■

Since we assumed that Θ is logically finite, $\mathbf{W}^{\equiv}$ is *finite*. Hence for any $\xi \in \mathbf{W}^{\equiv}$, $\mathcal{F}_{\xi}^{\equiv}$ is *finite*, so the next corollary follows from Lemma 4 and Lemma 6.

Corollary 2. *For any* $\xi \in \mathbf{W}^{\equiv}$, *Corollary 1 is provable in* $(\mathcal{F}_{\xi}^{\equiv}, \succ_{\xi})$. ■

Definition 17. *We define* $\mathcal{U}_{\sharp}^{\equiv} := (\mathbf{W}^{\equiv}, R^{\equiv}, V^{\equiv}, \rho^{\equiv})$ *as a structured Kripke model for preference where* $\rho^{\equiv}$ *is a preference space assignment that assigns to each* $\xi \in \mathbf{W}^{\equiv}$ $(\mathcal{F}_{\xi}^{\equiv}, \succ_{\xi})$. ■

Lemma 7 (Full Truth). *For any $\varphi \in \Theta$ and any $\Gamma \in \mathbf{W}$,*

$$(\mathcal{U}^{\equiv}_{\sharp}, [\Gamma]) \models_{\mathsf{TUMPL}} \varphi \text{ iff } \varphi \in \Gamma.$$

■

Remark 6. This lemma is the announced improvement of Lemma 3. ■

Theorem 4 (Completeness). *For any $\varphi \in \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$, if $\models_{\mathsf{TUMPL}} \varphi$, then $\vdash_{\mathsf{TUMPL}} \varphi$.* ■

Proof. Suppose that $\nvdash_{\mathsf{TUMPL}} \varphi_0$. Then $\{\neg\varphi_0\}$ is a TUMPL-consistent set. By Lemma 2 $\{\neg\varphi_0\}$ is a subset of a TUMPL-maximal consistent set Γ. Evidently $\varphi_0 \notin \Gamma$. Let Ψ be the set of subformulae of TUMPL that is finite and let Θ be the smallest stuffed extension of Ψ. By Lemma 1 Θ is logically finite with respect to TUMPL. Let $\mathbf{W}$ be as mentioned above, either $\mathbf{W}^C$ itself or else $\{\Delta : \Gamma = \Delta \text{ or } R^{C*}(\Gamma, \Delta)\}$ (where R^{C*} is the ancestral of R^C). By Lemma 7 $(\mathcal{U}^{\equiv}_{\sharp}, [\Gamma]) \not\models_{\mathsf{TUMPL}} \varphi$. Therefore $\not\models_{\mathsf{TUMPL}} \varphi$. ■

We can prove the decidability of TUMPL as follows.

Lemma 8 (Finite Model Property). TUMPL *has the finite model property that every non-theorem of* TUMPL *fails in a structured Kripke model for preference with only finitely many elements.* ■

Theorem 5 (Decidability). TUMPL *is decidable.* ■

Proof. Suppose that φ is not provable in TUMPL. By Lemma 8 φ fails in a structured Kripke model $\mathcal{U}^{\equiv}_{\sharp}$ for preference with only finitely many elements. If we take a domain $\mathbf{W}^{\equiv}$ with at most that number of elements, there are only finitely many ways in which accessibility relations of and truth assignments can be defined, and there are also only finite many ways to define the preference space assignment $\rho^{\equiv}$. Finally for any $\xi \in \mathbf{W}^{\equiv}$, there are only finitely many ways in which $\succ_{\xi}$ can be defined. To decide, of a given relation $\succ_{\xi}$, whether it satisfies Irreflexivity, Intervality and Semitransitivity can be done in finitely many steps. So we can find, in at most a finite number of steps, a counter-model of the unprovable formula. In fact, we can compute an upper bound to the number of steps needed. Thus TUMPL is decidable. ■

4 Concluding Remarks

In this paper we have proposed a new version of complete and decidable extrinsic preference logic–threshold utility maximiser's preference logic (TUMPL) that can solve the Nontransitivity Problem and avoid the Fundamental Problem of Intrinsic Preference.

References

1. Armstrong, E. W.: The Determinateness of the Utility Function. Economic Journal **49** (1939) 453–467.
2. Cantor, G.: Beiträge zur Begründung der Transfiniten Mengenlehre I. Mathematische Annalen **46** (1895) 481–512.
3. Chisholm, R. M. and Sosa, E.: On the Logic of Intrinsically Better. American Philosophical Quarterly **3** (1966) 244–249.
4. Keefe, R.: Theories of Vagueness. Cambridge UP, Cambridge (2000).
5. Luce, D.: Semiorders and a Theory of Utility Discrimination. Econometrica **24** (1956) 178–191.
6. Martin, R. M.: Intension and Decision. Prentice-Hall, Inc., Englewood Cliffs (1963).
7. Mullen, J. D.: Does the Logic of Preference Rest on a Mistake?. Metaphilosophy **10** (1979) 247–255.
8. Roberts, F. S.: Measurement Theory. Addison-Wesley, Reading (1979).
9. Scott, D. and Suppes, P.: Foundational Aspects of Theories of Measurement. Journal of Symbolic Logic **3** (1958) 113–128.
10. Scott, D.: Measurement Structures and Linear Inequalities. Journal of Mathematical Psychology **1** (1964) 233–247.
11. Segerberg, K.: Qualitative Probability in a Modal Setting. In: Fenstad, J. E. (ed.): Proceedings of the Second Scandinavian Logic Symposium. North-Holland, Amsterdam (1971) 341–352.
12. Simon, H. A.: Models of Bounded Rationality. MIT Press, Cambridge, Mass. (1982).
13. Van Rooij, R.: Revealed Preference and Satisficing Behavior. In: Accepted Papers of 8th Conference on Logic and the Foundations of Game and Decision Theory (LOFT 2008).
14. Von Wright, G. H.: The Logic of Preference. Edinburgh UP, Edinburgh (1963).
15. Von Wright, G. H.: The Logic of Preference Reconsidered. Theory and Decision **3** (1972) 140–169.

Measurement-Theoretic Foundation of Threshold Utility Maximiser's Preference Logic

SATORU SUZUKI, PhD.
bxs05253@nifty.com

Faculty of Arts and Sciences, Komazawa University, JAPAN,
Faculty of Political Science and Economics, Meiji University, JAPAN

SOCREAL 2010
27 March 2010, Hokkaido University

Structure of This Talk

Structure of This Talk

Counterexample to Transitivity of Indifference (1)

- The economist Armstrong ([Armstrong 1939]) was one of the first to argue that indifference is not always transitive.

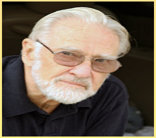
Armstrong, E. W.:
The Determinateness of the Utility Function.
Economic Journal **49** (1939) 453–467.

Counterexample to Transitivity of Indifference (1)

- The economist Armstrong ([Armstrong 1939]) was one of the first to argue that indifference is not always transitive.

Armstrong, E. W.:
The Determinateness of the Utility Function.
Economic Journal **49** (1939) 453–467.

- Luce ([[Luce 1956]: 179]) gave the following counterexample to the transitivity of indifference:



Luce, D.:
Semiorders and a Theory of Utility Discrimination.
Econometrica **24** (1956) 178–191.

Counterexample to Transitivity of Indifference (2)

Example (Avoidance of Sorites Paradox in Preference)

- If indifference were transitive, then he would be unable to detect any weight differences, however great, which is patently false. . . .

Counterexample to Transitivity of Indifference (2)

Example (Avoidance of Sorites Paradox in Preference)

- If indifference were transitive, then he would be unable to detect any weight differences, however great, which is patently false....
- Find a subject who prefers a cup of coffee with one cube of sugar to one with five cubes....

Counterexample to Transitivity of Indifference (2)

Example (Avoidance of Sorites Paradox in Preference)

- If indifference were transitive, then he would be unable to detect any weight differences, however great, which is patently false....
- Find a subject who prefers a cup of coffee with one cube of sugar to one with five cubes....
- Now prepare 401 cups of coffee with $(1 + \frac{i}{100})x$ grams of sugar, $i = 0, 1, \ldots, 400$, where x is the weight of one cube of sugar.

Counterexample to Transitivity of Indifference (2)

Example (Avoidance of Sorites Paradox in Preference)

- If indifference were transitive, then he would be unable to detect any weight differences, however great, which is patently false. . . .
- Find a subject who prefers a cup of coffee with one cube of sugar to one with five cubes. . . .
- Now prepare 401 cups of coffee with $(1 + \frac{i}{100})x$ grams of sugar, $i = 0, 1, \ldots, 400$, where x is the weight of one cube of sugar.
- It is evident that he will be indifferent between cup i and cup $i + 1$, for any i, but by choice he is not indifferent between $i = 0$ and $i = 400$.

■

Counterexample to Transitivity of Indifference (2)

Example (Avoidance of Sorites Paradox in Preference)

- If indifference were transitive, then he would be unable to detect any weight differences, however great, which is patently false. . . .
- Find a subject who prefers a cup of coffee with one cube of sugar to one with five cubes. . . .
- Now prepare 401 cups of coffee with $(1 + \frac{i}{100})x$ grams of sugar, $i = 0, 1, \ldots, 400$, where x is the weight of one cube of sugar.
- It is evident that he will be indifferent between cup i and cup $i + 1$, for any i, but by choice he is not indifferent between $i = 0$ and $i = 400$.

■

This example shows a situation where we would face the Sorites Paradox in preference if indifference were transitive.

Nontransitivity Problem

The first problem now arises:

Nontransitivity Problem

The first problem now arises:

Problem (Nontransitivity Problem)

What kind of preference logic can formalise inferences in which indifference is not transitive? ■

Nontransitivity Problem

The first problem now arises:

Problem (Nontransitivity Problem)

What kind of preference logic can formalise inferences in which indifference is not transitive? ■

We call it the Nontransitivity Problem.

The Aim of This Talk

The aim of this talk is to propose a new version of complete and decidable extrinsic preference logic–threshold utility maximiser's preference logic (TUMPL) that can solve the Nontransitivity Problem.

Fundamental Problem of Intrinsic Preference (1)

- Generally, preference logics are in danger of inviting a serious problem.

Fundamental Problem of Intrinsic Preference (1)

- Generally, preference logics are in danger of inviting a serious problem.
- Von Wright divided preferences into two categories: extrinsic and intrinsic preference.

Von Wright, G. H.:
The Logic of Preference.
Edinburgh UP, Edinburgh (1963).

Fundamental Problem of Intrinsic Preference (2)

- An agent is said to prefer φ extrinsically to ψ if φ is better than ψ in some explicit respect. So we can explain extrinsic preference from some explicit point of view.

Fundamental Problem of Intrinsic Preference (2)

- An agent is said to prefer φ extrinsically to ψ if φ is better than ψ in some explicit respect. So we can explain extrinsic preference from some explicit point of view.
- If we cannot explain preference from any explicit point of view, we call it intrinsic.

Fundamental Problem of Intrinsic Preference (3)

- Von Wright posed the following fundamental problem intrinsic preference logics faced.

Von Wright, G. H.:
The Logic of Preference Reconsidered.
Theory and Decision **3** (1972) 140–169.

Fundamental Problem of Intrinsic Preference (3)

- Von Wright posed the following fundamental problem intrinsic preference logics faced.

Von Wright, G. H.:
The Logic of Preference Reconsidered.
Theory and Decision **3** (1972) 140–169.

Problem (Fundamental Problem of Intrinsic Preference)

The development of a satisfactory logic of preference has turned out to be unexpectedly problematic. The evidence for this lies in the fact that almost every principle which has been proposed as fundamental to one preference logic has been rejected by another one. ■

Fundamental Problem of Intrinsic Preference (3)

- Von Wright posed the following fundamental problem intrinsic preference logics faced.

Von Wright, G. H.:
The Logic of Preference Reconsidered.
Theory and Decision **3** (1972) 140–169.

Problem (Fundamental Problem of Intrinsic Preference)

The development of a satisfactory logic of preference has turned out to be unexpectedly problematic. The evidence for this lies in the fact that almost every principle which has been proposed as fundamental to one preference logic has been rejected by another one. ■

- We call it the Fundamental Problem of Intrinsic Preference.

Fundamental Problem of Intrinsic Preference (4)

Example

	von Wright	Martin	Chisholm and Sosa
Transitivity	$+$	$+$	$+$
Contraposition	$-$	$+$	$-$
Conjunctive Expansion	$+$	$-$	$-$
Disjunctive Distribution	$-$	$-$	$-$
Conjunctive Distribution	$+$	$-$	$-$

Chisholm, R. M. and Sosa, E.:
On the Logic of "Intrinsically Better".
American Philosophical Quarterly **3** (1966) 244–249.

Martin, R. M.:
Intension and Decision.
Prentice-Hall, Inc., Englewood Cliffs (1963).

Mullen's Analysis of Cause of Fundamental Problem (1)

- According to Mullen ([Mullen 1979]), we can analyse the cause of the Fundamental Problem as follows.

Mullen, J. D.:
Does the Logic of Preference Rest on a Mistake?.
Metaphilosophy **10** (1979) 247–255.

Mullen's Analysis of Cause of Fundamental Problem (1)

- According to Mullen ([Mullen 1979]), we can analyse the cause of the Fundamental Problem as follows.

Mullen, J. D.:
Does the Logic of Preference Rest on a Mistake?.
Metaphilosophy **10** (1979) 247–255.

- The adequacy criteria for intrinsic preference principles considered by preference logicians have been whether the principles are consistent with their intuitions of reasonableness.

Mullen's Analysis of Cause of Fundamental Problem (1)

- According to Mullen ([Mullen 1979]), we can analyse the cause of the Fundamental Problem as follows.

Mullen, J. D.:
Does the Logic of Preference Rest on a Mistake?.
Metaphilosophy **10** (1979) 247–255.

- The adequacy criteria for intrinsic preference principles considered by preference logicians have been whether the principles are consistent with their intuitions of reasonableness.
- But each intuitions often disagrees even on the fundamental properties.

Mullen's Analysis of Cause of Fundamental Problem (2)

- Different theories, such as ethics, welfare economics, consumer demand theory, game theory and decision theory make different demands upon the fundamental properties of preference.

Mullen's Analysis of Cause of Fundamental Problem (2)

- Different theories, such as ethics, welfare economics, consumer demand theory, game theory and decision theory make different demands upon the fundamental properties of preference.
- So if we would like to propose preference logic that can avoid the Fundamental Problem of Intrinsic Preference, it should be constructed not from intuition but from a theory or a rule in a theory, that is, it should be extrinsic.

Measurement Theory (1)

- In order to avoid the Fundamental Problem of Intrinsic Preference, we resort to measurement theory.

Measurement Theory (1)

- In order to avoid the Fundamental Problem of Intrinsic Preference, we resort to measurement theory.
- There are two main problems with measurement theory:
  1. the representation problem–justifying the assignment of numbers to objects or propositions,

Measurement Theory (1)

- In order to avoid the Fundamental Problem of Intrinsic Preference, we resort to measurement theory.
- There are two main problems with measurement theory:
  1. the representation problem–justifying the assignment of numbers to objects or propositions,
  2. the uniqueness problem–specifying the transformation up to which this assignment is unique.

Measurement Theory (2)

- A solution to the first problem can be furnished by a representation theorem, which establishes that the chosen numerical system preserves the relations of the relational system.

Measurement Theory (2)

- A solution to the first problem can be furnished by a representation theorem, which establishes that the chosen numerical system preserves the relations of the relational system.
- When we provide TUMPL with a model based on semiorders, by virtue of a corollary of the Scott-Suppes representation theorem, we can adopt threshold utility maximisation as a rule in utility theory that makes demands upon the fundamental properties of preference, which enables TUMPL to avoid the Fundamental Problem of Intrinsic Preference.

Structure of This Talk

Global Rationality

- The standard model of economics is based on global rationality that requires an optimising behavior.

Global Rationality

- The standard model of economics is based on global rationality that requires an optimising behavior.
- Utility maximisation is a typical example of an optimising behavior.

Representation Theorem for Utility Maximisation

- Cantor ([Cantor 1895]) proved the representation theorem for utility maximisation.

Cantor, G.:
Beiträge zur Begründung der Transfiniten Mengenlehre I.
Mathematische Annalen **46** (1895) 481–512.

Representation Theorem for Utility Maximisation

- Cantor ([Cantor 1895]) proved the representation theorem for utility maximisation.

Cantor, G.:
Beiträge zur Begründung der Transfiniten Mengenlehre I.
Mathematische Annalen **46** (1895) 481–512.

Theorem (Representation for Utility Maximisation, Cantor ([Cantor 1895]))

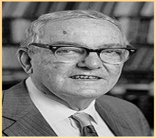
Suppose $\mathbf{A}$ *is a countable set and* $\succeq$ *is a binary relation on* $\mathbf{A}$*. Then* $\succeq$ *is a weak order (transitive and connected) iff there is a function* $u : \mathbf{A} \to \mathbb{R}$ *such that for any* $x, y \in \mathbf{A}$*,*

$$x \succeq y \text{ iff } u(x) \geq u(y).$$

■

Bounded Rationality

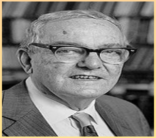
- But according to Simon ([Simon 1982]), cognitive and information-processing constrains on the capabilities of agents, together with the complexity of their environment, render an optimising behavior an unattainable ideal.



Simon, H. A.:
Models of Bounded Rationality.
MIT Press, Cambridge, Mass. (1982).

Bounded Rationality

- But according to Simon ([Simon 1982]), cognitive and information-processing constrains on the capabilities of agents, together with the complexity of their environment, render an optimising behavior an unattainable ideal.



Simon, H. A.:
Models of Bounded Rationality.
MIT Press, Cambridge, Mass. (1982).

- He dismissed the idea that agents should exhibit global rationality and suggested that they in fact exhibit bounded rationality that allows a satisficing behavior.

Semiorders (1)

- One explanation for Example 1 is that the nontransitivity of indifference results from the fact that we cannot generally discriminate very close quantities.

Semiorders (1)

- One explanation for Example 1 is that the nontransitivity of indifference results from the fact that we cannot generally discriminate very close quantities.
- The concept of a semiorder was introduced by Luce ([Luce 1956]) to construct a model to interpret situations like Example 1 of nontransitive indifference with a threshold of discrimination.

Semiorders (2)

- Scott and Suppes defined ([[Scott and Suppes 1958]: 117]) a semiorder as follows:

Scott, D. and Suppes, P.:
Foundational Aspects of Theories of Measurement.
Journal of Symbolic Logic **3** (1958) 113–128.

Semiorders (2)

- Scott and Suppes defined ([[Scott and Suppes 1958]: 117]) a semiorder as follows:

Scott, D. and Suppes, P.:
Foundational Aspects of Theories of Measurement.
Journal of Symbolic Logic **3** (1958) 113–128.

Definition (Semiorder)

$\succ$ on $\mathbf{A}$ is called a semiorder if, for any $w, x, y, z \in \mathbf{A}$, the following conditions are satisfied:

Semiorders (2)

- Scott and Suppes defined ([[Scott and Suppes 1958]: 117]) a semiorder as follows:

Scott, D. and Suppes, P.:
Foundational Aspects of Theories of Measurement.
Journal of Symbolic Logic **3** (1958) 113–128.

Definition (Semiorder)

$\succ$ on $\mathbf{A}$ is called a semiorder if, for any $w, x, y, z \in \mathbf{A}$, the following conditions are satisfied:

1. $x \not\succ x$. (Irreflexivity),

Semiorders (2)

- Scott and Suppes defined ([[Scott and Suppes 1958]: 117]) a semiorder as follows:

Scott, D. and Suppes, P.:
Foundational Aspects of Theories of Measurement.
Journal of Symbolic Logic **3** (1958) 113–128.

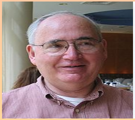
Definition (Semiorder)

$\succ$ on $\mathbf{A}$ is called a semiorder if, for any $w, x, y, z \in \mathbf{A}$, the following conditions are satisfied:

1. $x \not\succ x$. (Irreflexivity),
2. If $w \succ x$ and $y \succ z$, then $w \succ z$ or $y \succ x$. (Intervality),

Semiorders (2)

- Scott and Suppes defined ([[Scott and Suppes 1958]: 117]) a semiorder as follows:




Scott, D. and Suppes, P.:
Foundational Aspects of Theories of Measurement.
Journal of Symbolic Logic **3** (1958) 113–128.

Definition (Semiorder)

$\succ$ on $\mathbf{A}$ is called a semiorder if, for any $w, x, y, z \in \mathbf{A}$, the following conditions are satisfied:

1. $x \not\succ x$. (Irreflexivity),
2. If $w \succ x$ and $y \succ z$, then $w \succ z$ or $y \succ x$. (Intervality),
3. If $w \succ x$ and $x \succ y$, then $w \succ z$ or $z \succ y$. (Semitransitivity).

■

Representation Theorem for Threshold Utility Maximisation

- Threshold utility maximisation can be said to be a typical example of a satisficing behavior(bounded rationality).

Representation Theorem for Threshold Utility Maximisation

- Threshold utility maximisation can be said to be a typical example of a satisficing behavior(bounded rationality).
- Scott and Suppes ([Scott and Suppes 1958]) proved a representation theorem for threshold utility maximisation when **A** is finite.

Representation Theorem for Threshold Utility Maximisation

- Threshold utility maximisation can be said to be a typical example of a satisficing behavior(bounded rationality).
- Scott and Suppes ([Scott and Suppes 1958]) proved a representation theorem for threshold utility maximisation when **A** is finite.

Theorem (Representation for Threshold Utility Maximisation, Scott and Suppes ([Scott and Suppes 1958]))

Suppose that $\succ$ is a binary relation on a finite set $\mathbf{A}$ and δ is a positive number. Then $\succ$ is a semiorder iff there is a function $u : \mathbf{A} \to \mathbb{R}$ such that for any $x, y \in \mathbf{A}$,

$$x \succ y \text{ iff } u(x) > u(y) + \delta.$$

■

Structure of This Talk

1. Introduction
2. Weak Orders as Global Rationality and Semiorders as Bounded Rationality
3. Threshold Utility Maximiser's Preference Logic TUMPL
   - Language of TUMPL
   - Semantics of TUMPL
   - Syntax of TUMPL
   - Metalogic of TUMPL
4. Summary, Further Investigation and Our Related Work

Language of TUMPL

We define the language $\mathcal{L}_{\mathsf{TUMPL}}$ of TUMPL.

Language of TUMPL

We define the language $\mathcal{L}_{\mathsf{TUMPL}}$ of TUMPL.

Definition (Language)

- Let **S** denote a set of sentential variables, $\Box$ a necessity operator, **SPR** a strict preference relation symbol.

Language of TUMPL

We define the language $\mathcal{L}_{\text{TUMPL}}$ of TUMPL.

Definition (Language)

- Let **S** denote a set of sentential variables, $\Box$ a necessity operator, **SPR** a strict preference relation symbol.
- The language $\mathcal{L}_{\text{TUMPL}}$ of TUMPL is given by the following rule:

Language of TUMPL

We define the language $\mathcal{L}_{\text{TUMPL}}$ of TUMPL.

Definition (Language)

- Let **S** denote a set of sentential variables, $\Box$ a necessity operator, **SPR** a strict preference relation symbol.
- The language $\mathcal{L}_{\text{TUMPL}}$ of TUMPL is given by the following rule:

$$\varphi ::= s \mid \top \mid \neg\varphi \mid \varphi_1 \& \varphi_2 \mid \Box\varphi \mid \mathbf{SPR}(\varphi_1, \varphi_2),$$

where $s \in \mathbf{S}$, and nestings of **SPR** do not occur.

Language of TUMPL

We define the language $\mathcal{L}_{\mathsf{TUMPL}}$ of TUMPL.

Definition (Language)

- Let $\mathbf{S}$ denote a set of sentential variables, $\Box$ a necessity operator, **SPR** a strict preference relation symbol.
- The language $\mathcal{L}_{\mathsf{TUMPL}}$ of TUMPL is given by the following rule:

$$\varphi ::= s \mid \top \mid \neg\varphi \mid \varphi_1 \& \varphi_2 \mid \Box\varphi \mid \mathbf{SPR}(\varphi_1, \varphi_2),$$

  where $s \in \mathbf{S}$, and nestings of **SPR** do not occur.
- $\bot, \vee, \rightarrow, \leftrightarrow$ and $\Diamond$ are introduced by the standard definitions.

Language of TUMPL

We define the language $\mathcal{L}_{\mathsf{TUMPL}}$ of TUMPL.

Definition (Language)

- Let $\mathbf{S}$ denote a set of sentential variables, $\Box$ a necessity operator, $\mathbf{SPR}$ a strict preference relation symbol.
- The language $\mathcal{L}_{\mathsf{TUMPL}}$ of TUMPL is given by the following rule:

$$\varphi ::= s \mid \top \mid \neg\varphi \mid \varphi_1 \& \varphi_2 \mid \Box\varphi \mid \mathbf{SPR}(\varphi_1, \varphi_2),$$

  where $s \in \mathbf{S}$, and nestings of $\mathbf{SPR}$ do not occur.
- $\bot, \vee, \rightarrow, \leftrightarrow$ and $\Diamond$ are introduced by the standard definitions.
- We define an indifference relation symbol $\mathbf{IND}$ and a weak preference relation symbol $\mathbf{WPR}$ as follows:

Language of TUMPL

We define the language $\mathcal{L}_{\mathsf{TUMPL}}$ of TUMPL.

Definition (Language)

- Let **S** denote a set of sentential variables, $\Box$ a necessity operator, **SPR** a strict preference relation symbol.
- The language $\mathcal{L}_{\mathsf{TUMPL}}$ of TUMPL is given by the following rule:

$$\varphi ::= s \mid \top \mid \neg\varphi \mid \varphi_1 \& \varphi_2 \mid \Box\varphi \mid \mathbf{SPR}(\varphi_1, \varphi_2),$$

  where $s \in \mathbf{S}$, and nestings of **SPR** do not occur.
- $\bot, \vee, \rightarrow, \leftrightarrow$ and $\Diamond$ are introduced by the standard definitions.
- We define an indifference relation symbol **IND** and a weak preference relation symbol **WPR** as follows:

$$\begin{aligned}\mathbf{IND}(\varphi_1, \varphi_2) &:= \neg\mathbf{SPR}(\varphi_1, \varphi_2) \& \neg\mathbf{SPR}(\varphi_2, \varphi_1),\\ \mathbf{WPR}(\varphi_1, \varphi_2) &:= \mathbf{SPR}(\varphi_1, \varphi_2) \vee \mathbf{IND}(\varphi_1, \varphi_2).\end{aligned}$$

Language of TUMPL

We define the language $\mathcal{L}_{\text{TUMPL}}$ of TUMPL.

Definition (Language)

- Let $\mathbf{S}$ denote a set of sentential variables, $\Box$ a necessity operator, $\mathbf{SPR}$ a strict preference relation symbol.
- The language $\mathcal{L}_{\text{TUMPL}}$ of TUMPL is given by the following rule:

$$\varphi ::= s \mid \top \mid \neg\varphi \mid \varphi_1 \& \varphi_2 \mid \Box\varphi \mid \mathbf{SPR}(\varphi_1, \varphi_2),$$

  where $s \in \mathbf{S}$, and nestings of $\mathbf{SPR}$ do not occur.
- $\bot, \vee, \rightarrow, \leftrightarrow$ and $\Diamond$ are introduced by the standard definitions.
- We define an indifference relation symbol $\mathbf{IND}$ and a weak preference relation symbol $\mathbf{WPR}$ as follows:

$$\mathbf{IND}(\varphi_1, \varphi_2) := \neg\mathbf{SPR}(\varphi_1, \varphi_2) \& \neg\mathbf{SPR}(\varphi_2, \varphi_1),$$
$$\mathbf{WPR}(\varphi_1, \varphi_2) := \mathbf{SPR}(\varphi_1, \varphi_2) \vee \mathbf{IND}(\varphi_1, \varphi_2).$$

- The set of all well-formed formulae of $\mathcal{L}_{\text{TUMPL}}$ will be denoted by $\Phi_{\mathcal{L}_{\text{TUMPL}}}$. $\blacksquare$

The Point of Introducing $\Box$

Remark

We introduce $\Box$ to construct a Boolean algebra of subsets of $\mathbf{W}$ that is accessible from $w \in \mathbf{W}$. $\blacksquare$

Model

We define a structured Kripke model $\mathcal{M}$ for TUMPL.

Model

We define a structured Kripke model $\mathcal{M}$ for TUMPL.

Definition (Model)

$\mathcal{M}$ is a quadruple $(\mathbf{W}, R, V, \rho)$, where:

Model

We define a structured Kripke model $\mathcal{M}$ for TUMPL.

Definition (Model)

$\mathcal{M}$ is a quadruple $(\mathbf{W}, R, V, \rho)$, where:

- $\mathbf{W}$ is a nonempty set of possible worlds,

Model

We define a structured Kripke model $\mathcal{M}$ for TUMPL.

Definition (Model)

$\mathcal{M}$ is a quadruple $(\mathbf{W}, R, V, \rho)$, where:

- $\mathbf{W}$ is a nonempty set of possible worlds,
- R is a binary relation on $\mathbf{W}$,

Model

We define a structured Kripke model $\mathcal{M}$ for TUMPL.

Definition (Model)

$\mathcal{M}$ is a quadruple $(\mathbf{W}, R, V, \rho)$, where:

- $\mathbf{W}$ is a nonempty set of possible worlds,
- R is a binary relation on $\mathbf{W}$,
- V is a truth assignment to each $s \in \mathbf{S}$ for each $w \in \mathbf{W}$,

Model

We define a structured Kripke model $\mathcal{M}$ for TUMPL.

Definition (Model)

$\mathcal{M}$ is a quadruple $(\mathbf{W}, R, V, \rho)$, where:

- $\mathbf{W}$ is a nonempty set of possible worlds,
- R is a binary relation on $\mathbf{W}$,
- V is a truth assignment to each $s \in \mathbf{S}$ for each $w \in \mathbf{W}$,
- ρ is a preference space assignment that assigns to each $w \in \mathbf{W}$ a preference space $(\mathcal{F}_w, \succ_w)$ such that $\mathcal{F}_w$ is a Boolean algebra of subsets of $\{w' \in \mathbf{W} : R(w, w')\}$ and $\succ_w$ on $\mathcal{F}_w$ is a semiorder.

■

Representation on Finite Boolean Algebra

Since A is an arbitrary finite set, the next corollary follows directly from Theorem 2.

Representation on Finite Boolean Algebra

Since A is an arbitrary finite set, the next corollary follows directly from Theorem 2.

Corollary (Representation on Finite Boolean Algebra)

Suppose that $\mathbf{W}$ *is a finite set of possible worlds and* $\mathcal{F}$ *is a finite Boolean algebra of subsets of* $\mathbf{W}$ *and* $\succ$ *is a binary relation on* $\mathcal{F}$*, and* δ *is a positive number. Then* $\succ$ *is a semiorder iff there is a function* $u : \mathcal{F} \to \mathbb{R}$ *such that for any* $\alpha, \beta \in \mathcal{F}$*,*

$$\alpha \succ \beta \text{ iff } u(\alpha) > u(\beta) + \delta.$$

■

Representation on Finite Boolean Algebra

Since A is an arbitrary finite set, the next corollary follows directly from Theorem 2.

Corollary (Representation on Finite Boolean Algebra)

Suppose that $\mathbf{W}$ *is a finite set of possible worlds and* $\mathcal{F}$ *is a finite Boolean algebra of subsets of* $\mathbf{W}$ *and* $\succ$ *is a binary relation on* $\mathcal{F}$*, and* δ *is a positive number. Then* $\succ$ *is a semiorder iff there is a function* $u : \mathcal{F} \to \mathbb{R}$ *such that for any* $\alpha, \beta \in \mathcal{F}$*,*

$$\alpha \succ \beta \textit{ iff } u(\alpha) > u(\beta) + \delta.$$

■

Remark

This corollary, by virtue of filtration theory, can guarantee that $\succ_w$ *on* $\mathcal{F}_w$ *is a threshold utility maximiser's preference relation.* ■

Truth

We provide TUMPL with the following truth definition relative to $\mathcal{M}$:

Truth

We provide TUMPL with the following truth definition relative to $\mathcal{M}$:

Definition (Truth)

The notion of $\varphi \in \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$ being true at $w \in W$ in $\mathcal{M}$, in symbols $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi$ is inductively defined as follows:

Truth

We provide TUMPL with the following truth definition relative to $\mathcal{M}$:

Definition (Truth)

The notion of $\varphi \in \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$ being true at $w \in W$ in $\mathcal{M}$, in symbols $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi$ is inductively defined as follows:

- $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} s \quad \text{iff} \quad V(w)(s) = \mathbf{true}$,
- $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \top$,
- $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi_1 \& \varphi_2$
  iff $\quad (\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi_1$ and $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi_2$,
- $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \neg\varphi \quad \text{iff} \quad (\mathcal{M}, w) \not\models_{\mathsf{TUMPL}} \varphi$,
- $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \Box\varphi$
  iff, $\quad$ for any w' such that $R(w, w')$, $\quad (\mathcal{M}, w') \models_{\mathsf{TUMPL}} \varphi$,

Truth

We provide TUMPL with the following truth definition relative to $\mathcal{M}$:

Definition (Truth)

The notion of $\varphi \in \Phi_{\mathcal{L}_{\mathrm{TUMPL}}}$ being true at $w \in W$ in $\mathcal{M}$, in symbols $(\mathcal{M}, w) \models_{\mathrm{TUMPL}} \varphi$ is inductively defined as follows:

- $(\mathcal{M}, w) \models_{\mathrm{TUMPL}} s$ iff $V(w)(s) = \mathbf{true}$,
- $(\mathcal{M}, w) \models_{\mathrm{TUMPL}} \top$,
- $(\mathcal{M}, w) \models_{\mathrm{TUMPL}} \varphi_1 \& \varphi_2$ iff $(\mathcal{M}, w) \models_{\mathrm{TUMPL}} \varphi_1$ and $(\mathcal{M}, w) \models_{\mathrm{TUMPL}} \varphi_2$,
- $(\mathcal{M}, w) \models_{\mathrm{TUMPL}} \neg\varphi$ iff $(\mathcal{M}, w) \not\models_{\mathrm{TUMPL}} \varphi$,
- $(\mathcal{M}, w) \models_{\mathrm{TUMPL}} \Box\varphi$ iff, for any w' such that $R(w, w')$, $(\mathcal{M}, w') \models_{\mathrm{TUMPL}} \varphi$,
- $(\mathcal{M}, w) \models_{\mathrm{TUMPL}} \mathbf{SPR}(\varphi_1, \varphi_2)$ iff $[\![\varphi_1]\!]^{\mathcal{M}}_w \succ_w [\![\varphi_2]\!]^{\mathcal{M}}_w$,

where $[\![\varphi]\!]^{\mathcal{M}}_w := \{w' \in \mathbf{W} : R(w, w') \text{ and } (\mathcal{M}, w') \models_{\mathrm{TUMPL}} \varphi\}$. ■

Validity

Definition (Validity)

- If $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi$ for all $w \in \mathbf{W}$, we write $\mathcal{M} \models_{\mathsf{TUMPL}} \varphi$ and say that φ is valid in $\mathcal{M}$.

Validity

Definition (Validity)

- If $(\mathcal{M}, w) \models_{\mathsf{TUMPL}} \varphi$ for all $w \in \mathbf{W}$, we write $\mathcal{M} \models_{\mathsf{TUMPL}} \varphi$ and say that φ is valid in $\mathcal{M}$.
- If φ is valid in all structured Kripke models for TUMPL, we write $\models_{\mathsf{TUMPL}} \varphi$ and say that φ is valid.

■

Counter-Model of Transitivity of Indifference (1)

- We would like to provide a counter-model of the transitivity of indifference.

Counter-Model of Transitivity of Indifference (1)

- We would like to provide a counter-model of the transitivity of indifference.
- We now return to Example 1.

Counter-Model of Transitivity of Indifference (1)

- We would like to provide a counter-model of the transitivity of indifference.
- We now return to Example 1.
- Assume that $\mathcal{U} := (\mathbf{W}, R, V, \rho)$ is given, where:

Counter-Model of Transitivity of Indifference (1)

- We would like to provide a counter-model of the transitivity of indifference.
- We now return to Example 1.
- Assume that $\mathcal{U} := (\mathbf{W}, R, V, \rho)$ is given, where:
  - $\mathbf{W} := \{w_0, \ldots, w_{400}\}$, where w_i is a possible world in which you try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar, for any i $(0 \leq i \leq 400)$,

Counter-Model of Transitivity of Indifference (1)

- We would like to provide a counter-model of the transitivity of indifference.
- We now return to Example 1.
- Assume that $\mathcal{U} := (\mathbf{W}, R, V, \rho)$ is given, where:
  - $\mathbf{W} := \{w_0, \ldots, w_{400}\}$, where w_i is a possible world in which you try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar, for any i $(0 \leq i \leq 400)$,
  - R is a binary relation on $\mathbf{W}$,

Counter-Model of Transitivity of Indifference (1)

- We would like to provide a counter-model of the transitivity of indifference.
- We now return to Example 1.
- Assume that $\mathcal{U} := (\mathbf{W}, R, V, \rho)$ is given, where:
    - $\mathbf{W} := \{w_0, \ldots, w_{400}\}$, where w_i is a possible world in which you try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar, for any i $(0 \leq i \leq 400)$,
    - R is a binary relation on $\mathbf{W}$,
    - V is a truth assignment to each $s \in \mathbf{S}$ for each $w_i \in \mathbf{W}$,

Counter-Model of Transitivity of Indifference (1)

- We would like to provide a counter-model of the transitivity of indifference.
- We now return to Example 1.
- Assume that $\mathcal{U} := (\mathbf{W}, R, V, \rho)$ is given, where:
  - $\mathbf{W} := \{w_0, \dots, w_{400}\}$, where w_i is a possible world in which you try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar, for any i $(0 \leq i \leq 400)$,
  - R is a binary relation on $\mathbf{W}$,
  - V is a truth assignment to each $s \in \mathbf{S}$ for each $w_i \in \mathbf{W}$,
  - ρ is a preference space assignment that assigns to each $w_i \in \mathbf{W}$ a preference space $(\mathcal{F}_{w_i}, \succ_{w_i})$, where:

Counter-Model of Transitivity of Indifference (1)

- We would like to provide a counter-model of the transitivity of indifference.
- We now return to Example 1.
- Assume that $\mathcal{U} := (\mathbf{W}, R, V, \rho)$ is given, where:
  - $\mathbf{W} := \{w_0, \ldots, w_{400}\}$, where w_i is a possible world in which you try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar, for any i $(0 \leq i \leq 400)$,
  - R is a binary relation on $\mathbf{W}$,
  - V is a truth assignment to each $s \in \mathbf{S}$ for each $w_i \in \mathbf{W}$,
  - ρ is a preference space assignment that assigns to each $w_i \in \mathbf{W}$ a preference space $(\mathcal{F}_{w_i}, \succ_{w_i})$, where:
    - $\mathcal{F}_{w_i}$ is a Boolean algebra of subsets of $\{w_i' \in \mathbf{W} : R(w_i, w_i')\}$,

Counter-Model of Transitivity of Indifference (1)

- We would like to provide a counter-model of the transitivity of indifference.
- We now return to Example 1.
- Assume that $\mathcal{U} := (\mathbf{W}, R, V, \rho)$ is given, where:
  - $\mathbf{W} := \{w_0, \ldots, w_{400}\}$, where w_i is a possible world in which you try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar, for any i $(0 \leq i \leq 400)$,
  - R is a binary relation on $\mathbf{W}$,
  - V is a truth assignment to each $s \in \mathbf{S}$ for each $w_i \in \mathbf{W}$,
  - ρ is a preference space assignment that assigns to each $w_i \in \mathbf{W}$ a preference space $(\mathcal{F}_{w_i}, \succ_{w_i})$, where:
    - $\mathcal{F}_{w_i}$ is a Boolean algebra of subsets of $\{w_i' \in \mathbf{W} : R(w_i, w_i')\}$,
    - $\succ_{w_i}$ on $\mathcal{F}_{w_i}$ is a semiorder,

Counter-Model of Transitivity of Indifference (1)

- We would like to provide a counter-model of the transitivity of indifference.
- We now return to Example 1.
- Assume that $\mathcal{U} := (\mathbf{W}, R, V, \rho)$ is given, where:
  - $\mathbf{W} := \{w_0, \ldots, w_{400}\}$, where w_i is a possible world in which you try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar, for any i $(0 \leq i \leq 400)$,
  - R is a binary relation on $\mathbf{W}$,
  - V is a truth assignment to each $s \in \mathbf{S}$ for each $w_i \in \mathbf{W}$,
  - ρ is a preference space assignment that assigns to each $w_i \in \mathbf{W}$ a preference space $(\mathcal{F}_{w_i}, \succ_{w_i})$, where:
    - $\mathcal{F}_{w_i}$ is a Boolean algebra of subsets of $\{w_i' \in \mathbf{W} : R(w_i, w_i')\}$,
    - $\succ_{w_i}$ on $\mathcal{F}_{w_i}$ is a semiorder,
    - $\{w_j\} \sim_{w_i} \{w_{j+1}\}$, for any j $(0 \leq j \leq 400)$,

Counter-Model of Transitivity of Indifference (1)

- We would like to provide a counter-model of the transitivity of indifference.
- We now return to Example 1.
- Assume that $\mathcal{U} := (\mathbf{W}, R, V, \rho)$ is given, where:
  - $\mathbf{W} := \{w_0, \ldots, w_{400}\}$, where w_i is a possible world in which you try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar, for any i $(0 \leq i \leq 400)$,
  - R is a binary relation on $\mathbf{W}$,
  - V is a truth assignment to each $s \in \mathbf{S}$ for each $w_i \in \mathbf{W}$,
  - ρ is a preference space assignment that assigns to each $w_i \in \mathbf{W}$ a preference space $(\mathcal{F}_{w_i}, \succ_{w_i})$, where:
    - $\mathcal{F}_{w_i}$ is a Boolean algebra of subsets of $\{w_i' \in \mathbf{W} : R(w_i, w_i')\}$,
    - $\succ_{w_i}$ on $\mathcal{F}_{w_i}$ is a semiorder,
    - $\{w_j\} \sim_{w_i} \{w_{j+1}\}$, for any j $(0 \leq j \leq 400)$,
    - $\{w_0\} \nsim_{w_i} \{w_{400}\}$.

Counter-Model of Transitivity of Indifference (2)

- Let φ_i denote the sentence "You try a cup of coffee with $(1+\frac{i}{100})x$ grams of sugar", for any i $(0 \leq i \leq 400)$.

Counter-Model of Transitivity of Indifference (2)

- Let φ_i denote the sentence "You try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar", for any i $(0 \leq i \leq 400)$.
- Then we have, for any i $(0 \leq i \leq 400)$,

Counter-Model of Transitivity of Indifference (2)

- Let φ_i denote the sentence "You try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar", for any i $(0 \leq i \leq 400)$.
- Then we have, for any i $(0 \leq i \leq 400)$,

$$[\![\varphi_j]\!]^{\mathcal{U}}_{w_i} \sim_{w_i} [\![\varphi_{j+1}]\!]^{\mathcal{U}}_{w_i}, \text{ for any } j \ (0 \leq j \leq 400),$$

Counter-Model of Transitivity of Indifference (2)

- Let φ_i denote the sentence "You try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar", for any i $(0 \le i \le 400)$.
- Then we have, for any i $(0 \le i \le 400)$,

$$\begin{array}{l} [\![\varphi_j]\!]^{\mathcal{U}}_{w_i} \sim_{w_i} [\![\varphi_{j+1}]\!]^{\mathcal{U}}_{w_i}, \text{ for any } j \ (0 \le j \le 400), \\ [\![\varphi_0]\!]^{\mathcal{U}}_{w_i} \nsim_{w_i} [\![\varphi_{400}]\!]^{\mathcal{U}}_{w_i}, \end{array}$$

Counter-Model of Transitivity of Indifference (2)

- Let φ_i denote the sentence "You try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar", for any i $(0 \leq i \leq 400)$.
- Then we have, for any i $(0 \leq i \leq 400)$,

$$\begin{array}{l} [\![\varphi_j]\!]^{\mathcal{U}}_{w_i} \sim_{w_i} [\![\varphi_{j+1}]\!]^{\mathcal{U}}_{w_i}, \text{ for any } j \ (0 \leq j \leq 400), \\ [\![\varphi_0]\!]^{\mathcal{U}}_{w_i} \nsim_{w_i} [\![\varphi_{400}]\!]^{\mathcal{U}}_{w_i}, \end{array}$$

for $[\![\varphi_j]\!]^{\mathcal{U}}_{w_i} = \{w_j\}$ holds for any j $(0 \leq j \leq 400)$.

Counter-Model of Transitivity of Indifference (2)

- Let φ_i denote the sentence "You try a cup of coffee with $(1 + \frac{i}{100})x$ grams of sugar", for any i $(0 \leq i \leq 400)$.
- Then we have, for any i $(0 \leq i \leq 400)$,

$$\begin{array}{l} [\![\varphi_j]\!]^{\mathcal{U}}_{w_j} \sim_{w_i} [\![\varphi_{j+1}]\!]^{\mathcal{U}}_{w_j}, \text{ for any } j \ (0 \leq j \leq 400), \\ [\![\varphi_0]\!]^{\mathcal{U}}_{w_i} \nsim_{w_i} [\![\varphi_{400}]\!]^{\mathcal{U}}_{w_i}, \end{array}$$

  for $[\![\varphi_j]\!]^{\mathcal{U}}_{w_i} = \{w_j\}$ holds for any j $(0 \leq j \leq 400)$.
- It must be noted that, for any i $(0 \leq i \leq 400)$, because $\succ_{w_i}$ on $\mathcal{F}_{w_i}$ is a semiorder,
  $[\![\varphi_j]\!]^{\mathcal{U}}_{w_i} \sim_{w_i} [\![\varphi_{j+1}]\!]^{\mathcal{U}}_{w_i}$ for any j $(0 \leq j \leq 400)$ does not imply $[\![\varphi_0]\!]^{\mathcal{U}}_{w_i} \sim_{w_i} [\![\varphi_{400}]\!]^{\mathcal{U}}_{w_i}$.

Counter-Model of Transitivity of Indifference (3)

- So we have, or any i $(0 \leq i \leq 400)$,

Counter-Model of Transitivity of Indifference (3)

- So we have, or any i $(0 \leq i \leq 400)$,

$$(\mathcal{U}, w_i) \not\models_{\mathrm{TUMPL}} (\mathbf{IND}(\varphi_0, \varphi_1)\& \cdots \&\mathbf{IND}(\varphi_{399}, \varphi_{400})) \rightarrow \mathbf{IND}(\varphi_0, \varphi_{400}).$$

Counter-Model of Transitivity of Indifference (3)

- So we have, or any i $(0 \leq i \leq 400)$,

  $(\mathcal{U}, w_i) \not\models_{\text{TUMPL}} (\mathbf{IND}(\varphi_0, \varphi_1) \& \cdots \& \mathbf{IND}(\varphi_{399}, \varphi_{400})) \to \mathbf{IND}(\varphi_0, \varphi_{400})$.

- Therefore we obtain the following proposition.

Counter-Model of Transitivity of Indifference (3)

- So we have, or any i $(0 \leq i \leq 400)$,

  $$(\mathcal{U}, w_i) \not\models_{\mathrm{TUMPL}} (\mathbf{IND}(\varphi_0, \varphi_1) \& \cdots \& \mathbf{IND}(\varphi_{399}, \varphi_{400})) \to \mathbf{IND}(\varphi_0, \varphi_{400}).$$

- Therefore we obtain the following proposition.

Proposition (Nontransitivity of Indifference)

$$\not\models_{\mathrm{TUMPL}} (\mathbf{IND}(\varphi_0, \varphi_1) \& \cdots \& \mathbf{IND}(\varphi_{399}, \varphi_{400})) \to \mathbf{IND}(\varphi_0, \varphi_{400}).$$

■

Proof System

We provide TUMPL with a proof system.

Definition (Proof System)

The proof system of TUMPL consists of the following:

Proof System

We provide TUMPL with a proof system.

Definition (Proof System)

The proof system of TUMPL consists of the following:

1. all tautologies of classical sentential logic,

Proof System

We provide TUMPL with a proof system.

Definition (Proof System)

The proof system of TUMPL consists of the following:

1. all tautologies of classical sentential logic,
2. $\Box(\varphi_1 \to \varphi_2) \to (\Box\varphi_1 \to \Box\varphi_2) \qquad (K)$,

Proof System

We provide TUMPL with a proof system.

Definition (Proof System)

The proof system of TUMPL consists of the following:

1. all tautologies of classical sentential logic,
2. $\Box(\varphi_1 \to \varphi_2) \to (\Box\varphi_1 \to \Box\varphi_2) \qquad (K)$,
3. $\Box(\varphi_1 \leftrightarrow \varphi_2) \& \Box(\psi_1 \leftrightarrow \psi_2) \to (\mathbf{SPR}(\varphi_1, \psi_1) \leftrightarrow \mathbf{SPR}(\varphi_2, \psi_2))$ (Replacement of Necessary Equivalents),

Proof System

We provide TUMPL with a proof system.

Definition (Proof System)

The proof system of TUMPL consists of the following:

1. all tautologies of classical sentential logic,
2. $\Box(\varphi_1 \to \varphi_2) \to (\Box\varphi_1 \to \Box\varphi_2) \qquad (K)$,
3. $\Box(\varphi_1 \leftrightarrow \varphi_2) \& \Box(\psi_1 \leftrightarrow \psi_2) \to (\mathbf{SPR}(\varphi_1, \psi_1) \leftrightarrow \mathbf{SPR}(\varphi_2, \psi_2))$
   (Replacement of Necessary Equivalents),
4. $\neg\mathbf{SPR}(\varphi, \varphi)$
   (Syntactic Counterpart of Irreflexivity),

Proof System

We provide TUMPL with a proof system.

Definition (Proof System)

The proof system of TUMPL consists of the following:

1. all tautologies of classical sentential logic,
2. $\Box(\varphi_1 \to \varphi_2) \to (\Box\varphi_1 \to \Box\varphi_2) \qquad (K)$,
3. $\Box(\varphi_1 \leftrightarrow \varphi_2) \& \Box(\psi_1 \leftrightarrow \psi_2) \to (\mathbf{SPR}(\varphi_1, \psi_1) \leftrightarrow \mathbf{SPR}(\varphi_2, \psi_2))$ (Replacement of Necessary Equivalents),
4. $\neg\mathbf{SPR}(\varphi, \varphi)$ (Syntactic Counterpart of Irreflexivity),
5. $(\mathbf{SPR}(\varphi_1, \varphi_2) \wedge \mathbf{SPR}(\varphi_3, \varphi_4)) \to (\mathbf{SPR}(\varphi_1, \varphi_4) \vee \mathbf{SPR}(\varphi_3, \varphi_2))$ (Syntactic Counterpart of Intervality),

Proof System

We provide TUMPL with a proof system.

Definition (Proof System)

The proof system of TUMPL consists of the following:

1. all tautologies of classical sentential logic,
2. $\Box(\varphi_1 \to \varphi_2) \to (\Box\varphi_1 \to \Box\varphi_2) \qquad (K)$,
3. $\Box(\varphi_1 \leftrightarrow \varphi_2) \& \Box(\psi_1 \leftrightarrow \psi_2) \to (\mathbf{SPR}(\varphi_1, \psi_1) \leftrightarrow \mathbf{SPR}(\varphi_2, \psi_2))$
   (Replacement of Necessary Equivalents),
4. $\neg\mathbf{SPR}(\varphi, \varphi)$
   (Syntactic Counterpart of Irreflexivity),
5. $(\mathbf{SPR}(\varphi_1, \varphi_2) \wedge \mathbf{SPR}(\varphi_3, \varphi_4)) \to$
   $(\mathbf{SPR}(\varphi_1, \varphi_4) \vee \mathbf{SPR}(\varphi_3, \varphi_2))$
   (Syntactic Counterpart of Intervality),
6. $(\mathbf{SPR}(\varphi_1, \varphi_2) \wedge \mathbf{SPR}(\varphi_2, \varphi_3)) \to$
   $(\mathbf{SPR}(\varphi_1, \varphi_4) \vee \mathbf{SPR}(\varphi_4, \varphi_3))$
   (Syntactic Counterpart of Semitransitivity),

Proof System

We provide TUMPL with a proof system.

Definition (Proof System)

The proof system of TUMPL consists of the following:

1. all tautologies of classical sentential logic,
2. $\Box(\varphi_1 \to \varphi_2) \to (\Box\varphi_1 \to \Box\varphi_2) \qquad (K)$,
3. $\Box(\varphi_1 \leftrightarrow \varphi_2) \& \Box(\psi_1 \leftrightarrow \psi_2) \to (\mathbf{SPR}(\varphi_1, \psi_1) \leftrightarrow \mathbf{SPR}(\varphi_2, \psi_2))$
   (Replacement of Necessary Equivalents),
4. $\neg\mathbf{SPR}(\varphi, \varphi)$
   (Syntactic Counterpart of Irreflexivity),
5. $(\mathbf{SPR}(\varphi_1, \varphi_2) \wedge \mathbf{SPR}(\varphi_3, \varphi_4)) \to$
   $(\mathbf{SPR}(\varphi_1, \varphi_4) \vee \mathbf{SPR}(\varphi_3, \varphi_2))$
   (Syntactic Counterpart of Intervality),
6. $(\mathbf{SPR}(\varphi_1, \varphi_2) \wedge \mathbf{SPR}(\varphi_2, \varphi_3)) \to$
   $(\mathbf{SPR}(\varphi_1, \varphi_4) \vee \mathbf{SPR}(\varphi_4, \varphi_3))$
   (Syntactic Counterpart of Semitransitivity),
7. Modus Ponens,

Proof System

We provide TUMPL with a proof system.

Definition (Proof System)

The proof system of TUMPL consists of the following:

1. all tautologies of classical sentential logic,
2. $\Box(\varphi_1 \to \varphi_2) \to (\Box\varphi_1 \to \Box\varphi_2)$ $\quad (K)$,
3. $\Box(\varphi_1 \leftrightarrow \varphi_2) \& \Box(\psi_1 \leftrightarrow \psi_2) \to (\mathbf{SPR}(\varphi_1, \psi_1) \leftrightarrow \mathbf{SPR}(\varphi_2, \psi_2))$ (Replacement of Necessary Equivalents),
4. $\neg\mathbf{SPR}(\varphi, \varphi)$ (Syntactic Counterpart of Irreflexivity),
5. $(\mathbf{SPR}(\varphi_1, \varphi_2) \wedge \mathbf{SPR}(\varphi_3, \varphi_4)) \to (\mathbf{SPR}(\varphi_1, \varphi_4) \vee \mathbf{SPR}(\varphi_3, \varphi_2))$ (Syntactic Counterpart of Intervality),
6. $(\mathbf{SPR}(\varphi_1, \varphi_2) \wedge \mathbf{SPR}(\varphi_2, \varphi_3)) \to (\mathbf{SPR}(\varphi_1, \varphi_4) \vee \mathbf{SPR}(\varphi_4, \varphi_3))$ (Syntactic Counterpart of Semitransitivity),
7. Modus Ponens,
8. Necessitation. ■

Provability

Definition (Provability)

- A proof of $\varphi \in \Phi_{\mathsf{TUMPL}}$ is a finite sequence of $\mathcal{L}_{\mathsf{TUMPL}}$-formulae having φ as the last formula such that either each formula is an instance of an axiom, or it can be obtained from formulae that appear earlier in the sequence by applying an inference rule.

Provability

Definition (Provability)

- A proof of $\varphi \in \Phi_{\mathsf{TUMPL}}$ is a finite sequence of $\mathcal{L}_{\mathsf{TUMPL}}$-formulae having φ as the last formula such that either each formula is an instance of an axiom, or it can be obtained from formulae that appear earlier in the sequence by applying an inference rule.
- If there is a proof of φ, we write $\vdash_{\mathsf{TUMPL}} \varphi$. ■

Soundness

- We prove the metatheorems of TUMPL.

Soundness

- We prove the metatheorems of TUMPL.
- It is easy to prove the soundness of TUMPL.

Soundness

- We prove the metatheorems of TUMPL.
- It is easy to prove the soundness of TUMPL.

Theorem (Soundness)

For any $\varphi \in \Phi_{\mathcal{L}_{\mathsf{TUMPL}}}$, *if* $\vdash_{\mathsf{TUMPL}} \varphi$, *then* $\models_{\mathsf{TUMPL}} \varphi$. ■

Completeness

- We now turn to the task of proving the completeness of TUMPL.

Completeness

- We now turn to the task of proving the completeness of TUMPL.
- We prove it by developing the idea of Segerberg ([Segerberg 1971]) that we modify filtration theory in such a way that completeness can be established by Corollary(Representation on Finite Boolean Algebra).

Segerberg, K.:
Qualitative Probability in a Modal Setting. in Fenstad, J. E. (ed.):
Proceedings of the Second Scandinavian Logic Symposium. North-Holland, Amsterdam (1971) 341–352.

Completeness

- We now turn to the task of proving the completeness of TUMPL.
- We prove it by developing the idea of Segerberg ([Segerberg 1971]) that we modify filtration theory in such a way that completeness can be established by Corollary(Representation on Finite Boolean Algebra).

Segerberg, K.:
Qualitative Probability in a Modal Setting. in Fenstad, J. E. (ed.):
Proceedings of the Second Scandinavian Logic Symposium.
North-Holland, Amsterdam (1971) 341–352.

Theorem (Completeness)

For any $\varphi \in \Phi_{\mathcal{L}_{\mathrm{TUMPL}}}$, *if* $\models_{\mathrm{TUMPL}} \varphi$, *then* $\vdash_{\mathrm{TUMPL}} \varphi$. ■

Decidability

- TUMPL has the finite model property that every non-theorem of TUMPL fails in a structured Kripke model for preference with only finitely many elements.

Decidability

- TUMPL has the finite model property that every non-theorem of TUMPL fails in a structured Kripke model for preference with only finitely many elements.
- So we can prove the decidability of TUMPL.

Decidability

- TUMPL has the finite model property that every non-theorem of TUMPL fails in a structured Kripke model for preference with only finitely many elements.
- So we can prove the decidability of TUMPL.

Theorem (Decidability)

TUMPL *is decidable.* ■

Structure of This Talk

Summary and Further Investigation

- SUMMARY: In this talk we have proposed a new version of complete and decidable extrinsic preference logic–threshold utility maximiser's preference logic (TUMPL) that can solve the Nontransitivity Problem and avoid the Fundamental Problem of Intrinsic Preference.

Summary and Further Investigation

- SUMMARY: In this talk we have proposed a new version of complete and decidable extrinsic preference logic–threshold utility maximiser's preference logic (TUMPL) that can solve the Nontransitivity Problem and avoid the Fundamental Problem of Intrinsic Preference.
- FURTHER INVESTIGATION: This talk is only a part of a larger measurement-theoretic study. We are now trying to construct such logics as dyadic deontic logic, logic for goodness and badness, and logic of questions and answers by means of measurement theory.

Our Related Work

Satoru Suzuki:
Preference Logic and Its Measurement-Theoretic Semantics.
In: Accepted Papers of 8th Conference on Logic and the Foundations of Game and Decision Theory (LOFT 2008), Universiteit van Amsterdam (2008).

Satoru Suzuki:
Prolegomena to Dynamic Epistemic Preference Logic.
In: Hattori, H. et al. (eds.), New Frontiers in Artificial Intelligence, LNAI 5447, Springer-Verlag, Berlin (2009) 177–192.

Satoru Suzuki:
Measurement-Theoretic Foundation of Preference-Based Dyadic Deontic Logic.
In: He, X. et al. (eds.), Proceedings of the Second International Workshop on Logic, Rationality, and Interaction (LORI-II), LNAI 5834, Springer-Verlag, Berlin, (2009) 278–291.

Satoru Suzuki:
Measurement-Theoretic Foundation of Logic for Goodness and Badness.
In: Bekki, D. (ed.), Proceedings of the Sixth Workshop on Logic and Engineering of Natural Language Semantics (LENLS 2009), JSAI, (2009) 1–14.

Structure of This Talk

Thank You for Your Attention!

A Logical Typology of Normative Systems

Berislav Žarnić
University of Split, Croatia

This paper gives a first order formalization of the proposal put forward by John Broome[1] [2] and develops a typology on that basis. The three-place code function $\mathrm{k} : \mathrm{S} \times \mathrm{A} \times W \longrightarrow \wp\mathcal{L}_\mathrm{n}$ delivers the set $\mathrm{k}_\mathsf{s}(i, w) \subseteq \mathcal{L}_\mathrm{n}$ of propositions in the normative language $\mathcal{L}_\mathrm{n}$ that a normative source $\mathsf{s} \in \mathrm{S}$ requires of an agent $i \in \mathrm{A}$ in a world $w \in W$. The value of the code function $\mathrm{k}_\mathsf{s}(i, w)$ will be termed the 'set of requirements'. The vocabulary of the normative language $\mathcal{L}_\mathrm{n}$ will contain modal operators for belief, B, desire, D, and intention, I. The worlds are construed as subsets of normative language $\mathcal{L}_\mathrm{n}$ which are maximal consistent in propositional logic. Possible worlds may violate the laws of modal logics of intentionality according to the philosophical thesis that the essence of the mental is to be subject to norms, not to conform to them (Zangwill [6]).

Definition 1 *The normative language $\mathcal{L}_\mathrm{n}$ is built over the base language of propositional logic $\mathcal{L}_{PL}$. Let $i \in A$, $\mathrm{X} = \mathrm{B}, \mathrm{D}, \mathrm{I}$, and $p \in \mathcal{L}_{PL}$*

$$\textit{Sentences of } \mathcal{L}_\mathrm{n} ::= p \mid [\mathrm{X}_i]\varphi \mid \neg\varphi \mid (\varphi \wedge \psi)$$

The set of quasi-literals is the set of propositional letters and their negations, and modal formulas and their negations.

The T axiom ($\Box p \rightarrow p$) poses a serious threat to this kind of modeling that keeps modality and world apart. If modalities obeying axiom T were allowed (e.g. epistemic or praxeologic), then possible worlds, being defined as maximal consistent sets in propositional logic, would become intuitively impossible[2]. Since the corresponding T axioms seem to constitute an important part of the meaning of verbs of knowledge and of action, epistemic and praxeologic modalities must be excluded from the language of norms $\mathcal{L}_\mathrm{n}$. Von Wright [4] defined 'content of a norm' as "that which ought to or may or must not be or be done". The normative language $\mathcal{L}_\mathrm{n}$ departs from von Wright's definition by taking norm-content to be *the psychological state or relation of psychological states that ought to or may or must not be present in the mind of the norm addressee on a particular occasion.* The reduction and the switch may seem

[1] "We must allow for the possibility that the requirements you are under depend on your circumstances. ... There is a set of worlds, at each of which propositions have a truth value. The values of all propositions at a particular world conform to the axioms of propositional calculus. For each source of requirements s, each person i and each world w, there is a set of propositions $\mathrm{k}_\mathsf{s}(i, w)$, which is to be interpreted as the set of things that s requires of i at w. Each proposition in the set is a required proposition. The function k_s from i and w to $\mathrm{k}_\mathsf{s}(i, w)$ I shall call s's *code* of requirements". (Broome [2], p. 14) The symbols in the citation have been changed to match the symbols used in this paper.

[2] For example, although $\{\neg p, [\mathrm{K}]_i p\}$ is pl-consistent set, we do not want to have it included in any world since no false sentence can be known to be true.

drastic but there is a rationale for it. The requirement that agent i knows that p could be replaced by $p \to [\mathrm{B}_i]p$; a required action to see to it that p could be replaced by the required intention, i.e. $[\mathrm{I}_i]p$.

In order to achieve technical clarity we define a first-order metanormative many-sorted language $\mathcal{L}_{\mathrm{meta}}$ with the following extralogical **vocabulary** – individual constants for normative sources, agents and worlds: $\mathrm{s}, \mathrm{s}_1, \ldots, \mathrm{a}, \mathrm{a}_1, \ldots, \mathrm{v}, \mathrm{v}_1, \ldots$; function symbols for code of requirement, propositional logic consequence, and logic function: k^3, Cn^1, l^1; function symbols for the sentential forms: neg^1, conj^2, and a set of symbols of the type $\mathrm{mod}^1_{\mathrm{X}i}$; monadic predicate symbols expressing properties of being a normative source, an agent, a sentence in $\mathcal{L}_{\mathrm{n}}$, a possible world: $\mathrm{Sr}^1, Ag^1, \mathrm{Sen}^1, W^1$, and dyadic predicates expressing relations of an agent having i-th normative property (corresponding to i-th normative source) in a world, and relation of membership: $\mathrm{K}^2_s, \mathrm{K}^2_{s_1}, \ldots, \in^2$. The **structures** $\mathfrak{M}_{\mathrm{meta}} = \langle \mathrm{D}, \mathcal{I} \rangle$ are built over the domain $\mathrm{D} = \mathrm{S} \cup \mathrm{A} \cup \mathcal{L}_{\mathrm{n}} \cup \wp\mathcal{L}_{\mathrm{n}}$ where S and A are non-empty and disjoint sets, and $\mathcal{L}_{\mathrm{n}}$ is the set already defined (Definition 1). We use variables $w, w_1, \ldots$ to range over worlds; variables $p, p_1, \ldots, q, q_1, \ldots$ to range over sentences in $\mathcal{L}_{\mathrm{n}}$; variables $i, i_1, \ldots$ to range over agents; and variables $x, y, \ldots$ to range over everything. The shorthand notation for sentential form functions uses "Quine quotes", e.g. the shorthand notation for $\mathrm{neg}(x)$ is $\ulcorner \neg x \urcorner$. For the ease of reading, the universal closure of the formula will be notated by formula with free variables. The interpretation of the nonlogical vocabulary is almost straightforward. More complex cases are:

- interpretation of sentential form functions, which we introduce by the way of example – $\mathcal{I}(\mathrm{neg})$ is a function: $\mathcal{L}_{\mathrm{n}} \to \mathcal{L}_{\mathrm{n}}$ such that

  $$\mathcal{I}(\mathrm{neg})([\![x]\!]^{\mathfrak{M}_{\mathrm{meta}}}_g) = \begin{cases} \neg \frown [\![x]\!]^{\mathfrak{M}_{\mathrm{meta}}}_g & \text{if } [\![x]\!]^{\mathfrak{M}_{\mathrm{meta}}}_g \in \mathcal{L}_{\mathrm{n}}, \\ \text{undefined}, & \text{otherwise.} \end{cases}$$

  where g is an assignment function and $\frown$ is concatenation operation;

- interpretation of logic function l is function $\mathcal{I}(\mathrm{l}) : \wp\mathcal{L}_{\mathrm{n}} \to \wp\mathcal{L}_{\mathrm{n}}$ such that $\mathcal{I}(\mathrm{l})([\![x]\!]^{\mathfrak{M}_{\mathrm{meta}}}_g)$ is the set of all substitutional instances of the formula $[\![y]\!]^{\mathfrak{M}_{\mathrm{meta}}}_g \in \mathcal{L}_{\mathrm{n}}$ for each $y \in x$;

- interpretation of consequence function Cn is a set of consequences in classical propositional logic for a given set, i.e.

  $$\mathcal{I}(\mathrm{Cn})([\![x]\!]^{\mathfrak{M}_{\mathrm{meta}}}_g) = \begin{cases} \{y \in \mathcal{L}_{\mathrm{n}} \mid [\![x]\!]^{\mathfrak{M}_{\mathrm{meta}}}_g \vdash_{\mathrm{pl}} y\} & \text{if } [\![x]\!]^{\mathfrak{M}_{\mathrm{meta}}}_g \subseteq \mathcal{L}_{\mathrm{n}}, \\ \text{undefined}, & \text{otherwise.} \end{cases}$$

Definitions 2 *Quantifications over different argument positions in the code function enable a number of interesting type distinctions, some of which will be introduced below using a $\mathcal{L}_{\mathrm{meta}}$ formula in the definiens.*

- k_{s} *is a* pl-*congruent code iff* $\ulcorner p \leftrightarrow q \urcorner \in \mathrm{Cn}(\varnothing) \to (p \in \mathrm{k}_{\mathsf{s}}(i,w) \leftrightarrow q \in \mathrm{k}_{\mathsf{s}}(i,w))$*;*
- k_{s} *is a* pl-*consistent code iff* $\exists w_2\, \mathrm{k}_{\mathsf{s}}(i,w_1) \subseteq w_2$*;*
- k_{s} *is an achievable code iff* $\exists w\, \mathrm{k}_{\mathsf{s}}(i,w) \subseteq w$*;*
- k_{s} *is a* pl-*deductively closed iff* $\mathrm{k}_{\mathsf{s}}(i,w) = \mathrm{Cn}(\mathrm{k}_{\mathsf{s}}(i,w))$*;*

- $\mathrm{k_s}$ *is a relativistic code iff* $\exists i \exists w_1 \exists w_2\, \mathrm{k_s}(i, w_1) \neq \mathrm{k_s}(i, w_2)$*;*

- *a code is absolute iff it is not relativistic;*

- $\mathrm{k_s}$ *is a socially consistent code iff* $\exists w_2\, \mathrm{k_s}(i_1, w_1) \cup \mathrm{k_s}(i_2, w_1) \subseteq w_2$*;*

- *codes* $\mathrm{k_x}$ *and* $\mathrm{k_y}$ *are realization-equivalent iff* $\mathrm{k_x}(i, w) \subseteq w \leftrightarrow \mathrm{k_y}(i, w) \subseteq w$*;*

- *codes* $\mathrm{k_x}$ *and* $\mathrm{k_y}$ *are compatible iff* $\exists w_2\, \mathrm{k_x}(i, w_1) \cup \mathrm{k_y}(i, w_1) \subseteq w_2$*.*

Consistent and deductively closed codes seem to play an important role in our understanding of the basic normative concepts. For example, deontic KD logic without iterated deontic modalities may be conceived as logic of the specific type of code, namely of consistent pl-deductively closed code.

Definition 3 *Let* $p \in \mathcal{L}_{PL}$ *be a formula of propositional logic:*

$$\textit{Formulas of } \mathcal{L}^{\mathrm{O}}_{\mathrm{KD}} ::= p \mid \mathrm{O}p \mid \mathrm{P}p \mid \neg\varphi \mid (\varphi \wedge \psi)$$

Let us introduce the translation τ^1 from the restricted language $\mathcal{L}^{\mathrm{O}}_{\mathrm{KD}}$ to the metanormative language $\mathcal{L}_{\mathrm{meta}}$, with $\mathrm{O}\,p$ and $\mathrm{P}\,p$ standing for 'i in v has s-obligation (s-permission) to p'.

Definition 4 *Function* τ *maps sentences from the fragment* $\mathcal{L}^{\mathrm{O}}_{\mathrm{KD}} \cap \mathcal{L}_{PL}$ *to the set of sentential variables and sentential function terms of* $\mathcal{L}_{\mathrm{meta}}$*:*

$$\begin{array}{rcll}
\tau(l) & \in & \{p, p_1, \ldots, q, q_1, \ldots\} & \textit{for propositional letters } l \in \mathcal{L}_{PL} \\
\tau(\neg\varphi) & = & \neg\tau(\varphi) & \\
\tau(\varphi \wedge \psi) & = & (\tau(\varphi) \wedge \tau(\psi)) &
\end{array}$$

Definition 5 *Translation* $\tau^1 : \mathcal{L}^{\mathrm{O}}_{\mathrm{KD}} \to \mathcal{L}_{\mathrm{meta}}$

$$\begin{array}{rcll}
\tau^1(p) & = & \ulcorner\tau(p)\urcorner \in \mathrm{v} & \textit{if } p \in \mathcal{L}_{PL} \\
\tau^1(\mathrm{O}\varphi) & = & \ulcorner\tau(\varphi)\urcorner \in \mathrm{k_s}(\mathrm{a}, \mathrm{v}) & \\
\tau^1(\mathrm{P}\varphi) & = & \ulcorner\tau(\neg\varphi)\urcorner \notin \mathrm{k_s}(\mathrm{a}, \mathrm{v}) & \\
\tau^1(\neg\varphi) & = & \neg\tau^1(\varphi) & \\
\tau^1(\varphi \wedge \psi) & = & (\tau^1(\varphi) \wedge \tau^1(\psi)) &
\end{array}$$

The principles of the standard deontic logic[3] hold under the translation τ^1:

- "gaplessness" condition $\mathrm{P}p \vee \mathrm{O}\neg p$ translates to $\ulcorner\neg p\urcorner \notin \mathrm{k_s}(\mathrm{a}, \mathrm{v}) \vee \ulcorner\neg p\urcorner \in \mathrm{k_s}(\mathrm{a}, \mathrm{v})$ and that property obviously holds for any set of requirements;

- K axiom becomes $\ulcorner p \to q\urcorner \in \mathrm{k_s}(\mathrm{a}, \mathrm{v}) \to (p \in \mathrm{k_s}(\mathrm{a}, \mathrm{v}) \to q \in \mathrm{k_s}(\mathrm{a}, \mathrm{v}))$ and that property holds for any pl-deductively closed set;

- D axiom becomes $p \in \mathrm{k_s}(\mathrm{a}, \mathrm{v}) \to \ulcorner\neg p\urcorner \notin \mathrm{k_s}(\mathrm{a}, \mathrm{v})$ and that is just another way of stating pl-consistency;

[3]"...classical deontic logic, on the descriptive interpretation of its formulas, pictures a gapless and contradiction-free system of norms". (Von Wright [5] p. 32)
According to our translation scheme von Wright's claim should be appended: classical deontic logic "pictures a system of norms" that is deductively closed too.

- mutual definability, $\mathrm{P}_1 p \leftrightarrow \neg\mathrm{O}\neg p$ holds if the set of requirements is congruent.

Although iterated deontic operators receive no translation in the scheme proposed above, one may extend the line of thought by giving additional translation rules for language of standard deontic $\mathcal{L}_{\mathrm{KD}}^{\mathrm{OO}}$ restricted to the maximum of two iterations of deontic operators, treating iterated deontic modalities as a sequence of heterogenous operators and introducing the distinction into the syntax:

$$\mathcal{L}_{\mathrm{KD}}^{\mathrm{O_2O}} ::= p \in \mathcal{L}_{\mathrm{KD}}^{\mathrm{O}} \mid \mathrm{O}_2 p \mid \mathrm{P}_2 p \mid \neg\varphi \mid (\varphi \wedge \psi)$$

Definition 6 *Let $Sub(\varphi)_{[\frac{c_1}{x_1}\cdots\frac{c_n}{x_n}]}$ denote substitutional instance of $\varphi \in \mathcal{L}_{\mathrm{meta}}$ in which constants $c_1, ..., c_n$ are replaced by variables $x_1, ..., x_n$. Translation $\tau^2 : \mathcal{L}_{\mathrm{KD}}^{\mathrm{O_2O}} \to \mathcal{L}_{\mathrm{meta}}$*

$$\begin{array}{rcll} \tau^2(\mathrm{O}_2 p) & = & \forall i \forall w\ Sub(\tau^1(p))_{[\frac{\mathsf{a}}{i}\frac{\mathsf{v}}{w}]} & \text{for } p \in \mathcal{L}_{\mathrm{KD}}^{\mathrm{O}} \\ \tau^2(\mathrm{P}_2 p) & = & \exists i \exists w\ Sub(\tau^1(p))_{[\frac{\mathsf{a}}{i}\frac{\mathsf{v}}{w}]} & \text{for } p \in \mathcal{L}_{\mathrm{KD}}^{\mathrm{O}} \\ \tau^2(\neg\varphi) & = & \neg\tau^2(\varphi) & \\ \tau^2(\varphi \wedge \psi) & = & (\tau^2(\varphi) \wedge \tau^2(\psi)) & \end{array}$$

Such an approach to iterated deontic modalities departs from von Wright's [5] "second order descriptive interpretation" where e.g. O_2 would stand for existence of "normative demands on normative systems" ("norms for the norm givers"). The "first order" translation τ^1 as well as the "second order" translation τ^2 give us statements in metanormative language $\mathcal{L}_{\mathrm{meta}}$ both of which may "picture" some type of "normative system". The difference lies in the fact that τ^1 gives a local picture of a set of requirements (for a particular source, agent and world) while τ^2 gives a more global picture of a code function. In the second case the properties depicted are the properties of a code function for a particular source with respect to any agent and any world.

Let us consider KD45 deontic logic! The τ^2 translations of reinterpreted axioms 4, $\mathrm{O}_1 p \to \mathrm{O}_2\mathrm{O}_1 p$ and 5, $\mathrm{P}_1 p \to \mathrm{O}_2\mathrm{P}_1 p$ amount to stating that any s-obligation and any s-permission holds universally. So, the reinterpreted axioms will hold only if s-code is absolute.

Definition 7 *An agent i at world w has an "all-or-nothing" normative property K_{s} that corresponds to the source s iff the set of requirements $\mathrm{k}_{\mathsf{s}}(i, w)$ is satisfied in w, i.e. $\mathrm{K}_{\mathsf{s}}(i, w) \leftrightarrow \mathrm{k}_{\mathsf{s}}(i, w) \subseteq w$.*

If the only way to satisfy some relativistic code and some absolute code is to satisfy them simultaneously, then these codes define the same normative property. The question arises as to whether (non)absoluteness of a code function introduces a difference with respect to normative properties. The next theorem provides a negative answer.

Theorem 8 *For any code there is a realization equivalent absolute code.*

The proof requires extension of the normative language $\mathcal{L}_{\mathrm{n}}$ to the language $\mathcal{L}_{\mathrm{n}(\omega_1)}$ of a variant of infinitary logic which has the same vocabulary as $\mathcal{L}_{\mathrm{n}}$, but in $\mathcal{L}_{\mathrm{n}(\omega_1)}$ the conjunction symbol $\bigwedge$ may be applied to subsets of the set of *quasi-literals*. A function $\mathrm{k}_{\mathsf{s}}^{cond}$ is a conditionalized variant of a code k_{s} iff

$$\forall p \forall w_1 (p \in \mathrm{k}_{\mathsf{s}}^{cond}(i, w_1) \leftrightarrow \exists q \exists w_2 (p = \ulcorner \bigwedge \mathrm{lit}(w_2) \to q \urcorner \wedge q \in \mathrm{k}_{\mathsf{s}}(i, w_2)))$$

where $\mathrm{lit}(w_2)$ is the set of all *quasi-literals* belonging to w_2. The existence of conditionalized variant for any code proves the theorem. In the light of theorem 8, world and agent generalizing translation of axioms 4 and 5 do not introduce distinctions into logical typology of normative properties.

There are several plausible principles of intentionality and normativity: intentionality is normative, i.e. subjected to norms of different sources (e.g. [6]); rationality is one of the normative sources; some norms of rationality are based on logic of psychological modalities. If we accept these principles, then the codes that deliver some "logical" set of sentences deserve our attention. A number of authors take the closure under equivalence to be either unproblematic (e.g. [2]) or at least plausible minimal logical property of a code. In other words, the codes inherit some of the easily noticeable logical properties of the language in which norm-contents are stated. But then a question arises as to which properties are to be preserved in any code. E.g. if the truth-functional equivalence should be inherited, should not the modal congruence[4] be inherited as well, especially in the light of the widely accepted principle that propositions, and not sentences, are the objects of intentionality?

Definition 9 *Let $x \subseteq \mathcal{L}_\mathrm{n}$. The set of sentences $\mathrm{l}(x) \subseteq \mathcal{L}_\mathrm{n}$ is an axiomatic basis for a set of modal operators occurring in sentences in x ($\mathrm{l}(x)$ is the set of all the substitutional instances of sentences in the set x).*

Let us suppose that $\mathrm{l}(x)$ is also an axiomatic basis for the set of modal operators occurring in sentences in the sets of requirements delivered by k_s. Then we may distinguish several interesting types of codes that do not violate a logic of the modal part of its language:

- code is consistent with respect to $\mathrm{l}(x)$ iff $\exists w_2\ \mathrm{Cn}(\mathrm{l}(x) \cup \mathrm{k}_\mathsf{s}(i, w_1)) \subseteq w_2$;
- code is a logic iff $\exists x\ \mathrm{k}_\mathsf{s}(i, w) = \mathrm{Cn}(\mathrm{l}(x))$;
- code is "more than a logic" iff $\exists x \exists y(\neg y \subseteq \mathrm{Cn}(\mathrm{l}(x)) \wedge \mathrm{k}_\mathsf{s}(i, w) = \mathrm{Cn}(\mathrm{l}(x) \cup y))$;
- code is "less than a logic" iff $\exists x \exists y(\neg y \subseteq \mathrm{Cn}(\mathrm{l}(x)) \wedge \mathrm{k}_\mathsf{s}(i, w) = \mathrm{Cn}(\mathrm{l}(x) \cup y) - \mathrm{Cn}(\mathrm{l}(x)))$.

The second type of the logical code could be termed 'formal code', the third and the fourth — 'material codes'. All the four types exhibit some kind of "internal logicality".

One may distinguish two types of logical properties that a code may have. On the one hand, there are external properties of sets of requirements and code functions, like those given in the definitions 2. On the other hand, there is also an internal logicality of a code, pertaining to the modal logic of the code contents.

This approach relaxes the burden of unrealistic logical models of intentionality by their relocation to the normative side; e.g. it is nonsensical to attribute logical omniscience to real agents with finite resources available for reasoning, but one might argue that it is not nonsensical to consider logical omniscience as a normative requirement. The interaction between the normative and the real takes place on the level of agent properties. The widely accepted "ought

[4]If p and q are truth-functionally equivalent, then $[\mathrm{X}_i]p \in \mathrm{k}_\mathsf{s}(i, w)$ iff $[\mathrm{X}_i]q \in \mathrm{k}_\mathsf{s}(i, w)$.

implies can" principle holds if a code is achievable, i.e. if it is possible for an agent to have the normative property that corresponds to to the source of the code. A straightforward definition of the "all-or-nothing" normative property has been proposed by Broome (see definition 7 above). It is commonly held that rationality as a normative property is not all-or-nothing property but a matter of degree (e.g. Davidson [3]). Therefore, the set of requirements satisfied by an agent having the property of rationality need not include all the requirements delivered by rationality as a normative source. Consequently, the definition of achievability of the code should be modified for "extensive properties".

Further work. The typology of normative systems seems to need a supplementary typology of normative properties, most notably of those that are defined in terms of partial satisfaction. The motivation for the theory of belief revision came from legal context. AGM theory *inter alia* described the logical ways in which consistency of a theory should be maintained. The logical properties that define the state of equilibrium for "homeostatic dynamics" of normative codes should be determined. *Prima facie*, a number of other properties besides "external consistency" like social consistency, achievability, "internal consistency" should be included[5].

References

[1] Carlos E. Alchourrón and Eugenio Bulygin. The expressive conception of norms. In R. Hilpinen (ed.)*New Studies in Deontic Logic*, pp. 95–125, D. Reidel Publishing Company, Dordrecht, 1981.

[2] John Broome. Requirements. In T. Rønnow-Rasmussen, B. Petersson, J. Josefsson, and D. Egonsson, editors,*Homage a Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, pages 1–41. Lunds universitet, Lund, 2007. `http://www.fil.lu.se/hommageawlodek`.

[3] Donald Davidson. *Problems of Rationality.* Clarendon Press, Oxford, 2004.

[4] Georg Henrik von Wright. *Norm and Action : A Logical Enquiry.* Routledge and Kegan Paul, London, 1963.

[5] Georg Henrik von Wright. Deontic logic: a personal view. *Ratio Juris*, **12**: 26–38, 1999.

[6] Nick Zangwill. The normativity of the mental. *Philosophical Explorations*, **8**: 1–19, 2005.

[5] The author would like to thank anonymous referees for helpful comments. The investigation presented in the paper is supported by MZOŠ project "Logical structures and intentionality".

A Logical Typology of Normative Systems

(and its relation to deontic logic)

Berislav Žarnić

University of Split, Croatia

SOCREAL2010, Sapporo

The theoretical context

This talk is a continuation of the theoretical approach given in the following works:

Carlos E. Alchourrón and Eugenio Bulygin.
The expressive conception of norms.
In R. Hilpinen (ed.)*New Studies in Deontic Logic*, pp. 95–125, D. Reidel Publishing Company, Dordrecht, 1981.

John Broome.
Requirements.
In T. Rønnow-Rasmussen, B. Petersson, J. Josefsson, and D. Egonsson, editors,*Homage a Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, pages 1–41. Lunds universitet, Lund, 2007.
http://www.fil.lu.se/hommageawlodek.

Georg Henrik von Wright.
Deontic logic: a personal view.
Ratio Juris, **12**: 26–38, 1999.

Set theoretic approach in the logical theory of normative systems

Set theoretical approach

1. introduced by Alchourrón and Bulygin (1981): the "force" of norm is represented by the membership of its norm-content in the set (normative system);
2. discussed as a possible interpretation of deontic logic by von Wright (1999);
3. generalized by treating the sets of norm-contents as values of code functions by Broome (2007).

Quote (C. Alchourrón and E. Bulygin, 1981)

We . . . define the concept of a normative system as the set of all the propositions that are consequences of the explicitly commanded propositions.

Quote (G. H. von Wright, 2007)

. . . classic deontic logic, on the descriptive interpretation of its formulas, pictures a gapless and contradiction-free system of norms.

Generalization of the set-theoretic approach

Quote

We must allow for the possibility that the requirements you are under depend on your circumstances. ... There is a set of worlds, at each of which propositions have a truth value. The values of all propositions at a particular world conform to the axioms of propositional calculus. For each source of requirements s, each person *i* and each world *w*, there is a set of propositions $\mathrm{k_s}(i, w)$, which is to be interpreted as the set of things that s requires of *i* at *w*. Each proposition in the set is a required proposition. The function $\mathrm{k_s}$ from *i* and *w* to $\mathrm{k_s}(i, w)$ I shall call s's *code* of requirements".
Broome [2007] p. 14

Code function

- Code is ternary function:

$$k_{\boxed{1}}\left(\boxed{2},\boxed{3}\right)=\boxed{4}$$

- Input:
  1. A normative source[1]
  2. An agent
  3. A world
- Output
  4. A set of sentences.

[1] Broome does not explicate the notion of different normative sources but introduces the notion by the way of examples (”survival,” ”prudence” and ”rationality”). I will not give explication for the notion of normative sources either, but I will give a sketch of the notion that was implicit in my thoughts. Normative sources are: formal and material. Formal normative sources regulate relations between intentional states either within one category (e.g. theoretical rationality) or between categories (e.g. practical rationality). Material normative sources are those that require a specific content to be present in an intentional state. I posit theoretical type of normative source as requiring certain beliefs, and practical type of normative source as requiring certain desires and decisions (intentions).

Preliminary steps

- Metanormative theory speaks about the language $\mathcal{L}_{\mathrm{n}}$, the language in which the norm-contents are expressed.
- $\mathcal{L}_{\mathrm{n}}$ is a language of propositional logic whose formative syntax also allows modalities: B_i for 'i believes that', D_i for 'i desires that', I_i for 'i intends that'.

Definition

The normative language $\mathcal{L}_{\mathrm{n}}$ is built over the base language of propositional logic $\mathcal{L}_{PL}$:

$$\text{Sentences of } \mathcal{L}_{PL} ::= \text{propositional letters} \mid \neg\varphi \mid (\varphi \wedge \psi)$$

Let $i \in A$, $\mathrm{X} = \mathrm{B}, \mathrm{D}, \mathrm{I}$, and $p \in \mathcal{L}_{PL}$

$$\text{Sentences of } \mathcal{L}_{\mathrm{n}} ::= p \mid [\mathrm{X}_i]\varphi \mid \neg\varphi \mid (\varphi \wedge \psi)$$

The definitions of truth-functional connectives are standard.

Quasi-literals

Remark

The sentences of $\mathcal{L}_{\mathrm{n}}$ whose main operator is $[\mathrm{B}_i]$, $[\mathrm{D}_i]$, or $[\mathrm{I}_i]$ will be termed 'modals'.

Definition

The set $\mathrm{lit}(\mathcal{L}_{\mathrm{n}})$ of quasi-literals with respect to propositional logic is the smallest subset of $\mathcal{L}_{\mathrm{n}}$ containing the set of propositional letters and their negations, and the set of modals and their negations.

Considered in isolation, the language $\mathcal{L}_{\mathrm{n}}$ is not committed to any particular logic. Still, if a subset of $\mathcal{L}_{\mathrm{n}}$ has a logical property definable within some particular logic $< l >$, than that property will be noted as $< l >$-property.

Extension to infinitary language

For theoretical purposes the language $\mathcal{L}_{\mathrm{n}}$ will be extended to the language $\mathcal{L}_{\mathrm{n}(\omega_1)}$ of a variant of infinitary logic which has the same symbols as $\mathcal{L}_{\mathrm{n}}$, but in $\mathcal{L}_{\mathrm{n}(\omega_1)}$ the conjunction symbol $\bigwedge$ may be applied to subsets of the set of literals $\mathrm{lit}(\mathcal{L}_{\mathrm{n}})$.

Definition

Let $p \in \mathcal{L}_{\mathrm{n}}$ and let $x \subseteq \mathrm{lit}(\mathcal{L}_{\mathrm{n}})$ be countably infinite.

$$\text{Sentences of } \mathcal{L}_{\mathrm{n}(\omega_1)} ::= p \mid \bigwedge x \mid \neg\varphi \mid (\varphi \wedge \psi)$$

Extension of deduction rules and valuation function to infinitary propositional logic

For theoretical purposes the deductive system $\vdash_{\mathrm{pl}}$ (e.g. natural deduction) of propositional logic will be extended to an *ad hoc* variant of infinitary propositional logic $\vdash_{\mathrm{pl}(\omega_1)}$ containing the rules of $\vdash_{\mathrm{pl}}$ and the additional rules for the countably infinite conjunctions of literals.
According to the grammar of $\mathcal{L}_{\mathrm{n}}$ the introduction and elimination rules for $\bigwedge$ are applicable to the sets of literals only:
for $x \subseteq lit(\mathcal{L}_n)$,

1. $\Gamma, \bigwedge x \vdash_{\mathrm{pl}(\omega_1)} p$ for all $p \in x$, and
2. if $\Gamma \vdash_{\mathrm{pl}(\omega_1)} p$ for all $p \in x$, then $\Gamma \vdash_{\mathrm{pl}(\omega_1)} \bigwedge x$

On the side of semantics, the definition of the truth assignment $\hat{h}$ is extended in an obvious way: $\hat{h}(\bigwedge x) = \mathrm{t}$ iff $\hat{h}(p) = \mathrm{t}$ for all $p \in x$.

Conservative extension

The *ad hoc* system $\vdash_{\mathrm{pl}(\omega_1)}$ is a conservative extension of $\vdash_{\mathrm{pl}}$.

Proposition

For $x \cup \{p\} \subseteq \mathcal{L}_{\mathrm{n}}$, if $x \vdash_{\mathrm{pl}(\omega_1)} p$, then $x \vdash_{\mathrm{pl}} p$.

Proof.

The proof will be sketched. Assume $x \vdash_{\mathrm{pl}(\omega_1)} p$. The deductive system $\vdash_{\mathrm{pl}(\omega_1)}$ is sound, as can be easily checked. Therefore, $x \models_{\mathrm{pl}(\omega_1)} p$. Then also $x \models_{\mathrm{pl}} p$ thanks to coincidence of the semantic definitions for sentences in $\mathcal{L}_{\mathrm{n}}$. Finally, $x \vdash_{\mathrm{pl}} p$ by the completeness of the propositional logic. $\square$

Multi-sorted first order language $\mathcal{L}_{\mathrm{meta}}$

In order to achieve technical clarity we define a first-order metanormative language $\mathcal{L}_{\mathrm{meta}}$ in which variables of different sorts range over different objects in the domain with the following **extralogical vocabulary**:

- individual constants for normative sources, agents and worlds: $\mathrm{s}, \mathrm{s}_1, \ldots, \mathrm{a}, \mathrm{a}_1, \ldots, \mathrm{v}, \mathrm{v}_1, \ldots$;
- function symbols for code of requirement, propositional logic consequence, and "axiomatic basis (of a modal logic)" function: k^3, Cn^1, l^1;
- function symbols for generating sentential forms occurring in the object language: neg^1, conj^2, and a set of symbols $\mathrm{mod}^1_{\mathrm{B}i}$, $\mathrm{mod}^1_{\mathrm{D}i}$, $\mathrm{mod}^1_{\mathrm{I}i}$ for each $i = \mathrm{a}, \mathrm{a}_1, \ldots$, and $\mathrm{infconj}^2$;
- function symbols for extraction of "quasi-literals" from a given set: lt^1;
- dyadic predicate symbol for relation of membership: $\in^2$;
- and additionally we may introduce a dispensable part of vocabulary containing monadic predicate symbols expressing properties of being a normative source, an agent, a sentence in $\mathcal{L}_{\mathrm{n}}$, a possible world: Source^1, Ag^1, Sen^1, W^1.

Sorts of variables and shortcut notations

Variables comprise:

- "general variables" ranging over everything: $x, x_1, \ldots, y, y_1, \ldots$;
- sorts of variables:
  - $i, i_1, \ldots$ ranging over $\{x \in \mathrm{D} \mid Ag(x)\}$;
  - $p, p_1, \ldots, q, q_1, \ldots$ ranging over $\{x \in \mathrm{D} \mid \mathrm{Sent}(x)\}$;
  - $w, w_1, \ldots$ ranging over $\{x \in \mathrm{D} \mid W(x)\}$.

Notation

The shorthand notations for $\mathrm{neg}(p)$, $\mathrm{conj}(p, q)$, $\mathrm{mod}_{\mathrm{B}i}(p)$, $\mathrm{mod}_{\mathrm{D}i}(p)$, $\mathrm{mod}_{\mathrm{I}i}(p)$, $\mathrm{infconj}(x)$ are $\ulcorner \neg p \urcorner$, $\ulcorner p \wedge q \urcorner$, $\ulcorner [\mathrm{B}_i]p \urcorner$, $\ulcorner [\mathrm{D}_i]p \urcorner$, $\ulcorner [\mathrm{I}_i]p \urcorner$, $\ulcorner \bigwedge x \urcorner$. For the ease of reading "Quine quotes" will be used also for the standardly defined connectives, e.g. $\ulcorner p \rightarrow q \urcorner$ for $\mathrm{neg}(\mathrm{conj}(p, \mathrm{neg}(q)))$.
The sole variable written between "Quine quotes" is the same as the variable itself. Sometimes this redundant notation will be (ab)used in order to emphasize sentence variables and sentence functions within a formula.

Standard definitions of terms, formulas and sentences in $\mathcal{L}_{\mathrm{meta}}$

Definitions

Let c stand for individual constant, v for any variable, f for function symbol and P^n for predicate symbol.

$$\begin{array}{rcl} \text{terms } (t) & ::= & \mathrm{c} \mid \mathrm{v} \mid f(t_1, \ldots, t_n) \\ \text{atomic formulas } (p) & ::= & P^n(t_1, \ldots, t_n) \end{array}$$

Let p be an atomic formula

$$\text{Formulas of } \mathcal{L}_{\mathrm{meta}} ::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid \forall \mathrm{v}\ \varphi$$

Sentences of $\mathcal{L}_{\mathrm{meta}}$ are formulas of $\mathcal{L}_{\mathrm{meta}}$ with all variables bound.

Objects in the domain

The purpose of the metanormative language is to enable talking:

- about the syntax of sentences in $\mathcal{L}_{\mathrm{n}(\omega_1)}$,
- about the properties of sentences and their sets that they could have in different logics (most notably "world logic" and "intentionality logics"),
- about the semantics of sentences in $\mathcal{L}_{\mathrm{n}(\omega_1)}$, i.e. about sentence-world relation.

- The basic ontology for the code functions requires: normative sources, agents, worlds and sets of sentences.
- Some objects will be constructed using $\mathcal{L}_{\mathrm{n}(\omega_1)}$ sentences:
  - the worlds, which are theoretically identified with proposition-logically maximal consistent sets of $\mathcal{L}_{\mathrm{n}(\omega_1)}$,
  - code values, which are "logic free" sets of sentences,
  - axiomatic bases of logics, which are sets of substitutional instances of the sentences in a given set,
  - sentences, which **are** sentences.
  -

Worlds free of modal logic

Definition

$\mathrm{MaxCon}(\mathcal{L}_{\mathrm{n}(\omega_1)}) = \{x \subseteq \mathcal{L}_{\mathrm{n}(\omega_1)} \mid x \nvdash_{\mathrm{pl}(\omega_1)} \bot, \forall y \in \mathcal{L}_{\mathrm{n}(\omega_1)}(y \notin x \rightarrow x \cup \{y\} \vdash_{\mathrm{pl}(\omega_1)} \bot)\}$ is the set of possible worlds.

No modal axiom for belief, desire or intention does hold in all the possible worlds. Therefore, any kind and any measure of violations of logics of intentionality may occur.

Unbounded irrationality

Quote

What sets a limit to the amount of irrationality we can make psychological sense of is a purely conceptual or theoretical matter—the fact that mental states and events are the states and events they are by their location in a logical space. Donald Davidson (2004) *Problems of Rationality*. Clarendon Press, Oxford, p. 183

In the modeling the worlds characterized by an extreme "amount of irrationality" on the side of an agent *i* are admitted in the modeling. This fact should not be interpreted as an violation of Davidson's thesis, rather it should be understood as an unrealistic but harmless and dispensable theoretical possibility.

No place for axiom T

- The T axiom ($\Box p \rightarrow p$) poses a serious threat to the modeling that keeps modality and world apart.
- If modalities obeying "reflexive" axiom T are allowed, then possible worlds, *i.e.* maximal consistent sets in propositional logic, would become intuitively impossible.
- For example, although $\{p, [\mathrm{K}]_i \neg p\}$ is pl-consistent set, we do not want to have it included in any world since no false proposition may be known as a true proposition.
- T axioms constitute an important part of the meaning of verbs of knowledge and of action. So, epistemic and praxeological modalities must be excluded from the language of norms $\mathcal{L}_{\mathrm{n}}$.
- The exclusion strategy may seem drastic. The forthcoming analysis does not depend on the inclusion of "T modalities", so this strategy may be adopted as a provisional method.

Content of the norm

- Von Wright (1963, Norm and Action : A Logical Enquiry) defined 'content of a norm' as "that which ought to or may or must not be or be done".
- The normative language $\mathcal{L}_{\mathrm{n}(\omega_1)}$ departs from von Wright's definition by taking norm-content to be *the intentional state or relation of intentional states that ought to or may or must not be present in the mind of the norm addressee on a particular occasion*.
- The reduction and the switch may seem drastic but there is a rationale for it.
  - The requirement that agent i knows that p could be replaced by $p \to [\mathrm{B}_i]p$; a required action to see to it that p could be replaced by the required intention, i.e. $[\mathrm{I}_i]p$.
  - Actions if successful require "cooperation of the world," and "the world" is not a norm addressee.

Sets of requirements as logic free theories

- Broome claims that code values are closed under pl-equivalence and he seems to tacitly hold that this congruence[2] property constitutes the whole of the logic of "source requirements". A recent proponent is Loui Goble (2009, Normative Conflicts and The Logic of Ought, *Noûs*:**43**).
- Broome bases the acceptability of the congruence principle on the argument from the absence of contrary evidence, while Goble takes it for granted since "[it] seems [to be] a minimum requirement for a logic of ought," (p. 483).
- On the other hand, Alchourrón and Bulygin [1] propose an approach that is both more restrictive and more permissive. First, contrary to Broome's weak, "congruence logic", Alchourrón and Bulygin argue that there is *no logic of norms* since the existence of a norm depends on the empirical fact of promulgation. Second, they claim that *there is a logic of normative systems* since the set of norm-contents is deductively closed.
- It seems that there is no *consensus* and no conclusive reason for presupposing the existence of any particular logical property of code values. On the other hand, a "logic free" conception of code values operates at a higher level of generality.

[2] If p and q are equivalent in propositional logic, then p is a member of a code just in case q

Axiomatic bases for logics of intentionality

If the normative character the intentionality realm consists in its subjection to requirements of different normative sources and if rationality is a normative source, then some logic for rational relations between intentional states will be needed. On the other hand, a code function may deliver sets having logical properties other from those definable within propositional logic.

Sets of substitutional instances

The uniform substitutions are restricted to the finitary, i.e. $\mathcal{L}_{\mathrm{n}}$ part of the metanormative language since infintary sentences are not allowed to embed within each other.

Definitions

Any function g from $\mathcal{L}_{\mathrm{n}}$ to $\mathcal{L}_{\mathrm{n}(\omega_1)}$ is a substitution function iff (i) $g(p) \in \mathcal{L}_{\mathrm{n}(\omega_1)}$ if p is a propositional letter, (ii) $g(\neg p) = \neg g(p)$, (iii) $g(p \wedge q) = g(p) \wedge g(q)$, (iv) $g([\mathrm{X}_i]p) = [\mathrm{X}_i]g(p)$ for $\mathrm{X} = \mathrm{B}, \mathrm{D}, \mathrm{I},\ i \in \mathrm{A}$. The set Sb is the set of all substitution functions. The set of all substitutional instances of the sentences in a given set $x \subseteq \mathcal{L}_{\mathrm{n}}$ is the set $\mathrm{l}(x) = \{q \mid \exists p \exists f (p \in x \wedge f \in \mathsf{Sb} \wedge f(p) = q)\}$.

Definition

The set $\mathrm{Cn}(\mathrm{l}(x)) = \{p \mid \exists y (y \subseteq \mathrm{l}(x) \wedge y \vdash_{\mathrm{pl}(\omega_1)} p\}$ is the logic for axiomatic basis x.

Digression: normal logics

Definition

Let $\top_{[\mathrm{X}_i]}$ and $\mathrm{K}_{[\mathrm{X}_i]}$ denote axiom schemata $((p \vee \neg p) \leftrightarrow q) \rightarrow [\mathrm{X}_i]q$ and $[\mathrm{X}_i](p \rightarrow q) \rightarrow ([\mathrm{X}_i]p \rightarrow [\mathrm{X}_i]q)$ respectively. A set $\mathrm{Cn}(\mathrm{l}(x))$ is a normal logic for a set of modal operators $\mathrm{o}/x \subseteq \{[\mathrm{X}_i] \mid \mathrm{X} = \mathrm{B}, \mathrm{D}, \mathrm{I},\ i \in \mathrm{A},\ [\mathrm{X}_i]$ occurs in some $p \in x\}$ iff

$$\mathrm{Cn}(\mathrm{l}(\{\top_x \mid x \in \mathrm{o}/x\} \cup \{\mathrm{K}_x \mid x \in \mathrm{o}/x\})) \subseteq \mathrm{Cn}(\mathrm{l}(x))$$

The domain

- The domain for metanormative language $\mathcal{L}_{\mathrm{meta}}$ comprises the following objects:
  - normative sources, $x \in \mathrm{S}$
  - agents, $x \in \mathrm{A}$
  - sentences, $x \in \mathcal{L}_{\mathrm{n}(\omega_1)}$,
  - sets of sentences (code values, and axiomatic bases for logics), $x \in \wp\mathcal{L}_{\mathrm{n}(\omega_1)}$
  - worlds, $x \in \mathrm{MaxCon}(\mathcal{L}_{\mathrm{n}(\omega_1)}) \subseteq \wp\mathcal{L}_{\mathrm{n}(\omega_1)}$

Definition

$\mathrm{D} = \mathrm{S} \cup \mathrm{A} \cup \mathcal{L}_{\mathrm{n}(\omega_1)} \cup \wp\mathcal{L}_{\mathrm{n}(\omega_1)}$
where $\mathrm{S} \neq \emptyset$, $\mathrm{A} \neq \emptyset$, $\mathrm{O} \cap \mathrm{A} = \emptyset$.

Interpretation function (1st part)

Interpretation function $\mathcal{I}$ for $\mathcal{L}_{\mathrm{meta}}$:

- (Names of sources) $\mathcal{I}(\mathrm{s}_i) \in \mathrm{S}$,
- (Code function) $\mathcal{I}(\mathrm{k})$ is a function: $\mathrm{S} \times \mathrm{A} \times \mathrm{MaxCon}(\mathcal{L}_{\mathrm{n}(\omega_1)}) \to \wp\mathcal{L}_{\mathrm{n}(\omega_1)}$,
- (Axiomatic basis function) $\mathcal{I}(\mathrm{l})$ is a function: $\wp\mathcal{L}_{\mathrm{n}} \to \wp\mathcal{L}_{\mathrm{n}(\omega_1)}$, such that for any $x \subseteq \mathcal{L}_{\mathrm{n}}$, $\mathrm{l}(x) = \{f(p) \mid p \in x \wedge f \in \mathrm{Sb}\}$
- (pl-consequence function) $\mathcal{I}(\mathrm{Cn})$ is a function: $\wp\mathcal{L}_{\mathrm{n}(\omega_1)} \to \wp\mathcal{L}_{\mathrm{n}(\omega_1)}$, such that for any $x \subseteq \mathcal{L}_{\mathrm{n}(\omega_1)}$, $\mathrm{Cn}(x) = \{y \in \mathcal{L}_{\mathrm{n}(\omega_1)} \mid x \vdash_{\mathrm{pl}(\omega_1} y\}$.

Interpretation function (2nd part)

Sentence forms

- $\mathcal{I}(\mathrm{neg})$, $\mathcal{I}(\mathrm{conj})$, $\mathcal{I}(\mathrm{mod}_{\mathrm{B}i})$, $\mathcal{I}(\mathrm{mod}_{\mathrm{D}i})$, $\mathcal{I}(\mathrm{mod}_{\mathrm{I}i})$, $\mathcal{I}(\mathrm{infconj})$ are functions: $\mathcal{L}_{\mathrm{n}(\omega_1)} \to \mathcal{L}_{\mathrm{n}(\omega_1)}$, such that

$$
\begin{array}{rcl}
\mathcal{I}(\mathrm{neg}) & = & \{\langle x, y\rangle \mid y = \neg \frown x\} \\
\mathcal{I}(\mathrm{conj}) & = & \{\langle x, y, z\rangle \mid z = x \frown \wedge \frown y\} \\
\mathcal{I}(\mathrm{mod}_{\mathrm{B}i}) & = & \{\langle x, y\rangle \mid y = \mathrm{B}_i \frown x\} \\
\mathcal{I}(\mathrm{mod}_{\mathrm{D}i}) & = & \{\langle x, y\rangle \mid y = \mathrm{D}_i \frown x\} \\
\mathcal{I}(\mathrm{mod}_{\mathrm{I}i}) & = & \{\langle x, y\rangle \mid y = \mathrm{I}_i \frown x\} \\
\mathcal{I}(\mathrm{infconj}) & = & \left\{ \langle x, y\rangle \mid \begin{array}{c} x \subseteq \mathrm{lt}(\mathcal{L}_{\mathrm{n}}) \wedge \\ y = seq(x)(1) \frown \wedge \frown \\ \ldots \frown \wedge \frown seq(x)(n) \frown \wedge \frown \ldots) \end{array} \right\}
\end{array}
$$

where $seq(x) \in \prod_{i,j \in \mathbb{N}}^{funct} x$

and where $\prod_{i \in \mathbb{N}}^{funct} x = \{f : \mathbb{N} \to x | \forall i \forall j \; f(i) \neq f(j)\}$.

Interpretation function (3rd part)

- (extraction of quasi-literals) lt is the function: $\wp\mathcal{L}_{\mathrm{n}(\omega_1)} \to \wp\mathcal{L}_{\mathrm{n}(\omega_1)}$ such that for any $x \subseteq \mathcal{L}_{\mathrm{n}(\omega_1)}$, $\mathrm{lt}(x) = \{y \in \mathcal{L}_{\mathrm{n}} \mid y \in x \wedge y \in \mathrm{lit}(\mathcal{L}_{\mathrm{n}})\}$,
- (superfluous predicates)
  - (Source predicate) $\mathcal{I}(\mathrm{Source}) = \mathrm{S}$,
  - (Agent predicate) $\mathcal{I}(A) = \mathrm{A}$,
  - (Sentence predicate) $\mathcal{I}(\mathrm{Sen}) = \mathcal{L}_{\mathrm{n}}$,
  - (World predicate) $\mathcal{I}(\mathrm{W}) = \mathrm{MaxCon}(\mathcal{L}_{\mathrm{n}})$,
- (Relation of having a property corresponding to a source s in a world) $\mathcal{I}(\mathrm{K_s}) \subseteq \mathrm{O} \times \mathrm{MaxCon}(\mathcal{L}_{\mathrm{n}(\omega_1)})$,
- (Membership relation) $\mathcal{I}(\in) \subseteq \mathcal{L}_{\mathrm{n}(\omega_1)} \times \wp\mathcal{L}_{\mathrm{n}(\omega_1)} \cup \mathrm{A} \times \wp\mathrm{A} \cup \mathrm{S} \times \wp\mathrm{S}$.

Variable assignment

Definition

$\mathfrak{M}_{\mathrm{mn}} = \langle \mathrm{D}, \mathcal{I} \rangle$

Definition

Variable assignment g in $\mathfrak{M}_{\mathrm{mn}} = \langle \mathrm{D}, \mathcal{I} \rangle$ is possibly partial function g such that for any variable v

$$g(v) \in \mathrm{D} \text{ if } v \in \mathit{domain}(g)$$

For sorts of variables: (world variables) $g(v) \in \mathrm{MaxCon}(\mathcal{L}_{\mathrm{n}})$ if $v = w, w_1, \ldots$; (sentence variables) $g(v) \in \mathcal{L}_{\mathrm{n}}$ if $v = p, p_1, \ldots, q, q_1, \ldots$, (agent variables) $g(v) \in \mathrm{A}$.

Definition

The variable assignment g is appropriate for formula p iff all free variables in p are in the domain of g .

Singular terms

Notation

By $g_{[x/d]}$ we denote the variable assignment that differs from g on at most x:

$$g_{[x/d]}(v) = \begin{cases} g(v), \text{ if } x \neq v \\ d, \text{ otherwise.} \end{cases}$$

The special case of the empty variable assignment $g_{\emptyset}$:

$$range(g_{\emptyset}) = \emptyset$$

Definitions

$$\mathcal{I}(f)(x_1, .., x_n) = \begin{cases} y, \text{ if } \langle x_1, .., x_n, y \rangle \in \mathcal{I}(f) \\ \text{undefined, otherwise.} \end{cases}$$

$$[\![t]\!]_g^{\mathfrak{M}_{mn}} = \begin{cases} \mathcal{I}(t), \text{ if } t \text{ is an individual constant} \\ g(t), \text{ if } t \text{ is an individual variable} \\ \mathcal{I}(f)([\![t_1]\!]_g^{\mathfrak{M}_{mn}}, \ldots, [\![t_n]\!]_g^{\mathfrak{M}_{mn}}), \text{ if } t \text{ is } f(t_1, ..., t_n) \end{cases}$$

Satisfaction and truth in $\mathfrak{M}_{mn}$

Definition

(Satisfaction) Let g be an assignment in $\mathfrak{M}_{\mathrm{mn}}$ that is appropriate for the formulas being evaluated.

$$\begin{array}{lll}
\mathfrak{M}_{mn} & \models & P(t_1, ..., t_n) \; [g] \text{ iff } \langle [\![t_1]\!]_g^{\mathfrak{M}_{mn}}, \ldots, [\![t_n]\!]_g^{\mathfrak{M}_{mn}} \rangle \in \mathcal{I}(P) \\
\mathfrak{M}_{mn} & \models & \neg\varphi \; [g] \text{ iff not } \mathfrak{M}_{mn} \models \varphi \; [g] \\
\mathfrak{M}_{mn} & \models & (\varphi \wedge \psi) \; [g] \text{ iff } \mathfrak{M}_{mn} \models \varphi \; [g] \text{ and } \mathfrak{M}_{mn} \models \psi \; [g] \\
\mathfrak{M}_{mn} & \models & \forall v \, \varphi \; [g] \text{ iff for all } \mathrm{d} \in \mathrm{D}, \mathfrak{M}_{mn} \models \varphi \; [g_{[v/d]}]
\end{array}$$

Definition (Truth in a metanormative model)

Formula φ is true in $\mathfrak{M}_{mn}$ iff $g_{\emptyset}$ satisfies φ in $\mathfrak{M}_{mn}$: $\mathfrak{M}_{mn} \models \varphi$ iff $\mathfrak{M}_{mn} \models \varphi \, [g_{\emptyset}]$

Properties of a code within "world logic"

Quantifications over different argument positions in the code function enable a number of interesting type distinctions, some of which will be introduced below using a $\mathcal{L}_{\mathrm{meta}}$ formula in the *definiens*.

Some pl logical properties of codes:

- $\mathrm{k_s}$ is a pl-*congruent* code iff

$$\ulcorner p \leftrightarrow q \urcorner \in \mathrm{Cn}(\varnothing) \rightarrow (p \in \mathrm{k_s}(i, w) \leftrightarrow q \in \mathrm{k_s}(i, w))$$

- $\mathrm{k_s}$ is a pl-*consistent* code iff

$$\exists w_2 \, \mathrm{k_s}(i, w_1) \subseteq w_2$$

- $\mathrm{k_s}$ is a pl-*deductively closed* iff

$$\mathrm{k_s}(i, w) = \mathrm{Cn}(\mathrm{k_s}(i, w))$$

Codes and the world

Some world related code properties:

- $\mathrm{k_s}$ code is *futile* iff

$$\forall i \forall w \ \mathrm{k_s}(i, w) \subseteq w$$

- $\mathrm{k_s}$ is an *achievable* code iff

$$\exists w \, \mathrm{k_s}(i, w) \subseteq w$$

- $\mathrm{k_s}$ is a *relativistic* code iff

$$\exists i \exists w_1 \exists w_2 \, \mathrm{k_s}(i, w_1) \neq \mathrm{k_s}(i, w_2)$$

- a code is *absolute* iff it is not relativistic

Social aspects of codes

Some social code properties:

- $\mathrm{k_s}$ is a *socially consistent* code iff

$$\exists w_2 \, \mathrm{k_s}(i_1, w_1) \cup \mathrm{k_s}(i_2, w_1) \subseteq w_2$$

- $\mathrm{k_s}$ is an *socially achievable* code (for group G) iff

$$\exists w \, \forall i (i \in G \rightarrow \mathrm{k_s}(i, w) \subseteq w)$$

Relations between normative sources

Some relations between normative sources:

- codes k_{s_m} and k_{s_n} are *realization-equivalent* iff

$$\mathrm{k}_{s_m}(i, w) \subseteq w \leftrightarrow \mathrm{k}_{s_n}(i, w) \subseteq w$$

- codes k_{s_m} and k_{s_n} are *compatible* iff

$$\exists w_1 \exists w_2\, \mathrm{k}_{s_m}(i, w_1) \cup \mathrm{k}_{x_n}(i, w_1) \subseteq w_2$$

- code k_{s_m} is *maximally compatible* iff

$$\forall x \exists w_1 \exists w_2\, \mathrm{k}_{s_m}(i, w_1) \cup \mathrm{k}_{x}(i, w_1) \subseteq w_2$$

Codes and logics

Some $\mathrm{l}(x)$ logical properties of codes:

- code is *consistent with respect to* $\mathrm{l}(x)$ iff
$$\exists w_2 \ \mathrm{Cn}(\mathrm{l}(x) \cup \mathrm{k_s}(i, w_1)) \subseteq w_2$$
- code is *logic* iff
$$\exists x \ \mathrm{k_s}(i, w) = \mathrm{Cn}(\mathrm{l}(x))$$
- code is *deductively closed with respect to logic* $\mathrm{Cn}(\mathrm{l}(x))$ ("more than a logic") iff
$$\exists x \exists y (\neg y \subseteq \mathrm{Cn}(\mathrm{l}(x)) \wedge \mathrm{k_s}(i, w) = \mathrm{Cn}(\mathrm{l}(x) \cup y))$$
- code is *less than a logic* iff
$$\exists x \exists y (\neg y \subseteq \mathrm{Cn}(\mathrm{l}(x)) \wedge \mathrm{k_s}(i, w) = \mathrm{Cn}(\mathrm{l}(x) \cup y) - \mathrm{Cn}(\mathrm{l}(x))).$$

KD deontic logic without iterated modalities

Definition

Let $p \in \mathcal{L}_{PL}$ be a formula of propositional logic:

$$\text{Formulas of } \mathcal{L}^{\mathrm{O}}_{\mathrm{KD}} ::= p \mid \mathrm{O}p \mid \mathrm{P}p \mid \neg\varphi \mid (\varphi \wedge \psi)$$

Let us introduce the translation τ^1 from the restricted language $\mathcal{L}^{\mathrm{O}}_{\mathrm{KD}}$ to the metanormative language $\mathcal{L}_{\mathrm{meta}}$, where $\mathrm{O}\,p$ and $\mathrm{P}\,p$ will be translated as 'a in v has s-obligation (s-permission) to p'.

Two-step translation: typo

Remark

Typo in the booklet: instead in τ "Quine quotes" should have been introduced in τ^1.

Two-step translation

Definitions

Function τ maps sentences from the fragment $\mathcal{L}^{\mathrm{O}}_{\mathrm{KD}} \cap \mathcal{L}_{PL}$ to the set of sentential variables and sentential function terms of $\mathcal{L}_{\mathrm{meta}}$:

$$\begin{array}{rcll} \tau(l) & \in & \{p, p_1, \ldots, q, q_1, \ldots\} & \text{for propositional letters } l \in \mathcal{L}_{PL} \\ \tau(\neg\varphi) & = & \neg\tau(\varphi) & \\ \tau(\varphi \wedge \psi) & = & (\tau(\varphi) \wedge \tau(\psi)) & \end{array}$$

Translation $\tau^1 : \mathcal{L}^{\mathrm{O}}_{\mathrm{KD}} \to \mathcal{L}_{\mathrm{meta}}$

$$\begin{array}{rcll} \tau^1(p) & = & \ulcorner\tau(p)\urcorner \in \mathrm{v} & \text{if } p \in \mathcal{L}_{PL} \\ \tau^1(\mathrm{O}\varphi) & = & \ulcorner\tau(\varphi)\urcorner \in \mathrm{k_s}(\mathrm{a}, \mathrm{v}) & \\ {}^1(\mathrm{P}\varphi) & = & \ulcorner\tau(\neg\varphi)\urcorner \notin \mathrm{k_s}(\mathrm{a}, \mathrm{v}) & \\ \tau^1(\neg\varphi) & = & \neg\tau^1(\varphi) & \\ \tau^1(\varphi \wedge \psi) & = & (\tau^1(\varphi) \wedge \tau^1(\psi)) & \end{array}$$

An example

Example

$$
\begin{array}{rcl}
\tau^1\left(\mathrm{O}\neg p \rightarrow \mathrm{O}\neg(p \wedge q)\right) & \Leftrightarrow & \tau^1(\mathrm{O}\neg p) \rightarrow \tau^1(\mathrm{O}\neg(p \wedge q)) \\
& \Leftrightarrow & \ulcorner\tau(\neg p)\urcorner \in \mathrm{k_s}(\mathrm{a},\mathrm{v}) \rightarrow \ulcorner\tau(\neg(p \wedge q))\urcorner \in \mathrm{k_s}(\mathrm{a},\mathrm{v}) \\
& \Leftrightarrow & \ulcorner\neg\tau(p)\urcorner \in \mathrm{k_s}(\mathrm{a},\mathrm{v}) \rightarrow \ulcorner\neg\tau(p \wedge q)\urcorner \in \mathrm{k_s}(\mathrm{a},\mathrm{v}) \\
& \Leftrightarrow & \ulcorner p\urcorner \in \mathrm{k_s}(\mathrm{a},\mathrm{v}) \rightarrow \ulcorner\neg(\tau(p) \wedge \tau(q))\urcorner \in \mathrm{k_s}(\mathrm{a},\mathrm{v}) \\
& \Leftrightarrow & \ulcorner p\urcorner \in \mathrm{k_s}(\mathrm{a},\mathrm{v}) \rightarrow \ulcorner\neg(p \wedge q)\urcorner \in \mathrm{k_s}(\mathrm{a},\mathrm{v})
\end{array}
$$

Narrow and wide scope reading of conditional obligation

Example

There are two interpretations of conditional obligation in standard deontic logic: (N-scope) "narrow scope interpretation" ("if p is the case, then q ought to be the case"): $p \rightarrow \mathrm{O}q$, 2. (W-scope) "wide scope interpretation" ("it ought to be the case that: if p is the case, then q is the case") or $\mathrm{O}(p \rightarrow q)$. Narrow scope formula, i.e. $p \rightarrow \mathrm{O}q$ is translated as $p \in \mathrm{v} \rightarrow q \in \mathrm{k_s}(\mathrm{a}, \mathrm{v})$. Wide scope formula, i.e. $\mathrm{O}(p \rightarrow q)$ is translated as $\ulcorner p \rightarrow q \urcorner \in \mathrm{k_s}(\mathrm{a}, \mathrm{v})$.
There is a tendency between natural language speakers to consider (N-scope) and (W-scope) expressions as equivalent.

Wide and narrow scope reading *vs.* absolute and relativistic codes

The impression of equivalence in meaning is justified by two theoretically derived facts.

1. A code $\mathrm{k_s}(\mathrm{a}, \mathrm{v})$ has its conditionalized variant $\mathrm{k_s}(\mathrm{a}, \mathrm{v})$ (Proposition below)

$$\underbrace{(\ulcorner p \urcorner \in w \rightarrow \ulcorner q \urcorner \in \mathrm{k_s}(\mathrm{a}, w))}_{\text{narrow scope}}$$
$$\leftrightarrow$$
$$\underbrace{(\ulcorner p \urcorner \in \mathrm{Cn}(\{\ulcorner \bigwedge \mathrm{lt}(w) \urcorner\}) \rightarrow \ulcorner \bigwedge \mathrm{lt}(w) \rightarrow q \urcorner \in \mathrm{k}_{\mathsf{s}}^{cond}(\mathrm{a}, w))}_{\text{wide scope (generalized)}}$$

   (right side translation: in all *p*-worlds there is a norm content stating that in a such a world *q* is a case).

2. A code and its conditionalized variant are realization equivalent. Therefore, from the behavioristic of view there is no difference between them.

KD as a descriptive theory of consistent and deductively closed code values

Quote

. . . classical deontic logic, on the descriptive interpretation of its formulas, pictures a gapless and contradiction-free system of norms.
Von Wright [1999] p. 32

According to our translation scheme von Wright's claim should be appended: classical deontic logic "pictures a system of norms" that is deductively closed too.

The translation of the axioms

- "gaplessness" condition $\mathrm{P}p \vee \mathrm{O}\neg p$ translates to $\ulcorner \neg p \urcorner \notin \mathrm{k_s}(\mathrm{a}, \mathrm{v}) \vee \ulcorner \neg p \urcorner \in \mathrm{k_s}(\mathrm{a}, \mathrm{v})$ and that property obviously holds for any set of requirements whatsoever;
- K axiom becomes $\ulcorner p \rightarrow q \urcorner \in \mathrm{k_s}(\mathrm{a}, \mathrm{v}) \rightarrow (p \in \mathrm{k_s}(\mathrm{a}, \mathrm{v}) \rightarrow q \in \mathrm{k_s}(\mathrm{a}, \mathrm{v}))$ and that property holds for any pl-deductively closed set;
- D axiom becomes $p \in \mathrm{k_s}(\mathrm{a}, \mathrm{v}) \rightarrow \ulcorner \neg p \urcorner \notin \mathrm{k_s}(\mathrm{a}, \mathrm{v})$ and that is just another way of stating pl-consistency;
- mutual definability, $\mathrm{P}_1 p \leftrightarrow \neg \mathrm{O} \neg p$ holds if the set of requirements is congruent.

Are there any problems?

One may ask whether these properties provide "an adequate" description of a formally sound set of requirements.

Example

For example, τ^1 translation for the D does not allow

$$[\mathrm{B}_i]p \wedge \neg[\mathrm{B}_i]p$$

to enter the set of requirements, but it does allow

$$[\mathrm{B}_i]p \wedge [\mathrm{B}_i]\neg p$$

So the question arises whether consistency property of a set of requirements is — a property that is connected to the logic of reality, or rather — a property that a set inherits when it obeys the logic of its contents (i.e. Logic of intentionality)?

Iterated operators

Although iterated deontic operators receive no translation in the scheme proposed above, one may extend the line of thought by giving additional translation rules for language of standard deontic $\mathcal{L}_{\mathrm{KD}}^{\mathrm{OO}}$ restricted to the maximum of two iterations of deontic operators, treating iterated deontic modalities as a sequence of heterogenous operators and introducing the distinction into the syntax:

$$\mathcal{L}_{\mathrm{KD}}^{\mathrm{O_2O}} ::= p \in \mathcal{L}_{\mathrm{KD}}^{\mathrm{O}} \mid \mathrm{O}_2 p \mid \mathrm{P}_2 p \mid \neg\varphi \mid (\varphi \wedge \psi)$$

Definition

Let $Sub(\varphi)_{[\frac{c_1}{x_1} \ldots \frac{c_n}{x_n}]}$ denote substitutional instance of $\varphi \in \mathcal{L}_{\mathrm{meta}}$ in which constants $c_1, \ldots, c_n$ are replaced by variables $x_1, \ldots, x_n$. Translation $\tau^2 : \mathcal{L}_{\mathrm{KD}}^{\mathrm{O_2O}} \rightarrow \mathcal{L}_{\mathrm{meta}}$

$$\begin{array}{rcll}
\tau^2(\mathrm{O}_2 p) & = & \forall i \forall w \; Sub(\tau^1(p))_{[\frac{\mathrm{a}}{i} \frac{\mathrm{v}}{w}]} & \text{for } p \in \mathcal{L}_{\mathrm{KD}}^{\mathrm{O}} \\
\tau^2(\mathrm{P}_2 p) & = & \exists i \exists w \; Sub(\tau^1(p))_{[\frac{\mathrm{a}}{i} \frac{\mathrm{v}}{w}]} & \text{for } p \in \mathcal{L}_{\mathrm{KD}}^{\mathrm{O}} \\
\tau^2(\neg\varphi) & = & \neg\tau^2(\varphi) & \\
\tau^2(\varphi \wedge \psi) & = & (\tau^2(\varphi) \wedge \tau^2(\psi)) &
\end{array}$$

Generalization over agents and worlds

Such an approach to iterated deontic modalities departs from von Wright's [3] "second order descriptive interpretation" where e.g. O_2 would stand for existence of "normative demands on normative systems" ("norms for the norm givers"). The "first order" translation τ^1 as well as the "second order" translation τ^2 give us statements in metanormative language $\mathcal{L}_{\mathrm{meta}}$ both of which may "picture" some type of "normative system". The difference lies in the fact that τ^1 gives a local picture of a set of requirements (for a particular source, agent and world) while τ^2 gives a more global picture of a code function. In the second case the properties depicted are the properties of a code function for a particular source with respect to any agent and any world.

Are iterations significant under translation?

Let us consider KD45 deontic logic! The τ^2 translations of reinterpreted axioms 4, $\mathrm{O}_1 p \to \mathrm{O}_2\mathrm{O}_1 p$ and 5, $\mathrm{P}_1 p \to \mathrm{O}_2\mathrm{P}_1 p$ amount to stating that any s-obligation and any s-permission holds universally. So, the reinterpreted axioms will hold only if s-code is absolute.

Definition

An agent i at world w has an "all-or-nothing" normative property K_s that corresponds to the source s iff the set of requirements $\mathrm{k}_\mathsf{s}(i, w)$ is satisfied in w, i.e. $\mathrm{K}_\mathsf{s}(i, w) \leftrightarrow \mathrm{k}_\mathsf{s}(i, w) \subseteq w$.

If the only way to satisfy some relativistic code and some absolute code is to satisfy them simultaneously, then these codes define the same normative property. The question arises as to whether (non)absoluteness of a code function introduces a difference with respect to normative properties. The next theorem provides a negative answer.

Conditionalization of codes

There is a number of ways to define a conditionalized variant of a code. Below we introduce one of the variants using an infinite conjunction of literals to single out a world, and assigning a conditional for each requirement.

Definition

A code k_s^{cond} is the conditionalized variant of a code k_s iff

$$\forall p \forall w_1 (p \in \mathrm{k}_s^{cond}(i, w_1) \leftrightarrow \exists q \exists w_2 (q \in \mathrm{k}_s(i, w_2) \wedge p = \ulcorner \bigwedge \mathrm{lt}(w_2) \rightarrow q \urcorner))$$

*Quasi*literals determine the world

Lemma

For all $p \in \mathcal{L}_{\mathrm{n}(\omega_1)}$, $\ulcorner p \urcorner \in \mathrm{Cn}(\mathrm{lt}(w))$ *or* $\ulcorner \neg p \urcorner \in \mathrm{Cn}(\mathrm{lt}(w))$.

Proof.

We use transfinite induction on the pl-complexity of formulas[a] We will consider the cases of limit ordinals. (0) The lemma holds for propositional letters and modal formulas in virtue of pl-maximality of w. (ω) Suppose p is $\bigwedge x$. According to the definition, any $p_i \in x$ is a quasi-literal, and by inductive hypothesis the lemma holds for each p_i. Either all quasi-literals in x are consequences of $\mathrm{lt}(w)$, and therefore $\ulcorner \bigwedge x \urcorner \in \mathrm{Cn}(\mathrm{lt}(w))$, or some of quasi-literals are not consequences of $\mathrm{lt}(w)$, and therefore $\ulcorner \neg \bigwedge x \urcorner \in \mathrm{Cn}(\mathrm{lt}(w))$. □

[a] We define the complexity of modal formulas and propositional letters to be 0; the complexity of $\neg p$ to be one greater than complexity of p; the complexity of $(p \wedge q)$ to be one greater than the maximum of that of p and q, the complexity of $\bigwedge x$ to be ω.

Proposition

$\mathrm{Cn}(\mathrm{lt}(w)) = w$

Proof.

First, suppose $p \in \mathrm{Cn}(\mathrm{lt}(w))$. Then, $p \in w$ since w is deductively closed. Second, suppose $p \in w$. By lemma, $\ulcorner p \urcorner \in \mathrm{Cn}(\mathrm{lt}(w)) \vee \ulcorner \neg p \urcorner \in \mathrm{Cn}(\mathrm{lt}(w))$, and so $\ulcorner p \urcorner \in \mathrm{Cn}(\mathrm{lt}(w))$ since w is consistent. $\square$

Definition

A code $\mathrm{k}_{\mathsf{s}}^{cond}$ is the conditionalized variant of a code k_{s} iff

$$\forall p \forall w_1 (p \in \mathrm{k}_{\mathsf{s}}^{cond}(i, w_1) \leftrightarrow \exists q \exists w_2 (q \in \mathrm{k}_{\mathsf{s}}(i, w_2) \wedge p = \ulcorner \bigwedge \mathrm{lt}(w_2) \rightarrow q \urcorner))$$

Lemma

Any conditionalized code is absolute.

Proof.

Let w_1 and w_2 be arbitrary worlds. Assume $p \in \mathrm{k}_{\mathsf{s}}^{cond}(i, w_1)$. Then, by definition of conditionalization, $\exists q \exists w_3 (q \in \mathrm{k}_{\mathsf{s}}(i, w_3) \wedge p = \ulcorner \bigwedge \mathrm{lt}(w_2) \rightarrow q \urcorner)$. Then, by (universal instantiation of) the same definition, $p \in \mathrm{k}_{\mathsf{s}}^{cond}(i, w_2)$. Obviously the same holds in the opposite direction. $\square$

Theorem

For each relativistic code there is a realization equivalent absolute code.

Proof.

Each relativistic code has its conditionalized counterpart. By lemma, each conditionalized code is absolute. It remains to prove that:

$$\mathrm{k}_{\mathsf{s}}(i, w) \subseteq w \leftrightarrow k_{\mathsf{s}}^{cond}(i, w) \subseteq w$$

□

L-R, 1st pt.

1		$\mathrm{k_s}(i, w_1) \subseteq w_1$	
2	p	$p \in \mathrm{k}_s^{cond}(i, w)$	
3		$\exists q \exists w_2 (q \in \mathrm{k_s}(i, w_2) \land p = \ulcorner \bigwedge \mathrm{lt}(w_2) \to q \urcorner)$	2/ def. of cond. code
4	q, w_2	$q \in \mathrm{k_s}(i, w_2) \land p = \ulcorner \bigwedge \mathrm{lt}(w_2) \to q \urcorner$	
5		$p = \ulcorner \bigwedge \mathrm{lt}(w_2) \to q \urcorner$	4/ $\land$ Elim
6		$\mathrm{Cn}(\mathrm{lt}(w_2)) = w_1 \lor \mathrm{Cn}(\mathrm{lt}(w_2)) \neq w_1$	Taut.
7		$\mathrm{Cn}(\mathrm{lt}(w_2)) = w_1$	
8		$q \in \mathrm{k_s}(i, w_1)$	4, 7/ =Elim; lemma
9		$q \in w_1$	1, 8/ *FO*con
10		$\ulcorner \bigwedge \mathrm{lt}(w_2) \to q \urcorner \in w_1$	9/ *w* closure
11		$p \in w_1$	5, 10/ =Elim

L-R, 2nd pt.

12	$\mathrm{Cn}(\mathrm{lt}(w_2)) \neq w_1$	
13	$\ulcorner \bigwedge \mathrm{lt}(w_2) \urcorner \notin w_1$	12/ lemma, w closure
14	$\ulcorner \neg \bigwedge \mathrm{lt}(w_2) \urcorner \in w_1$	13/ w completeness
15	$\ulcorner \bigwedge \mathrm{lt}(w_2) \to q \urcorner \in w_2$	14/ w closure
16	$p \in w_1$	5, 15/ $=$Elim
17	$p \in w_1$	6, 7–11, 12–16/ $\vee$ Elim
18	$p \in w_1$	3, 4–17/ $\exists$ Elim
19	$\mathrm{k}_s^{cond}(i, w_1) \subseteq w_1$	2–18/ $\forall$Intro

□

R-L.

1		$\mathrm{k}_s^{cond}(i,w)) \subseteq w$	
2	p	$p \in \mathrm{k}_s(i,w))$	
3		$\ulcorner \bigwedge \mathrm{lt}(w) \to p \urcorner \in \mathrm{k}_s^{cond}(i,w)$	2/ def. cond. code.
4		$\ulcorner \bigwedge \mathrm{lt}(w) \to p \urcorner \in w$	1, 3/ *FO* con.
5		$\ulcorner \bigwedge \mathrm{lt}(w) \urcorner \in w$	lemma
6		$p \in w$	4, 5/ w closure
7		$\mathrm{k}_s(i,w)) \subseteq w$	2–6/ $\forall$Intro

$\square$

It seems that generalized set theoretic approach opens up a number of interesting topics:

- historical,
- theoretical and philosophical,
- ethical.

Historical topic : Leibniz and the relation between the normative properties and requirements

Gottfried Wilhelm Leibniz.
Leibniz an Antoine Arnauld [Anfang November 1671].
Saemtliche Schriften Und Briefe. Zweite Reihe: Philosophischer Briefwechsel. Erster Band 1663-1685, 274–286, Akademie Verlag, 2006, Berlin

Quote

Licitum enim est, quod viro bono possibile est. Debitum sit, quod viro bono necessarium est.[a]

[a] That is permitted which a good man possibly is. That is obligatory which a good man necessary is.

Theoretical and philosophical topics

- Develop a typology of normative properties.
- Determine the deontic logic that describes the structure of the "property requirements".
- Use the "code approach" to explicate the notion of the "normativity of the mental" (e.g. Zangwill's thesis that it is of essence of the mental to be *subject* to the norms of rationality, not to conform to them).
    - If rationality is a normative source, what is its logical phenomenology (maximal compatibility?, formality?)?
    - What kind of property the rationality is?

Ethical topics

- Determination of the logical type of a "valid" code and of configurations of codes taking into account distinctions such as "code compatibility", "social consistency", "achievability", "logicality" etc.

Logical Framework for Normative Systems

Yasuo NAKAYAMA
Osaka University, Graduate School of Human Sciences
1-2 Yamada-oka, Suita, Osaka, Japan
nakayama@hus.osaka-u.ac.jp

In this paper, I propose a new logical framework that can be used to analyze normative phenomena in general. I call this framework a *Logic for Normative Systems* (LNS). I also demonstrate how to solve some paradoxes of Standard Deontic Logic (SDL). A characteristic of LNS is its dynamic behavior. LNS is flexible, hence it can be applied to describe complex normative problems including ethical problems.

1 Definition of Normative Systems

A normative system can be defined as follows, where $\vdash$ means the inference in the first-order logic:

(1a) Let T and OB be sets of sentences. A pair $\langle T, OB\rangle$ consisting of *propositional system* T and *obligation space* OB is called a *normative system* (NS).

(1b) A sentence p belongs to the *propositional context* of normative system $\langle T, OB\rangle$ if and only if (*iff*) $T \vdash p$.

(1c) A sentence p belongs to the *obligation context* of normative system $\langle T, OB\rangle$ (abbreviated as $\mathbf{O}_{\langle T,OB\rangle}p$) *iff* $T \cup OB \vdash p$ & $T \nvdash p$ & $T \cup OB \nvdash \perp$.

(1d) A sentence p belongs to the *prohibition context* of normative system $\langle T, OB\rangle$ (abbreviated as $\mathbf{F}_{\langle T,OB\rangle}p$) *iff* $\mathbf{O}_{\langle T,OB\rangle}\neg p$.

(1e) A sentence *p belongs to the permission context of normative system* $\langle T, OB\rangle$ (abbreviated as $\mathbf{P}_{\langle T,OB\rangle}p$) *iff* $T \cup OB \cup \{p\} \nvdash \perp$ & $T \nvdash p$.

(1f) A group G has (*normative*) *power* of doing act_1 in $\langle T, OB\rangle$ *iff*
$T \vdash \forall x(member(x, G) \rightarrow agent(x))$ &
$\mathbf{P}_{\langle T,OB\rangle}\forall x(member(x, G) \rightarrow do(x, act_1))$ &
$\mathbf{F}_{\langle T,OB\rangle}\forall x(\neg member(x, G) \rightarrow do(x, act_1))$. [1]

These definitions presuppose that we insert *what we believe to be true* into the propositional system and *what we believe ought to be done* into the obligation space.

[1] The notion of (legal) *power* plays an essential role in Hart (1961). This shows that this notion is inevitable for describing legal systems. According to definition (1f), a group G has power of doing act_1 *iff* all members of G and only members of G are allowed to perform act_1.

From these definitions immediately follow the following three fundamental characterizations for LNS.

(2a) If $\mathbf{O}_{\langle T,OB\rangle}p$, then $\mathbf{P}_{\langle T,OB\rangle}p$. [2]

(2b) If $\mathbf{F}_{\langle T,OB\rangle}p$, then not $\mathbf{P}_{\langle T,OB\rangle}p$.

(2c) The propositional context of a NS is independent of its obligation space, while the obligation context depends on the propositional system.

To see how LNS works, let us consider an example.

(3) You should not kill any human beings. Peter is a human being. So you should not kill Peter.

According to our presupposition mentioned before, the first sentence in (3) expresses the content of one component of OB_1 and the second sentence expresses the content of one component in T_1 of normative system $\langle T_1, OB_1\rangle$:

(4a) $\forall x \forall y(agent(x) \wedge human(y) \rightarrow \neg kill(x,y))$ is an element of OB_1.

(4b) $human(Peter)$ is an element of T_1.

From (1a), (1c), and (4a) immediately follow (4c) and (4d), where (4c) corresponds to the conclusion of (3).

(4c) $\mathbf{O}_{\langle T_1,OB_1\rangle}\forall x(agent(x) \rightarrow \neg kill(x, Peter))$.

(4d) For any agent A, $\mathbf{O}_{\langle T_1,OB_1\rangle}\neg kill(A, Peter)$.

This result can be summarized as follows:

(4e) If $\{human(Peter), agent(A)\} \subseteq T_1$ & $\{\forall x \forall y(agent(x) \wedge human(y) \rightarrow \neg kill(x,y))\} \subseteq OB_1$, then
$\mathbf{O}_{\langle T_1,OB_1\rangle}\forall x(agent(x) \rightarrow \neg kill(x, Peter))$ & $\mathbf{O}_{\langle T_1,OB_1\rangle}\neg kill(A, Peter)$ & $\mathbf{F}_{\langle T_1,OB_1\rangle}kill(A, Peter)$.

In this way, the informal reasoning in (3) can be formally justified within LNS.

2 Paradoxes of Deontic Logic and Their Solutions

In this section, I propose how to solve some paradoxes of SDL. First, let us consider Ross's paradox (Ross (1941), McNamara (2006) sec. 4.3, Åqvist (2002) sec. 6):

[2] This characterization corresponds to an axiom of SDL, namely $\mathbf{O}p \rightarrow \neg\mathbf{O}\neg p$. However, within LNS the iteration of normative sentences is not possible. This is no failure of LNS, because many normative systems in our ordinary life are expressible without iterative normative expressions. It is interesting that LNS fulfills the *principle of deontic contingency* required by Von Wright (1951). Because of (1c), we can easily prove that no tautology belongs to an obligation context.

(5a) It is obligatory that the letter is mailed.

(5b) It is obligatory that the letter is mailed or the letter is burned.

Within SDL, (5b) seems to follow from (5a), because $\mathbf{O}m \rightarrow \mathbf{O}(m \vee b)$ is a theorem of SDL, where $\mathbf{O}p$ means "It is obligatory that p". However, it seems rather odd to say that an obligation to mail the letter entails an obligation that can be fulfilled by burning the letter. Within LNS, a similar inference is valid:

(6*) If $\mathbf{O}_{\langle T,OB\rangle}p$ & $T \nvdash p \vee q$, then $\mathbf{O}_{\langle T,OB\rangle}(p \vee q)$.

The source of the paradoxical appearance of this example consists in ignoring the incompatibility among two types of actions. In this case, it will be appropriate to assume that mailing a letter is incompatible with burning it. This fact justifies to accept that $\forall x(letter(x) \rightarrow \neg(mailed(x) \wedge burned(x)))$ is a component of the propositional system of the given normative system $\langle T_2, OB_2\rangle$.

(6a) $mailed(l_1)$ is an element of OB_2. (From (5a))

(6b) $T_2 \nvdash mailed(l_1) \vee burned(l_1)$. (Observation)

(6c) $letter(l_1)$ and $\forall x(letter(x) \rightarrow \neg(mailed(x) \wedge burned(x)))$ are elements of T_2. (Observation)

From these conditions follows that $\neg burned(l_1)$ belongs to the obligation context of $\langle T_2, OB_2\rangle$ (see (6d)). Thus, it is forbidden to burn the letter. [3] [4]

(6d) If $\{letter(l_1), \forall x(letter(x) \rightarrow \neg(mailed(x) \wedge burned(x)))\} \subseteq T_2$ & $\{mailed(l_1)\} \subseteq OB_2$ & $T_2 \nvdash (mailed(l_1) \vee burned(l_1))$, then $T_2 \vdash \neg(mailed(l_1) \wedge burned(l_1))$ & $\mathbf{O}_{\langle T_2,OB_2\rangle}(mailed(l_1) \vee burned(l_1))$ & $\mathbf{O}_{\langle T_2,OB_2\rangle}\neg burned(l_1)$ & $\mathbf{F}_{\langle T_2,OB_2\rangle}burned(l_1)$.

The *Good Samaritan Paradox* pointed out by Prior (1958) can be solved in a similar way. Let us consider the following sentences:

(7a) It ought to be the case that John helps Smith who has been robbed.

(7b) John helps Smith who has been robbed *iff* John helps Smith and Smith has been robbed.

$\mathbf{O}(h \wedge r) \rightarrow \mathbf{O}(h)$ is a theorem of SDL. Thus, if we represent sentence (7a) as $\mathbf{O}(h \wedge r)$, (7c) follows from (7a) and (7b). However, (7c) seems hardly right.

(7c) It ought to be the case that Smith has been robbed.

Within LNS, (7a) can be analyzed as the combination of two conditions (8a) and (8c), where (8c) follows from (8a) and (8b).

[3]Belnap et al (2001) also discusses how to solve Ross's paradox within Stit logic (p. 84f). Stit logic can take future developments into consideration and they solve Ross's paradox using this property of Stit logic. They demonstrate that $[\alpha\ stit : p \vee q]$ does not follow from $[\alpha\ stit : p]$, where $[\alpha\ stit : p]$ is an abbreviation of [α sees to it that p].

[4]One of the reviewers pointed out my misinterpretation of Ross's paradox in the first draft of this paper. According to him, Ross's paradox is $\mathbf{O}m \rightarrow \mathbf{O}(m \vee b)$ itself. In this case, I will state that (6*) is not so bad, because realizing p remains still as an obligation after realizing q.

(8a) $agent(John)$ and $robbed(Smith)$ are elements of T_3.

(8b) $\forall x \forall y(agent(x) \wedge robbed(y) \rightarrow help(x, y))$ is an element of OB_3.

(8c) $\mathbf{O}_{\langle T_3, OB_3 \rangle}(robbed(Smith) \rightarrow help(John, Smith))$.

(8d) If $\{agent(John), robbed(Smith)\} \subseteq T_3$ &
$\{\forall x \forall y(agent(x) \wedge robbed(y) \rightarrow help(x, y))\} \subseteq OB_3$, then
$\mathbf{O}_{\langle T_3, OB_3 \rangle}(robbed(Smith) \rightarrow help(John, Smith))$ &
$\mathbf{O}_{\langle T_3, OB_3 \rangle} help(John, Smith)$.

From (8a) and (8c), follows that "John helps Smith" belongs to the obligation context of $\langle T_3, OB_3 \rangle$ ((8d)). This result means that *It ought to be the case that John helps Smith*, which is the result we sought.

Note that a normative system can express explicitly who the bearers of an obligation are, where we assume that they always try to fulfill their obligations, if they accept them:

(9a) Given a normative system $\langle T, OB \rangle$, "Teachers should prepare for their lectures" can be expressed as follows: $\mathbf{O}_{\langle T, OB \rangle}\ \forall x \forall y(agent(x) \wedge teacher(x) \wedge lecture\text{-}of(y, x) \rightarrow prepare(x, y))$.

(9b) Given a normative system $\langle T, OB \rangle$, "Students should study hard" can be expressed as follows: $\mathbf{O}_{\langle T, OB \rangle}\ \forall x(agent(x) \wedge student(x) \rightarrow study\text{-}hard(x))$.

3 Conflicts in Normative Systems

In normative system $\langle T, OB \rangle$, two kinds of contradictions are distinguished: the contradiction in propositional system T and that in $\langle T, OB \rangle$. T is *inconsistent as the propositional system of* $\langle T, OB \rangle$ *iff* T is inconsistent. However, $\langle T, OB \rangle$ is *inconsistent iff* $T \cup OB$ is inconsistent.

A characteristic of NSs is the property that any violation of the presupposed obligations produces inconsistency in NSs. Let us consider the following example:

(10a) It ought to be that Jones does go (to the assistance of his neighbors).

(10b) Jones doesn't go.

This situation can be represented as follows:

(10c) $\{agent(Jones), \neg go(Jones)\} \subseteq T_4$ & $\{go(Jones)\} \subseteq OB_4$.

This kind of violation of obligations could threaten the significance of a NS. If people perform their actions without respecting a given NS, that system can lose its significance. However, a small violation of a NS can sometimes be managed through taking the dynamic aspect of the reality into consideration (see section 4).

Next, let us consider a case of conflicts in an obligation space. Suppose that Tom is required to do act_1 as a member of group A. Furthermore, suppose that he is also required not to do act_1 as a member of group B. This situation produces inconsistency in the presupposed normative system $\langle T_5, OB_5 \rangle$:

(11) If $\{agent(Tom), member(Tom, A), member(Tom, B)\} \subseteq T_5$ & $\{\forall x(agent(x) \wedge member(x, A) \rightarrow do(x, act_1)),$ $\forall x(agent(x) \wedge member(x, B) \rightarrow \neg do(x, act_1))\} \subseteq OB_5$, then $\mathbf{O}_{\langle T_5, OB_5 \rangle} do(Tom, act_1)$ & $\mathbf{O}_{\langle T_5, OB_5 \rangle} \neg do(Tom, act_1)$.

One possible solution for Tom is to drop out from one of these groups. In that case, this decision reproduces a consistent NS. For example, if Tom drops out from group B, he need no more consider the prohibition of doing act_1.

4 Dynamic Aspects and Future Orientation of LNS

A normative system $\langle T, OB \rangle$ can function like a *conversational score* proposed by David Lewis (1979). A normative system can be updated to describe a development of a situation. Let us reconsider the case of a violating action described in (10a) and (10b). To describe the shift of time, we introduce here the *past-tense operator* P. Then, we can consider the shift of time and update normative system $\langle T_4, OB_4 \rangle$ and create $\langle T_{4up}, OB_4 \rangle$:

(12a) $\{agent(Jones), \neg go(Jones)\} \subseteq T_4$ & $\{go(Jones)\} \subseteq OB_4$.

(12b) $T_{4up} = (T_4 - \{\neg go(Jones)\}) \cup \{P(\neg go(Jones))\}$. (Information updated)

This example shows that a violated NS can sometimes automatically recover its consistency when its propositional system is updated.

Next, let us consider a case where an obligation becomes applicable through a change of the given situation.

(13a) We should help suffering neighbors.

(13b) Mary, who is a neighbor of John, was not suffering, but is now suffering.

(13c) So John should help Mary now.

Within LNS, the informal inference "(13c) follows from (13a) and (13b)" can be explicitly described as the inference of (13f) from (13d) and (13e):

(13d) $\{agent(John, neighbor(Mary, John)\} \subseteq T_6$ & $\{\forall x \forall y(agent(x) \wedge neighbor(y, x) \wedge suffering(y) \rightarrow help(x, y))\} \subseteq OB_6$.

(13e) $T_{6up} = T_6 \cup \{suffering(Mary)\}$. (Information updated)

(13f) $\mathbf{O}_{\langle T_{6up}, OB_6 \rangle} help(John, Mary)$.

Normative sentences are normally future oriented. Some problems can be solved by considering this property. Consider the following sentences: [5]

(14a) It should be the case that from now on any two countries have peaceful relations ($\{\forall t \forall x \forall y(t_{now} \leq t \wedge country(x) \wedge country(y) \wedge x \neq y \rightarrow peaceful(x, y, t))\} \subseteq OB_7$).

[5] This problem is pointed out by one of the reviewers.

(14b) A and B are countries ($\{country(A) \land country(B)\} \subseteq T_7$).

(14c1) A and B have now peaceful relations ($\{peaceful(A, B, t_{now})\} \subseteq T_7$).

(14c2) A and B have always peaceful relations ($\{\forall t\ peaceful(A, B, t)\} \subseteq T_7$)

(14d) It should be the case that from now on A and B have peaceful relations ($\mathbf{O}_{\langle T_7, OB_7 \rangle} \forall t(t_{now} \leq t \rightarrow peaceful(A, B, t))$).

Within LNS, (14d) follows from (14a), (14b), and (14c1), while (14d) does not follow from (14a), (14b), and (14c2). However, the second invalidity can be justified, because (14c2) expresses that the goal of (14d) has been already satisfied. [6]

5 Conclusions

It is well known that SDL has many theoretical difficulties (Åqvist (2002), McNamara (2006)). Recently, Stit logic was proposed and researchers have shown many interesting results (Belnap et al (2001), Horty (2001)). LNS is an alternative framework that can explicitly express both propositional and normative constraints. This property of explicitness makes LNS applicable to numerous classes of normative problems. For example, a legal system could be described as a normative system in the sense of LNS. The method developed in this paper can be modified to explain inferences among speech acts. However, this remains a future task. [7]

References

[1] Åqvist, L. (2002) "Deontic Logic", D. Gabbay and F. Guenthner (eds.) *Handbook of Philosophical Logic*, Vol. 8, Kluwer Academic Pub., pp. 147-264.
[2] Belnap, N., Perloff, M. and Xu, M. (2001) *Facing the Future: Agents and Choices in Our Indeterminist World*, Oxford University Press.
[3] Hart, H. L. A. (1961) *The Concept of Law*, Clarendon Press.
[4] Horty, J. F. (2001) *Agency and Deontic Logic*, Oxford University Press.
[5] Lewis, D. (1979) "Scorekeeping in a Language Game", *Journal of Philosophical Logic* 8, pp. 339-359.
[6] McNamara, P. (2006) "Deontic Logic", *Stanford Encyclopedia of Philosophy.*
[7] Prior, A. N. (1958) "Escapism: The Logical Basis of Ethics", In A.I. Melden (1958), *Essays in Moral Philosophy*, University of Washington Press, pp. 135-146.
[8] Ross, A. (1941) "Imperatives and Logic", *Theoria* 7, pp. 53-71.
[9] Von Wright, G. H. (1951) "Deontic Logic", *Mind* 60, pp. 1-15.

[6] As this section shows, LNS can deal with interactions between deontic states and temporal developments. Thus, LNS can be seen as a framework that satisfies the following requirements mentioned in Åqvist (2002): "for any serious perposes of application, the expressive resources of deontic languages must be enriched so as to include temporal and quantificational ones" (p. 150).

[7] I would like to thank two reviewers for many insightful comments.

Logical Framework for Normative Systems

Yasuo NAKAYAMA
Osaka University, Graduate School of Human Sciences
1-2 Yamada-oka, Suita, Osaka, Japan
nakayama@hus.osaka-u.ac.jp

Logical Framework for Normative Systems

- In this presentation, I propose a new logical framework that can be used to analyze normative phenomena in general.
- I call this framework a Logic for Normative Systems (LNS).
- I also demonstrate how to solve some paradoxes of Standard Deontic Logic (SDL).
- A characteristic of LNS is its dynamic behavior. LNS is flexible, hence it can be applied to describe complex normative problems including ethical problems.

Overview of my talk

1. Definition of Normative Systems
2. Paradoxes of Deontic Logic and Their Solutions
3. Conflicts in Normative Systems
4. Dynamic Aspects and Future Orientation of LNS
5. Conclusions

1. Definition of Normative Systems

Propositional Systems and Propositional Contexts

- A normative system can be defined as follows, where □ means the inference in the first-order logic:

(1a) Let *T* and *OB* be sets of sentences. A pair <*T*, *OB*> consisting of *propositional system* *T* and *obligation space* *OB* is called a *normative system* (NS).

(1b) A sentence p belongs to the *propositional context* of normative system <*T*, *OB*> iff *T* □ *p*.

1. Definition of Normative Systems

Obligation Contexts

(1c) A *sentence* p *belongs to the obligation context* of normative system $<T, OB>$ (abbreviated as $\mathbf{O}_{<T,\, OB>}\, p$) iff $T \cup OB \,\square\, p$ & not$(T \,\square\, p)$ & not$(T \cup OB \,\square\, \perp)$.

(1d) A sentence p belongs to the *prohibition context* of normative system $<T, OB>$ (abbreviated as $\mathbf{F}_{<T,\, OB>}\, p$) iff $\mathbf{O}_{<T,\, OB>} \neg p$.

(1e) A sentence p belongs to the *permission context* of normative system $<T, OB>$ (abbreviated as $\mathbf{P}_{<T,\, OB>}\, p$) iff not$(T \cup OB \cup \{\mathrm{p}\} \,\square\, \perp)$ & not$(T \,\square\, p)$.

1. Definition of Normative Systems

(Normative) Power

(1f) A group G has *(normative) power* of doing act_1 in $<T, OB>$ iff

$$T \ \Box\ \forall x\,(member(x,G) \rightarrow agent(x))\ \&$$

$$\mathbf{P}_{<T,\,OB>}\ \forall x\,(member(x,G) \rightarrow do(x,\, act_1))\ \&$$

$$\mathbf{F}_{<T,\,OB>}\ \forall x\,(\neg\, member(x,G) \rightarrow do(x,\, act_1))\ .$$

– The notion of (legal) *power* plays an essential role in Hart (1961). This shows that this notion is inevitable for describing legal systems. According to definition (1f), a group G has power of doing act_1 *iff* all members of G and only members of G are allowed to perform act_1.

1. Definition of Normative Systems

(Normative) Power: Example

- According to definition (1f), a group G has power of doing act_1 *iff* all members of G and only members of G are allowed to perform act_1.

- The chairman of a session has (normative) power of opening and closing the session. Only the chairman may close the session.

1. Definition of Normative Systems

Presuppositions of NSs

- These definitions presuppose that we insert *what we believe to be true* into the propositional system and *what we believe ought to be done* into the obligation space.
    - *T* consists of *what we believe to be true.*
    - *OB* consists of *what we believe ought to be done.*

1. Definition of Normative Systems

Obligation Implies Permission

- From these definitions immediately follow the following three fundamental characterizations for LNS.

(2a) If $\mathbf{O}_{<T,\ OB>}\, p$, then $\mathbf{P}_{<T,\ OB>}\, p$.

– This characterization corresponds to an axiom of SDL, namely $\mathbf{O}p \rightarrow \neg \mathbf{O} \neg p$. However, within LNS the iteration of normative sentences is not possible. This is no failure of LNS, because many normative systems in our ordinary life are expressible without iterative normative expressions. It is interesting that LNS fulfills the principle of deontic contingency required by Von Wright (1951). Because of (1c), we can easily prove that no tautology belongs to an obligation context.

1. Definition of Normative Systems

Three Fundamental Characterizations

(2a) If $\mathbf{O}_{<T,\, OB>}\, p$, then $\mathbf{P}_{<T,\, OB>}\, p$.

(2b) If $\mathbf{F}_{<T,\, OB>}\, p$, then not $\mathbf{P}_{<T,\, OB>}\, p$.

(2c) The propositional context of a NS is independent of its obligation space, while the obligation context depends on the propositional system.

1. Definition of Normative Systems

How LNS Works: An Example

- To see how LNS works, let us consider an example.

(3)

You should not kill any human beings.

Peter is a human being.

So you should not kill Peter.

1. Definition of Normative Systems

Analyzing Example (3)

- According to our presupposition mentioned before, the first sentence in (3) expresses the content of one component of OB_1 and the second sentence expresses the content of one component in T_1 of normative system $<T_1, OB_1>$:

(4a) $\forall x\, \forall y\, (agent(x) \wedge human(y) \rightarrow \neg\, kill(x,y))$ is an element of OB_1.

(4b) $human(Peter)$ is an element of T_1.

(3) You should not kill any human beings.
Peter is a human being.
So you should not kill Peter.

1. Definition of Normative Systems

Analyzing Example (3)

- From (1a), (1c), and (4a) immediately follow (4c) and (4d), where (4c) corresponds to the conclusion of (3).

(4c) $\mathbf{O}_{<T_1,\, OB_1>} \forall x\, (agent(x) \rightarrow \neg\, kill(x, Peter))$.

(4d) For any agent A, $\mathbf{O}_{<T_1,\, OB_1>} \neg\, kill(A, Peter)$.

(3) You should not kill any human beings.
Peter is a human being.
So you should not kill Peter.

1. Definition of Normative Systems

Analyzing Example (3)

(3) You should not kill any human beings.

Peter is a human being.

So you should not kill Peter.

(4e) If $\{human(Peter), agent(A)\} \subseteq T_1$ &

$\{\forall x\, \forall y\, (agent(x) \wedge human(y) \rightarrow \neg\, kill(x,y))\} \subseteq OB_1$,

then

$\mathbf{O}_{<T_1, OB_1>}\, \forall x\, (agent(x) \rightarrow \neg\, kill(x,Peter))$ &

$\mathbf{O}_{<T_1, OB_1>}\, \neg\, kill(A,Peter)$ & $\mathbf{F}_{<T_1, OB_1>}\, kill(A,Peter)$.

1. Definition of Normative Systems

Analyzing Example (3)

- This result can be summarized as follows:

(4e) If $\{human(Peter), agent(A)\} \subseteq T_1$ &

$\{\forall x\, \forall y\, (agent(x) \wedge human(y) \rightarrow \neg\, kill(x,y))\} \subseteq OB_1$,

then

$\mathbf{O}_{<T_1,\, OB_1>} \forall x\, (agent(x) \rightarrow \neg\, kill(x,Peter))$ &

$\mathbf{O}_{<T_1,\, OB_1>} \neg\, kill(A,Peter)$ & $\mathbf{F}_{<T_1,\, OB_1>}\, kill(A,Peter)$.

- In this way, the informal reasoning in (3) can be formally justified within LNS.

2. Paradoxes of Deontic Logic and Their Solutions

Ross's paradox

- In this section, I propose how to solve some paradoxes of SDL. First, let us consider Ross's paradox (Ross (1941), McNamara (2006) sec. 4.3, Aqvist (2002) sec. 6):

(5a) It is obligatory that the letter is mailed.

(5b) It is obligatory that the letter is mailed or the letter is burned.

2. Paradoxes of Deontic Logic and Their Solutions

Ross's paradox

- Within SDL, (5b) seems to follow from (5a), because $\mathbf{O}m \rightarrow \mathbf{O}(m \vee b)$ is a theorem of SDL, where $\mathbf{O}p$ means "It is obligatory that p".
- However, it seems rather odd to say that an obligation to mail the letter entails an obligation that can be fulfilled by burning the letter. Within LNS, a similar inference is valid:

(6*) If $\mathbf{O}_{<T,\ OB>}\ p$ & not$(T \Box\ p \vee q)$,

then $\mathbf{O}_{<T,\ OB>}(p \vee q)$.

2. Paradoxes of Deontic Logic and Their Solutions

Ross's paradox

- The source of the paradoxical appearance of this example consists in ignoring the incompatibility among two types of actions. In this case, it will be appropriate to assume that mailing a letter is incompatible with burning it. This fact justifies to accept that $\forall x\ (\mathit{letter}(x) \rightarrow \neg\ (\mathit{mailed}(x) \wedge \mathit{burned}(x)))$ is a component of the propositional system of the given normative system $<T_2, OB_2>$.

(6a) $\mathit{mailed}(l_1)$ is an element of OB_2. (From (5a))

(6b) $\mathrm{not}(T_2 \ \square\ \neg\ (\mathit{mailed}(l_1) \vee \mathit{burned}(l_1))$. (Observation)

(6c) $\mathit{letter}(l_1)$ and $\forall x\ (\mathit{letter}(x) \rightarrow \neg\ (\mathit{mailed}(x) \wedge \mathit{burned}(x)))$ are elements of T_2. (Observation)

(5a) It is obligatory that the letter is mailed.

2. Paradoxes of Deontic Logic and Their Solutions

Ross's paradox

(6d) If $\{letter(l_1),\ \forall x\,(letter(x) \rightarrow \neg\,(mailed(x) \wedge burned(x)))\} \subseteq T_2$ & $\{mailed(l_1)\} \subseteq OB_2$ &

not($T_2 \;\Box\; (mailed(l_1) \vee burned(l_1))$),

then $T_2 \;\Box\; \neg(mailed(l_1) \wedge burned(l_1))$ &

$\mathbf{O}_{<T_2,\, OB_2>}\ mailed(l_1)$ &

$\mathbf{O}_{<T_2,\, OB_2>}\,(mailed(l_1) \vee burned(l_1))$ &

$\mathbf{O}_{<T_2,\, OB_2>} \neg\, burned(l_1)$ & $\mathbf{F}_{<T_2,\, OB_2>}\, burned(l_1)$.

2. Paradoxes of Deontic Logic and Their Solutions

Ross's paradox

- From these conditions follows that $\neg$ *burned*(l_1) belongs to the obligation context of $<T_2, OB_2>$ (see (6d)). Thus, it is forbidden to burn the letter.
  - Belnap et al (2001) also discusses how to solve Ross's paradox within Stit logic (p. 84f). Stit logic can take future developments into consideration and they solve Ross's paradox using this property of Stit logic. They demonstrate that $[\alpha\ stit : p \vee q]$ does not follow from $[\alpha\ stit : p]$, where $[\alpha\ stit : p]$ is an abbreviation of $[\alpha$ sees to it that $p]$.
  - One of the reviewers pointed out my misinterpretation of Ross's paradox in the first draft of this paper. According to him, Ross's paradox is $\mathbf{O}m \rightarrow \mathbf{O}(m \vee b)$ itself. In this case, I will state that (6*) is not so bad, because realizing p remains still as an obligation after realizing q.

2. Paradoxes of Deontic Logic and Their Solutions

Good Samaritan Paradox

- The Good Samaritan Paradox pointed out by Prior (1958) can be solved in a similar way. Let us consider the following sentences:

(7a) It ought to be the case that John helps Smith who has been robbed.

(7b) John helps Smith who has been robbed *iff* John helps Smith and Smith has been robbed.

- $\mathbf{O}(h \wedge r) \rightarrow \mathbf{O}(r)$ is a theorem of SDL. Thus, if we represent sentence (7a) as $\mathbf{O}(h \wedge r)$, (7c) follows from (7a) and (7b). However, (7c) seems hardly right.

(7c) It ought to be the case that Smith has been robbed.

2. Paradoxes of Deontic Logic and Their Solutions

Good Samaritan Paradox

- Within LNS, (7a) can be analyzed as the combination of two conditions (8a) and (8c), where (8c) follows from (8a) and (8b).

(8a) *agent*(*John*) and *robbed*(*Smith*) are elements of T_3.

(8b) $\forall x\, \forall y\, (agent(x) \wedge robbed(y) \rightarrow help(x, y))$ is an element of OB_3.

(8c) $\mathbf{O}_{<T_3,\, OB_3>}(robbed(Smith) \rightarrow help(John, Smith))$.

(7a) It ought to be the case that John helps Smith who has been robbed.

2. Paradoxes of Deontic Logic and Their Solutions

Good Samaritan Paradox

(8d) If $\{agent(John), robbed(Smith)\} \subseteq T_3$ & $\{\forall x\, \forall y\, (agent(x) \wedge robbed(y) \rightarrow help(x, y))\} \subseteq OB_3$, then
$\mathbf{O}_{<T_3,\, OB_3>}(robbed(Smith) \rightarrow help(John, Smith))$ &
$\mathbf{O}_{<T_3,\, OB_3>}\, help(John, Smith)$.

- From (8a) and (8c), follows that "John helps Smith" belongs to the obligation context of $<T_3, OB_3>$ ((8d)). This result means that *It ought to be the case that John helps Smith*, which is the result we sought.

2. Paradoxes of Deontic Logic and Their Solutions

Good Samaritan Paradox

- Note that a normative system can express explicitly who the bearers of an obligation are, where we assume that they always try to fulfill their obligations, if they accept them:

(9a) Given a normative system $<T, OB>$, "Teachers should prepare for their lectures" can be expressed as follows:

$$\mathbf{O}_{<T,\, OB>} \forall x\, \forall y\, (agent(x) \wedge teacher(x) \wedge lecture\text{-}of(y, x)) \rightarrow prepare(x, y)).$$

(9b) Given a normative system $<T, OB>$, " Students should study hard" can be expressed as follows:

$$\mathbf{O}_{<T,\, OB>} \forall x\, (agent(x) \wedge student(x) \rightarrow study\text{-}hard(x)).$$

3. Conflicts in Normative Systems
Two Kinds of Contradictions

- In normative system $<T, OB>$, two kinds of contradictions are distinguished:

  the contradiction in propositional system T and that in $<T, OB>$.

- *T is inconsistent as the propositional system of $<T, OB>$ iff* T is inconsistent.
- However, $<T, OB>$ is *inconsistent iff* $T \cup OB$ is inconsistent.

3. Conflicts in Normative Systems

Violation of the Presupposed Obligations

- A characteristic of NSs is the property that any violation of the presupposed obligations produces inconsistency in NSs. Let us consider the following example:

(10a) It ought to be that Jones does go (to the assistance of his neighbors).

(10b) Jones doesn't go.

- This situation can be represented as follows:

(10c) $\{agent(Jones), \neg\, go(Jones)\} \subseteq T_4$ & $\{go(Jones)\} \subseteq OB_4$.

3. Conflicts in Normative Systems

Significance of a NS

- This kind of violation of obligations could threaten the significance of a NS. If people perform their actions without respecting a given NS, that system can lose its significance.
- However, a small violation of a NS can sometimes be managed through taking the dynamic aspect of the reality into consideration (see section 4).

3. Conflicts in Normative Systems

Conflicts in an Obligation Space

- Next, let us consider a case of conflicts in an obligation space. Suppose that Tom is required to do act_1 as a member of group A. Furthermore, suppose that he is also required not to do act_1 as a member of group B.
- This situation produces inconsistency in the presupposed normative system $<T_5, OB_5>$:

(11) If $\{agent(Tom), member(Tom,A), member(Tom,B)\} \subseteq T_5$ &

$\{\forall x\, (agent(x)\ member(x,A) \wedge\ \rightarrow do(x, act_1)),$

$\{\forall x\, (agent(x)\ member(x,B) \wedge\ \rightarrow \neg\, do(x, act_1))\} \subseteq OB_5$,

then $\mathbf{O}_{<T_5,\, OB_5>}\, do(Tom, act_1)$ & $\mathbf{O}_{<T_5,\, OB_5>} \neg\, do(Tom, act_1)$.

3. Conflicts in Normative Systems

One Possible Solution

- One possible solution for Tom is to drop out from one of these groups. In that case, this decision reproduces a consistent NS.
- For example, if Tom drops out from group *B*, he need no more consider the prohibition of doing act_1.

4. Dynamic Aspects and Future Orientation of LNS

Conversational Score

- A normative system $<T, OB>$ can function like a conversational score proposed by David Lewis (1979).
  - A normative system can be updated to describe a development of a situation.
- Let us reconsider the case of a violating action described in (10a) and (10b). To describe the shift of time, we introduce here the past-tense operator P. Then, we can consider the shift of time and update normative system $<T_4, OB_4>$ and create $<T_{4up}, OB_4>$:

(12a) $\{agent(Jones), \neg\, go(Jones)\} \subseteq T_4 \;\&\; \{go(Jones)\} \subseteq OB_4$.

(12b) $T_{4up} = (T_4 - \{\neg\, go(Jones)\}) \cup \{P(\neg\, go(Jones))\}$.
(Information updated)

(10a) It ought to be that Jones does go.

(10b) Jones doesn't go.

4. Dynamic Aspects and Future Orientation of LNS

Recovering the Consistency

(12a) $\{agent(Jones), \neg\, go(Jones)\} \subseteq T_4$ & $\{go(Jones)\} \subseteq OB_4$.

(12b) $T_{4up} = (T_4 - \{\neg\, go(Jones)\}) \cup \{P(\neg\, go(Jones))\}$. (Information updated)

- This example shows that a violated NS can sometimes automatically recover its consistency when its propositional system is updated.

4. Dynamic Aspects and Future Orientation of LNS

Change of a Situation

- Next, let us consider a case where an obligation becomes applicable through a change of the given situation.

(13a) We should help suffering neighbors.

(13b) Mary, who is a neighbor of John, was not suffering, but is now suffering.

(13c) So John should help Mary now.

4. Dynamic Aspects and Future Orientation of LNS

Explanation of an Informal Inference

- Within LNS, the informal inference " (13c) follows from (13a) and (13b)" can be explicitly described as the inference of (13f) from (13d) and (13e):

(13d) $\{agent(John), neighbor(Mary, John)\} \subseteq T_6$ &

T_6

$\{\forall x\, \forall y\, (agent(x) \wedge neighbor(y, x) \wedge suffering(y) \rightarrow help(x, y))\} \subseteq OB_6$.

(13e) $T_{6up} = T_6 \cup \{suffering(Mary)\}$. (Information updated)

(13f) $\mathbf{O}_{<T_{6up},\, OB_6>} help(John, Mary)$.

4. Dynamic Aspects and Future Orientation of LNS

Future Orientation of LNS

- Normative sentences are normally future oriented. Some problems can be solved by considering this property. Consider the following sentences:

(14a) It should be the case that from now on any two countries have peaceful relations.

$$\{\forall t\ \forall x\ \forall y\ (t_{now} \leq t \wedge country(x) \wedge country(y) \wedge x \neq y \rightarrow peaceful(x, y, t))\} \subseteq OB_7$$

(14b) A and B are countries.

$$\{country(A) \wedge country(B)\} \subseteq T_7$$

Future Orientation of LNS

(14c1) A and B have now peaceful relations.

$$\{peaceful(A, B, t_{now})\} \subseteq T_7$$

(14c2) A and B have always peaceful relations.

$$\{\forall t\ peaceful(A, B, t)\} \subseteq T_7$$

(14d) It should be the case that from now on A and B have peaceful relations.

$$\mathbf{O}_{<T_7,\, OB_7>} \forall t\ (t_{now} \leq t \rightarrow peaceful(A, B, t))$$

4. Dynamic Aspects and Future Orientation of LNS

Future Orientation of LNS

- Within LNS, (14d) follows from (14a), (14b), and (14c1), while (14d) does not follow from (14a), (14b), and (14c2).
- However, the second invalidity can be justified, because (14c2) expresses that the goal of (14d) has been already satisfied.
  - As this section shows, LNS can deal with interactions between deontic states and temporal developments. Thus, LNS can be seen as a framework that satisfies the following requirements mentioned in Aqvist (2002): " for any serious prepuces of application, the expressive resources of deontic languages must be enriched so as to include temporal and quantificational ones" (p. 150).

5. Conclusions
Theoretical Difficulties of SDL

- It is well known that SDL has many theoretical difficulties (Aqvist (2002), McNamara (2006)).
- Recently, Stit logic was proposed and researchers have shown many interesting results (Belnap et al (2001), Horty (2001)).

5. Conclusions

Explicitness of LNS

- LNS is an alternative framework that can explicitly express both propositional and normative constraints.
- This property of explicitness makes LNS applicable to numerous classes of normative problems.
- For example, a legal system could be described as a normative system in the sense of LNS.
- The method developed in this paper can be modified to explain inferences among speech acts. However, this remains a future task.

5. Conclusions

How to Express Speech Acts

- Instead of obligations, we may use imperatives or intentions.

$$\mathbf{O}_{<T,\ OB>}$$

$$\mathbf{COMMAND}_{<T,\ IMP>}$$

$$\mathbf{INTEND}_{<T,\ INT>}$$

- As these cases show, the method presented here is quite general.

References

[1] Aqvist, L. (2002) "Deontic Logic", D. Gabbay and F. Guenthner (eds.) Handbook of Philosophical Logic, Vol. 8, Kluwer Academic Pub., pp. 147-264.

[2] Belnap, N., Perlo, M. and Xu, M. (2001*) Facing the Future: Agents and Choices in Our Indeterminist World*, Oxford University Press.

[3] Hart, H. L. A. (1961) *The Concept of Law*, Clarendon Press.

[4] Horty, J. F. (2001*) Agency and Deontic Logic*, Oxford University Press.

[5] Lewis, D. (1979) "Scorekeeping in a Language Game*", Journal of Philosophical Logic* 8, pp. 339-359.

[6] McNamara, P. (2006) "Deontic Logic"*, Stanford Encyclopedia of Philo*sophy.

[7] Prior, A. N. (1958) "Escapism: The Logical Basis of Ethics", In A.I. Melden (1958), *Essays in Moral Philosophy*, University of Washington Press, pp. 135-146.

[8] Ross, A. (1941) "Imperatives and Logic", *Theoria* 7, pp. 53-71.

[9] Von Wright, G. H. (1951) "Deontic Logic", *Mind* 60, pp. 1-15.

Responsibility in Games

Matthew Braham
Faculty of Philosophy
University of Groningen
The Netherlands
m.braham@rug.nl

Martin van Hees
Faculty of Philosophy
University of Groningen
The Netherlands
Martin.van.Hees@rug.nl

January 10, 2010

When outcomes result from the joint actions of two or more people it is generally believed that we face major difficulties in ascribing responsibility. Thompson (1980) has christened it the 'many hands problem'. One part of the problem is determining the causal contributions given that individual efforts may be like strands in a rope: together each strand makes up the rope but each particular strand may be dispensable. In joint actions, the role of each individual in bringing about an outcome appears to be lost in a complex process. Another part of the problem is that even when causal contributions can be determined, there may not actually be anything wrong with each of the actions per se. Determining 'whose hands' will not necessarily pick out 'whose hands were dirty', i.e. the set of individuals who should – as is generally the case – be punished, held liable, or subjected to moral criticism.

The problems are germane and pervasive. Feinberg (1968) discusses the example of the 'Jesse James train robbery' in which an armed man holds up a car full of passengers. If the passengers had risen up as one and rushed the robber, one or two of them would have perhaps been shot, but collectively they would have overwhelmed him and saved their own and other's property. Feinberg asks to what extent are any of the passengers responsible for the loss of their property given that none alone could have prevented Jesse James walking off with it? Copp (2006) and Pettit (2007) have discussed a quirk of collective decision-making known as the 'discursive paradox' in which members of a committee each have their reasons to reject a particular proposal put before them but the proposal passes nevertheless given the way in which the decision procedure works. In what way can the committee members be held responsible for the outcome?

A more down to earth example concerns the group of managers and engineers at MacDonnell-Douglas who knew of the design faults in early DC-10s that led to these planes dropping out of the sky in the 1970s but nevertheless allowed these planes to, fatally, go into service. As it turned out no single individual was declared as having been directly decisive for the harm that occurred. Then there are the My Lai or Sebrenica massacres; the infamous murder of Kitty Genovese in New York in which onlookers watched her slow death; and the run of recent bank collapses that have greatly damaged the international financial system. The list is endless (for a catalogue of further cases see Bovens, 1998).

One solution to the problem has been the introduction of the concept of *collective responsibility*. The idea here is to argue that the outcomes are the result of a form of 'collective agency', which supervenes on individual members of a group. The task has been to show first which types of groups can be treated as 'moral agents' and then assign any personal responsibility on the basis of voluntary membership of the collective agent. Proponents of this idea of collective responsibility are French (1984) and Pettit (2007). The concept of collective responsibility is, however, not undisputed. The hypostasization of groups is anything but straightforward. It raises a host of metaphysical quandaries about the ontological status of agency. Normatively speaking there are equally difficult problems, one of which is that is that the 'membership' or 'shared attitudes' criteria has the highly unpalatable consequence that it can result in holding people responsible for states of affairs for which they played no part in bring about.

The purpose of this paper is to tackle the problem anew and demonstrate that it is in fact possible in principle to ascribe moral responsibility to individual agents in complex joint activities. To do so, we assume an agent can be held responsible for the realization of a state of affairs A if the following criteria are satisfied:

Agency Condition — The person is an autonomous, intentional, and planning agent who is capable of distinguishing right and wrong and good and bad.

Causal Relevancy Condition — There should be a causal relation between the action of the agent and the resultant state of affairs.

Avoidance Opportunity Condition — The agent should have had a reasonable opportunity to have done otherwise.

Building on previous work (Braham and Holler, 2008; Braham, 2008; Braham and van Hees, 2009), we show how these conditions can be recast in a very natural habitat: a game theoretic framework. We then define the components of what we call a 'responsibility game' and examine how different allocations of responsibility can be associated with different classes of responsibility games. Subsequently we show that, from a theoretical viewpoint, the 'many hands problem' is not as severe as it may at first appear.

References

Bovens, M. (1998). *The Quest for Responsibility*. Cambridge: Cambridge University Press.

Braham, M. (2008). Social Power and Social Causation: Towards a Formal Synthesis. In Braham, M. and Steffen, F. (eds), *Power, Freedom, and Voting*. Heidelberg: Springer.

Braham, M. and Hees, M. van (2009). Degrees of Causation. *Erkenntnis* 73: 323–344.

Braham, M. and Holler, M. J. (2008). Distributing Causal Responsibility in Collectivities. In Boylan, T. and Gekker, R. (eds), *Economics, Rational Choice, and Normative Philosophy*. London: Routledge.

Copp, D. (2006). On the Agency of Certain Collective Entities: An Argument from "Normative Autonomy". *Midwest Studies in Philosophy* 30: 194–221.

Feinberg, J. (1968). Collective Responsibility. *Journal of Philosophy* 65: 674–688.

French, P. A. (1984). *Collective and Corporate Responsibility*. New York: Columbia University Press.

Pettit, P. (2007). Responsibility Incorporated. *Ethics* 117: 171–201.

Thompson, D. F. (1980). Moral Responsibility of Public Officials: The Problem of Many Hands. *American Political Science Review* 74: 905–916.

Is Intention *sufficient* to explicate Collective Agency?

Biswanath Swain
Department of Humanities and Social Sciences
Indian Institute of Technology Kanpur, India

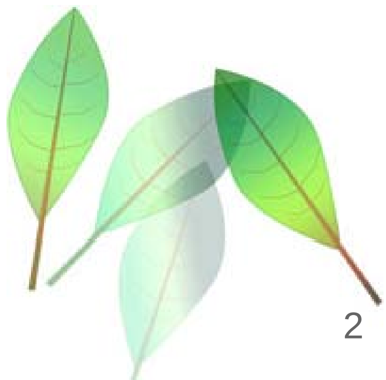
“…What is left over, if I subtract the fact that my arm goes up from the fact that I raised my arm?”

-:**Ludwig Wittgenstein**

Philosophical Investigations

Translated by G. E. M. Anscombe

1953, 621, P – 161



The Pervasive Nature of Change

Except the change itself, which is not subject to change, every other things in this world is subject to change. At a particular time, it comes into being and at another particular time, it goes out of existence.

Two types of changes across this world

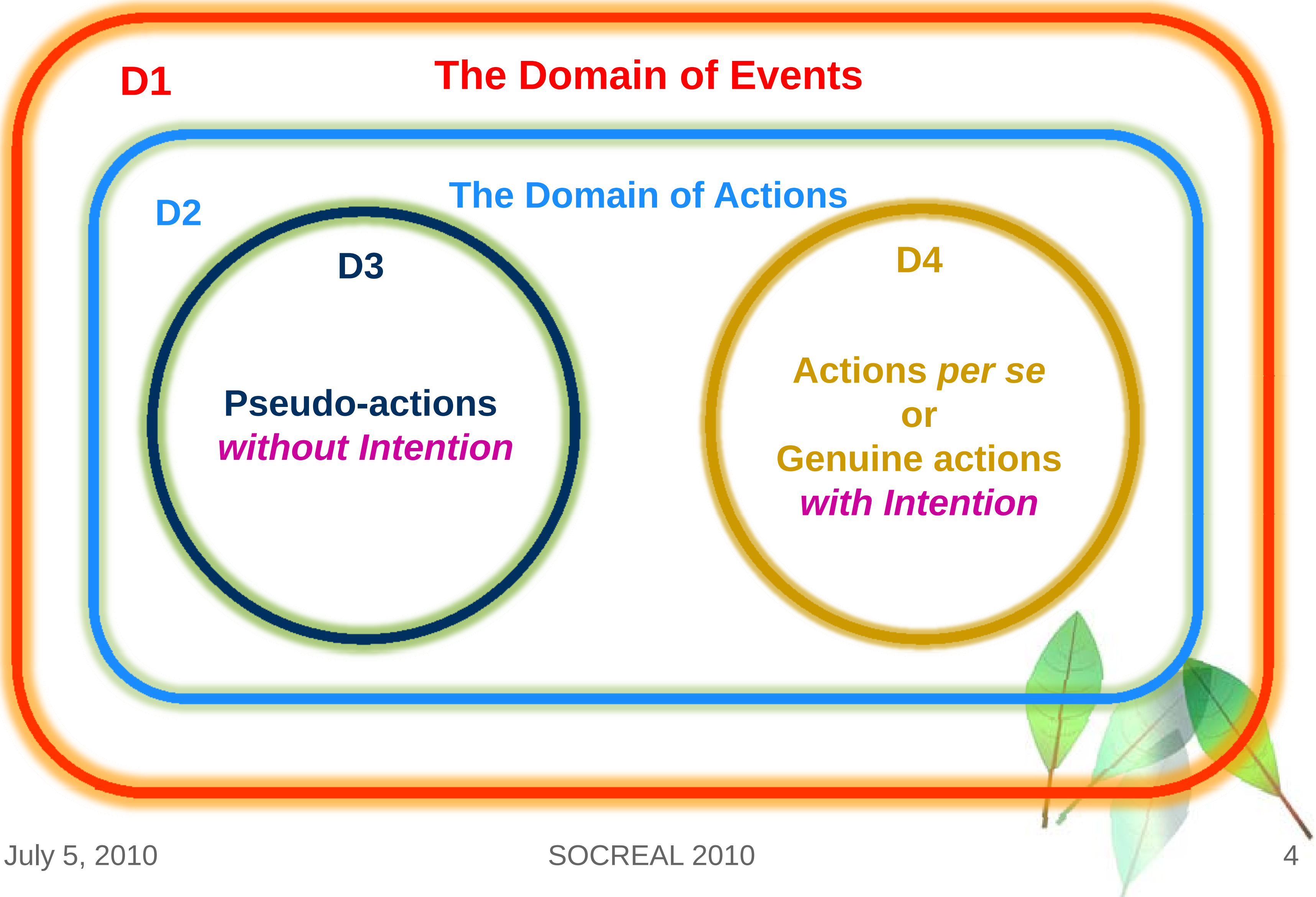
- Changes that are *brought about* by *Human Being*

  i.e., a uncoerced, unconditional, purposive behavior being guided by a free and a self-motivational force within a human being

- Changes that are *caused* by *Natural Forces*

  i.e., a purely conditional happening being guided by the *Laws of Nature*, and thus available for *causal explanation*.

$(\alpha \wedge \sim\alpha)$ The World of Changes $(\sim\alpha \wedge \alpha)$



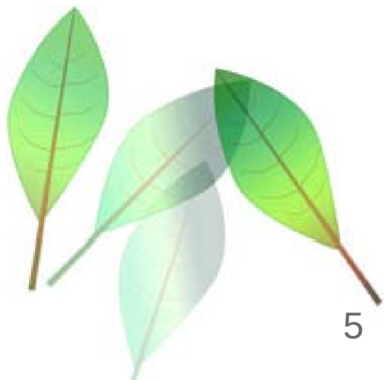
Intention:
The Differentiating Mark

- That bisects the world of changes into two parts.

- That differentiates Actions from the Events.

- That keeps apart the mere bodily behaviors from the behaviors that makes an action possible.

- That draws a dividing line between the answerability and unanswerability.



Continued…

- That keeps aside the agent and all the agential forms from all the form of non-agents and pseudo-agents.

- That produces more often moral or non-moral actions than *amoral* actions.

- Eventually, that shapes up an individual into a just individual or into a righteous person.

- Sometimes, Intention gets instantiated by some mental states like desire, pro-attitude, beliefs, predispositions etc.

Collective Intention behind
Collective Agency

- **We** are **social beings** by the very existence.

- Where we witness the **gregarious instinct** outrightly.

- Performance of **Inter**-dependent actions, apart from doing his/her **In**-dependent actions.

- Involvement in **Inter**-dependent actions by performing **In**-dependent actions.

- By **sharing a public place** with others under a **common goal**.

Continued…

- Keeping individuality and individual goal(s) aside, plunging into **commonality**.

- Becoming a part of that **shared intention** for the betterment of that group where he/she hails from.

- So the shift from **I-Intention** to **We-intention** sprouts the idea of **collective agency**, where all the I-Intentions are merged into a whole/ single.

- This collective intentional goal-directedness is a **necessary condition** for explicating collective agency.

Intention and Collective Action

- .Distinction between Intending-to and Intending-that.

- Individual autonomy vs. Collective autonomy.

- Distributive conception of collective intention.

- Instance of *seemingly* collective action where the collective intention is missing.

- Instance of a group action which is guided by an individual intention – cannot be a collective action as there is no collective intention.

Features of Collective Action

To evade the concept of collective intention from misuse, M. Bratman comes up with four features.

- Members of the collective each intend that they engage in the intended action.

- They each intend to do their part in this engagement.

- They each believe that the other members intend to do their part as well.

- Because of believing this, they each intend to do their part of the collective action.

Continued…

Stressing more on shared character of the members, he proposes *three* *basic features* of shared character which is indispensable for collective intention.

❶ Mutual responsiveness to each other's intention and action.

❷ Commitment to the joint activity.

❸ Commitment to mutual support in performing he joint activity.

Joint Commitment

Proposed by Margaret Gilbert (1989).

❖ Commitment is a **matter of will** to commit to oneself or to others.

❖ **Shift** from Individual commitment to Joint commitment.

❖ Plurality of individual wills intermingle to form **a single will**.

❖ Later, becomes a **single commitment** made by all of them, and are committed for the same intended goal.

❖ Members are **jointly committed** to one another.

❖ Each one is **supposedly obliged** to do his/her part of the action to make it the case that **they all act as a body**.

❖ It would necessarily invite **obligatory** or **normative considerations**.

❖ **Mutual reliance** and **expectation** would ensue from this commitment.

Continued…

Is normativity intrinsic or a matter of consequence?

- Account of Bratman, Gilbert, Davidson and Dennett.
- For Bratman, it is a matter of consequence, so it is contingent.
- For Gilbert, it cannot be contingent, rather it is intrinsic to collective action.
- For Davidson, normativity or rationality is constitutive of the idea of the intentional or mental.
- For Dennett, rationality in intentional behavior is interpreted in a epistemic or instrumental sense.

Internal Evaluative Mechanism

- A phenomenological shift of stance from *individual evaluative mechanism* to *collective evaluative mechanism*.

- This collective evaluative mechanism is not conditioned or constrained by the external forces, rather by the *ideals of rationality* within that group.

- *Joint evaluative perspective* is to be framed judiciously by each member of the group.

- Try to *evaluate* the significance of the goal of their collective intention in the light of this mechanism.

Continued…

- Take a course of *appropriate responsive behaviors* in regard to that envisaged end.

- *My* sense of concern and care for the desired end is for *our* sake, and is for my sake as well, because I feel myself *inalienably bound up* with the collective *we-identity*.

- Become *rationally beholden*, and also *answerable* for the result of *that* collective action jointly.

- Least fear for the unwanted outcome, and that makes all the members *fully devoted* towards that joint pursuit.

Agential Ability

- There should be a reasonable match between the content of the intention and agential ability to execute an action plan.
- Intention and possibility are supposed to go side by side.
- The collective intention must be in line with the possibility of executing the intention.
- There should not be a mismatch between the intention and the capacity.
- The collective agent i.e., we should realize the magnitude of the goal in relation to the standard agential limits.

Continued…

- We should be rationally self-confident enough to execute an action plan.
- That means rationally we should fathom our agential ability, before stepping into a course of action to realize the intended goal.
- If the match between the agential ability and the intention stands to suffer due to ailed foresight, the grave failure will be obvious.
- It is for sure that, intention cannot be irrational. It can only be rational.
- So, there may be the possibility of mismatch between desire and agential capacity. Only desire can be rational or irrational.

Key References

✓ Bratman, M. E. (1999). *Faces of Intention*. Cambridge, MA: Cambridge University Press.

✓Gilbert, M. (1989). *On Social Facts*. New Jersey: Princeton University Press.

✓ -- (2005). “Rationality in Collective Action.” *Philosophy of the Social Sciences*, Vol. 36, No. 1, pp. 3 – 17.

✓ Helm, B. W. (2008). “Plural Agents.” *Nous*, 42, (1), pp. 94 – 114.

✓Tuomela, R. (1977). *Human Action and its Explanation: A Study on the Philosophical Foundations of Psychology*, Dordrecht, Boston: D. Reidel Publishing Company.

✓ Wittegenstein, L. (1953). *Philosophical Investigations*, translated by Anscombe, G. E. M. Englewood Cliffs/N.J.: Prentice Hall.

Acknowledgment

- Organizers of SOCREAL 2010 for selecting me, and providing the travel grant.
- Indian Institute of Technology Kanpur, India for letting me be.
- Anonymous reviewers of my manuscript for their incisive comments and valuable suggestions.
- My thesis supervisors (Prof. C. A. Tomy, Prof. Bijoy H. Boruah and Prof. Munmun Jha) for their stimulating ideas and the words of encouragement, and Akhaya for his timely help and suggestion.


Thank you

Interpretation of Action and Sociality of Action

SOCREAL 2010

Fujimoto Ryoji and Choi Chang-Bong

Interpretation

- We use the word 'interpretation' (of action) as to make a choice from possible intentions of an agent which explain his action.
- If the intention which an interpreter decides as the interpretation of an action has the same type as the intention which *actually* motivates it, we say that the interpretation is *correct*.

Case A

- Suppose that Mike wants to take a taxi. Then, he intends to stop it, and so he raises his hand toward a taxi. Let John be the driver of the taxi. He interprets Mike's action *correctly*, so he stops his car, and Mike successfully takes a taxi.

- In this case, everything goes well.

Case B

- Suppose that Mike wants to take some exercise. At first, he intends to work upper body, and he raises his hand. At the moment John happens to pass by, and he watches Mike's raising hand. John interprets Mike's action *wrongly*, and He stops his car. But Mike, of course, does not get into the car.

- John may complain about Mike's action.

Case C

- Let us revise Case A with following modification. Suppose that John has a strong desire to go home. So, John is aware of Mike's action, whereas John does not stop his car.

- Mike may complain about John's action.

Interpretation of raising hand (1)

- Mike's intention to stop a taxi [I1]
- Mike's intention to take some exercise [I2]

- With appropriate mental states of Mike, both [I1] and [I2] why he raises hand.
- John can believe that
  - Mike raise hand because of [I1]
  - Mike does so because of [I2]

Interpretation of raising hand (2)

- Why does john interpret raising hand as the expression of the intention to stop a taxi?
- In a case that involves a pedestrian and a taxi, *the possibility* of raising hand's being the expression of the intention to stop a taxi is stronger than *possibilities* of the action's being the expression of other intentions.

Restriction on interpretation

(RI)

(Ⅰ) [Y] interprets [A] as the representation of [I]

only if

(Ⅱ) [Y] *can rationally believe* that [X] makes [A] because of [I]

[A]: action [X]: agent of [A] [Y]: interpreter
[I]: intention

1. Mike [P1]
2. John [P2]
3. Mike's intention to stop a taxi [I1]
4. Mike's intention to take some exercise [I2]
5. Raising hand [E1]
6. Interpretation of [E1] [E2]
7. Stopping a taxi by [P1] [E3]

- Both I1 and I2 meet condition (Ⅱ) in (RI)
- There may be other intentions which meet (Ⅱ). But, here, suppose that they do not occur to [P2].

The Direction of the complaint in the Case B

- In the Case B,

John([P2]) may complain about [E1] of Mike([P1]).

- The Case B contains the intentional action (= Mike's raising hand [E1]) involving the unintentional result (= to stop the taxi [E3]).
- It occurs because of John's *mis*interpretation.
- But it is not John but Mike who ought to excuse himself for his action.

Predictability of interpretation

- If we admit that our actions are founded on *predictabilities of interpretations* (PI), the problem of direction may disappear.
- For, if Mike[P1] takes the possibility of misinterpretation into consideration, he did not raise hand in the situation of the Case B. For, in such a situation, it is strongly possible that a taxi driver interprets raising hand[E1] as the expression of [I1].
- So, we can say that Mike also makes mistake. And, if the possibility of misinterpretation is strong, we would say that one who is wrong is Mike.

From intention to action (1)

(A) Mike believes that

(1) the taxi driver of the taxi (which gets closer to Mike by chance)

(a) interprets raising hand as the expression of the intention to stop a taxi, *and*

(b) is willing to stop a taxi, when he is aware of such an expression.

From intention to action (2)

(B) Mike believes that

(2) all taxi drivers

(c) interprets raising hand as the expression of the intention to stop a taxi, *and*

(d) is willing to stop his car, when he is aware of such an expression.

From intention to action (3)

(C) Mike believes that

(3) most taxi drivers

(e) interpret raising hand as the representation of the intention to stop a taxi, *and*

(f) are willing to stop his car, when they are aware of such an expression.

From intention to action (4)

- It may work well. Here, one may point out that (C) does not entail (A) unless we add some incidental presuppositions. But even so, it is plausible that Mike raises his hand as the expression of his intention. For, if (C) is true, Mike will still act so as a *reasonable bet*.

From intention to action (5)

- However, (C) still has one drawback. Let us revise Mike's case with following modification. Suppose that John is aware of Mike's action, whereas John does not stop his car. In this case, Mike may complain about John's action. And if he does, Mike's complaint is directed to John.

From intention to action (6)

- Even if actually (3) holds and Mike believes so, John is just an exceptional man to (2). So, Mike's complaint must be something like "How bad luck it is today!" which does not have the directedness. Thus, we can say that (C) explains Mike's action, but it does not explain his complaint.

From intention to action (7)

(D) Mike expects that

(4) all taxi drivers

(g) interprets raising hand as the expression of the intention to stop a taxi, *and*

(h) is willing to stop his car, when he is aware of such an expression.

Case Study [Kidnap]

A kidnap occurred in a big city. This city is abounding in crime for several years and local residents are conscious of crime. But, kidnappers demanded ransom and parents paid ransom for their son. After all, kidnappers carried out their kidnapping successfully. In this crime kidnappers used a person (P) as a messenger. They asked P to send a letter in which they demanded ransom and indicated the locations of transactions. P contributed to kidnapping-for-ransom by sending the letter.

Case [1] child

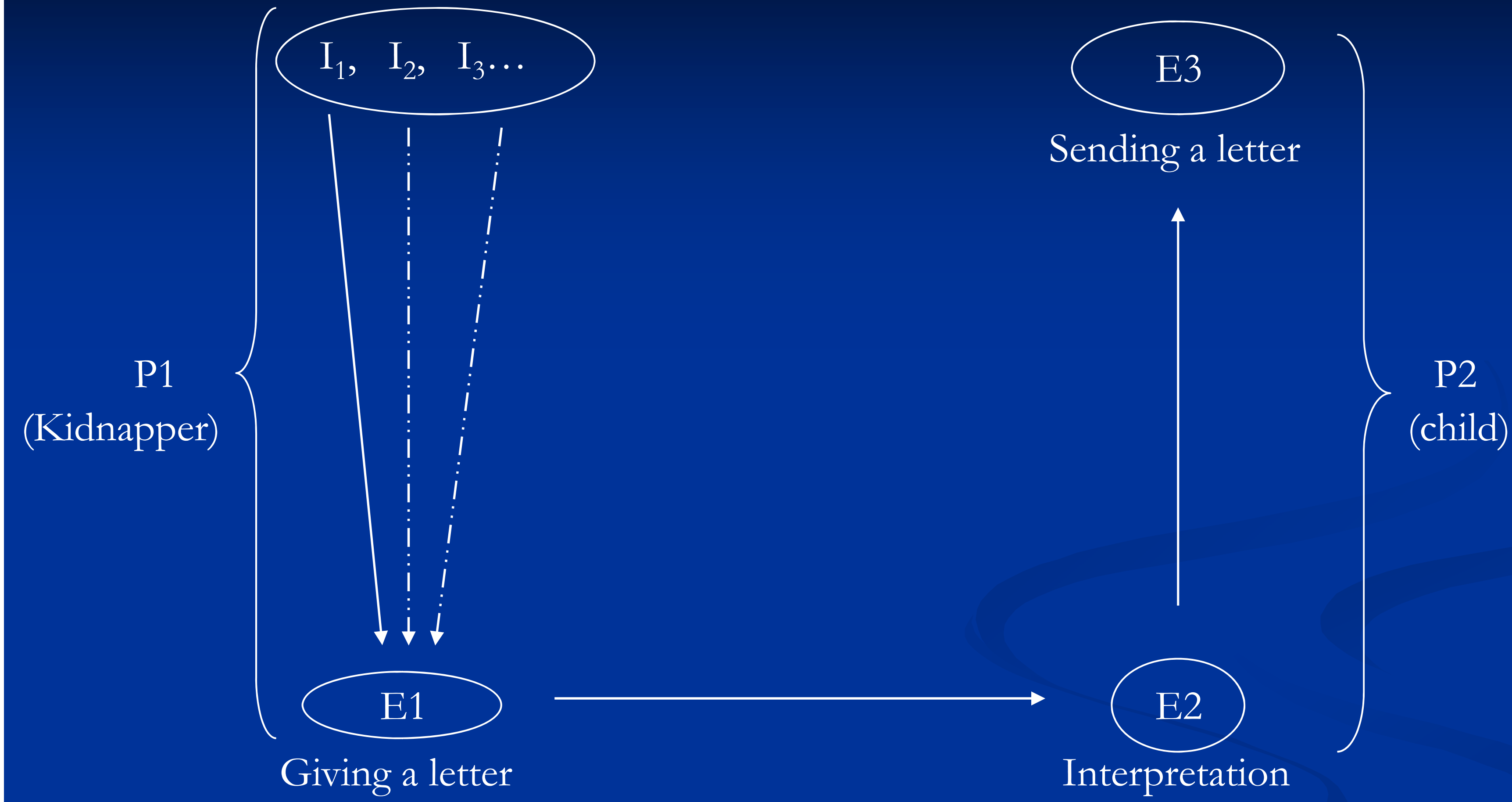
A kidnap occurred in a big city. This city is abounding in crime for several years and local residents are conscious of crime. But, kidnappers demanded ransom and parents paid ransom for their son. After all, kidnappers carried out their kidnapping successfully. In this crime kidnappers used *a child* as a messenger. They asked *the child* to send a letter, in which they demanded ransom and indicated the locations of transactions. *The child* contributed to kidnapping-for-ransom by sending the letter. But, *the child* didn't know contents of the letter. Because *the child* was asked to send a letter when he happened to pass a kidnapper by.

Case [2] adult

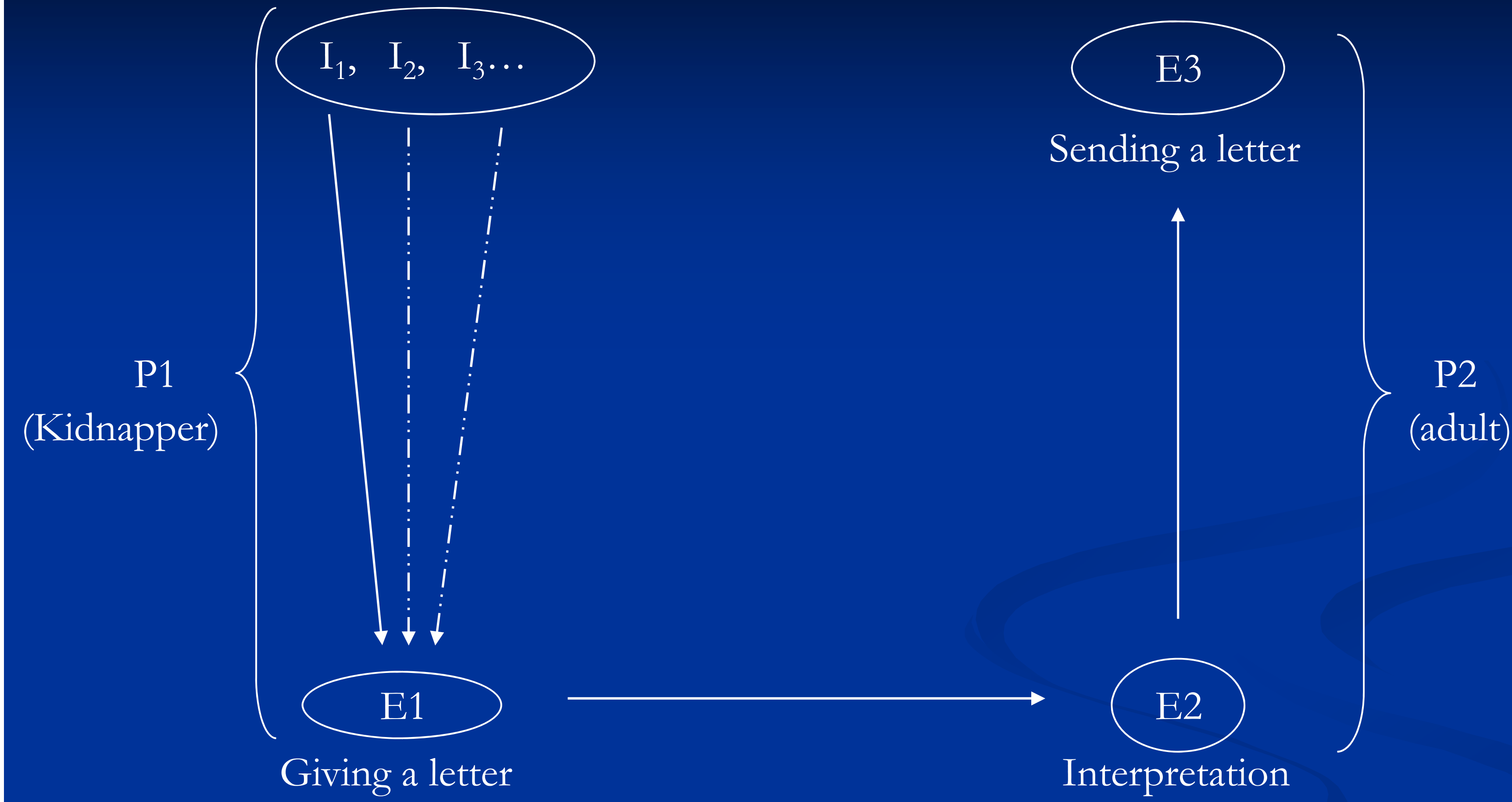
A kidnap occurred in a big city. This city is abounding in crime for several years and local residents are conscious of crime. But, kidnappers demanded ransom and parents paid ransom for their son. After all, kidnappers carried out their kidnapping successfully. In this crime kidnappers used *a adult* as a messenger. They asked *the adult* to send a letter, in which they demanded ransom and indicated the locations of transactions. *The adult* contributed to kidnapping-for-ransom by sending the letter. But, *the adult* didn't know contents of the letter. Because *the adult* was asked to send a letter when he happened to pass a kidnapper by.

Case [3] member of kidnappers

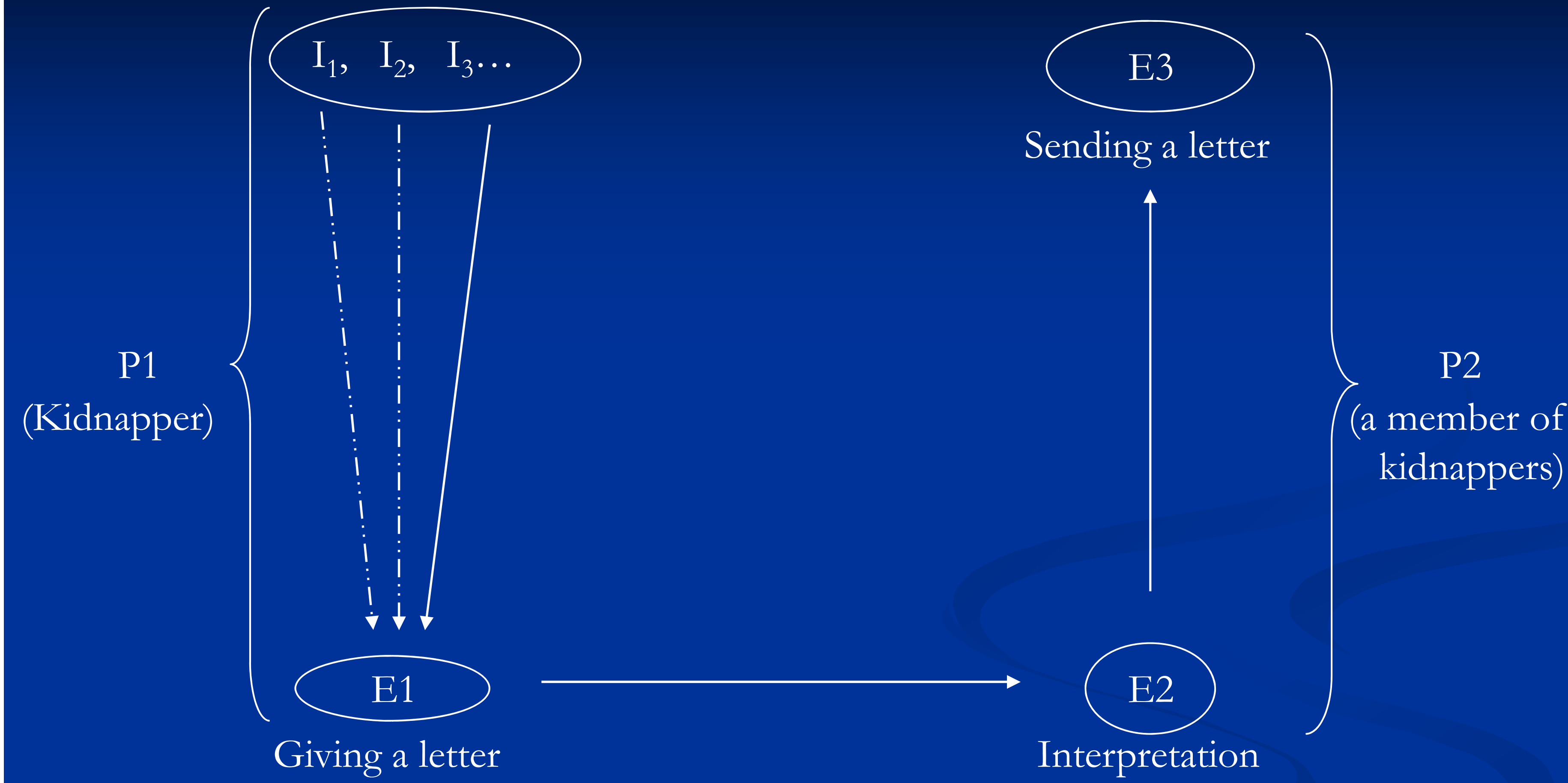
A kidnap occurred in a big city. This city is abounding in crime for several years and local residents are conscious of crime. But, kidnappers demanded ransom and parents paid ransom for their son. After all, kidnappers carried out their kidnapping successfully. In this crime kidnappers used *a member of kidnappers* as a messenger. They asked *the member* to send a letter, in which they demanded ransom and indicated the locations of transactions. *The member* contributed to kidnapping-for-ransom by sending the letter. As a matter of course, *the member* did know contents of the letter.



The child case



The adult case



The member case

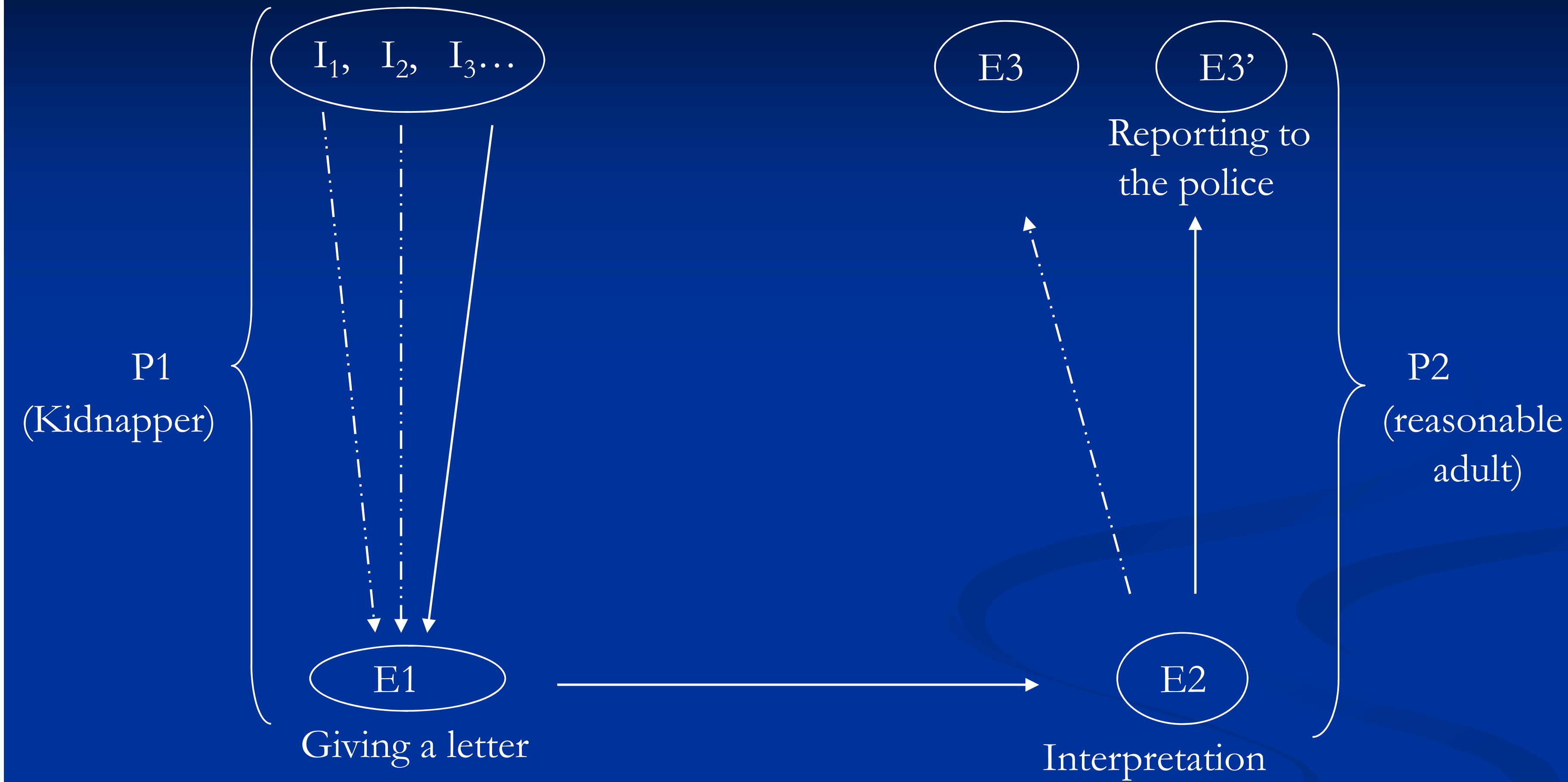
case	child	adult	a member of kidnappers
Causal connection	レ	レ	レ
Cooperative	NG	NG	レ
Share end	NG	NG	レ

Case	child	adult	a member
Causal connection	レ	レ	レ
Cooperative	NG	NG	レ
Share end	NG	NG	レ
Seriousness	High	High	High
Predictability	?	?	Full

What is the predictability of intention ?

[Predictability of Intention] (“the P2’s predictability of intention of P1’s action”) in our cases is …

the possibility of picking up I_3 (the intention of demanding ransom) from a set of intentions which are candidates for the reason of P1’s action.



Reasonable adult case

Impoverished interpretation

- The point of the matter is *not* whether P2 discovers the purpose of P1's action.
- The point is…

"P2 picks up [I_1] from *limited resources*, viz. a small number of candidates for the reason of P1".

- But P2 could have expanded its resources if he had tried. But he had not tried.

Expected predictability of intention

[the child case]

Demanding ransom (I3)		other intention (I1 or I2 or…)
$\downarrow$	$<$	$\downarrow$
Giving a letter (E1)		Giving a letter(E2)

[the adult case]

Demanding ransom (I3)		other intention (I1 or I2 or…)
$\downarrow$	$>$	$\downarrow$
Giving a letter (E1)		Giving a letter(E2)

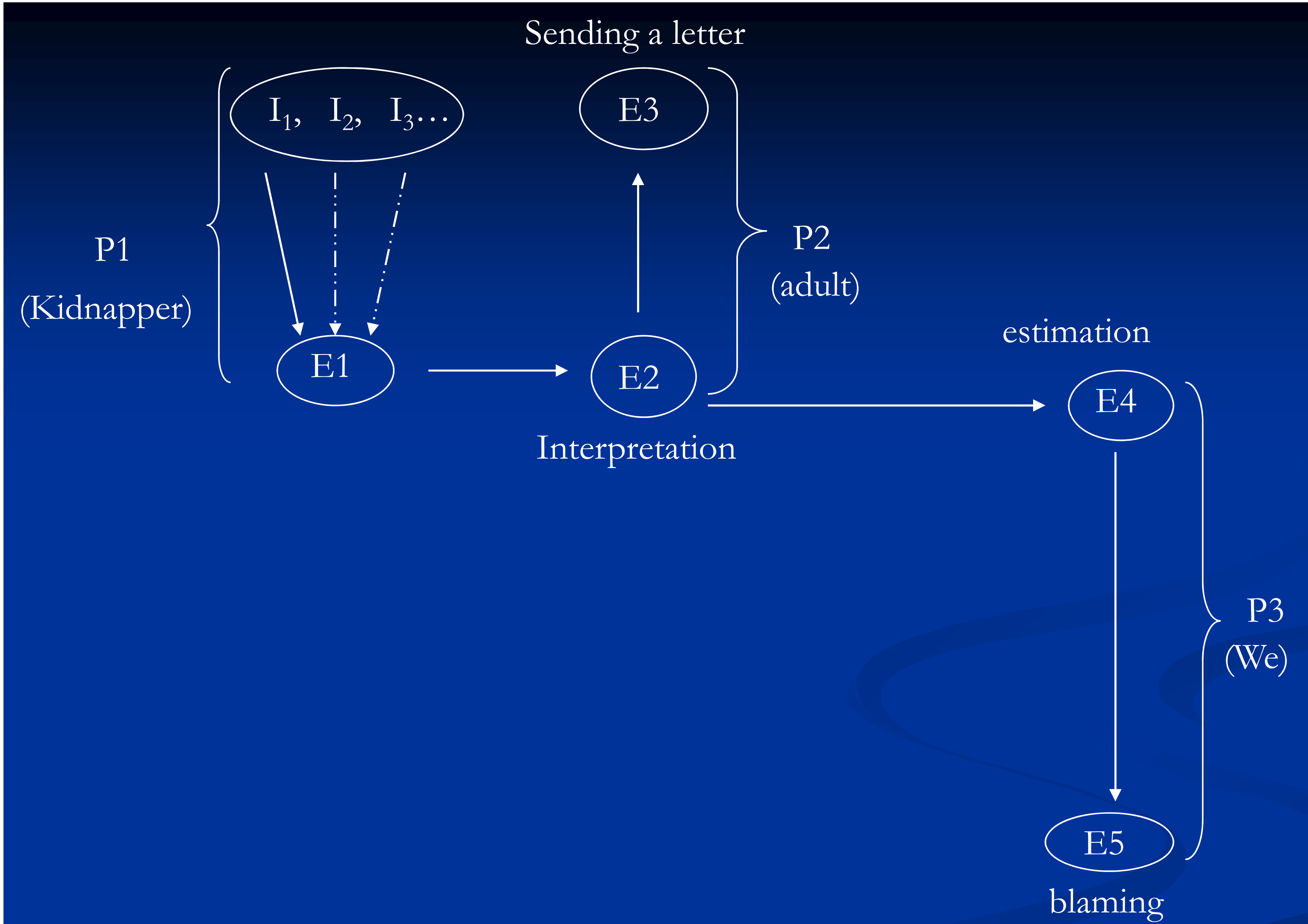
- *We* estimate that for the child it is so hard to predict that I_3 is the intention of P1's action (E1) in the child case.
- *We* expect the adult to predict that I_3 is the intention of P1's action (E1) in the adult case.

Expected predictability of intention [EPI]

- EPI depends on *our evaluating the situation* in which a person makes any action. Our evaluating the situation in the adult case, viz. big city where is abounding in crime, makes EPI strong.
- But, when the adult's action is located in another situation, for example a tranquil country town, we may say that he is not responsible for his action.

Thank you

for your attention


Sending a letter
I_1, I_2, I_3…
E3
P1
(Kidnapper)
P2
(adult)
E1
E2
estimation
E4
Interpretation
P3
(We)
E5
blaming

Social Commitments Expectations, obligations and entitlements

SOCREAL2010 – Sapporo, 27th-28th March

Miranda del Corral
Universidad Nacional de Educación a Distancia
mdelcorral@bec.uned.es

Outline

- Background: Individual, Social and Collective Commitments
- Theoretical framework for social commitments
  - Three features: expectations, creation process, operations on commitments
  - Towards a definition of social commitment
- The specificity of commitment: social norms and speech acts
- Conclusions

Outline

- Background: Individual, Social and Collective Commitments
- Theoretical framework for social commitments
  - Three features: expectations, creation process, operations on commitments
  - Towards a definition of social commitment
- The specificity of commitment: social norms and speech acts
- Conclusions

Background Issues

- The concept of commitment is widely used, but it has not been deeply analyzed.
- Commitment as a binding device ("glue"):
  - Between the agent and her goals (intention = stick to a plan). Implicit normativity in instrumental reasoning.
  - Between two (or more) agents in order to attain a shared goal. Social normativity.

Background Issues

- Commitment is used to explain
  - Intentional actions (Bratman, MAS)
  - Strategic actions (Schelling, Elster)
  - Counter-preferential actions (Sen, Searle)
  - Collective agency (Tuomela, Gilbert, Bratman)
- When explaining collective agency, [joint] commitment is used to explain how I-Intentions become We-Intentions.
- But this link is problematic: as it is not explained, it may make the argument circular.

Background Issues

- Kinds of commitments (Castelfranchi, 1995):
  - Individual (an individual agent's intentions)
  - **Social**
  - Collective (a collective agent's intentions)
- Social commitments (S-COMM) take this form:
  - (S-COMMxya) -> (OUGHTx(DOESxa))
  - "If x is S-Committed to y, he has a duty, an obligation, he ought to do what he is committed to" (Castelfranchi, 1995)
- Narrow/Wide formulations:
  - Wide: $O(A \rightarrow B)$ (e.g. von Wright)
  - Narrow: $A \rightarrow OB$ (e.g. Castelfranchi)

Outline

- Background: Individual, Social and Collective Commitments
- Theoretical framework for social commitments
  - Three features: expectations, creation process, operations on commitments
  - Towards a definition of social commitment
- The specificity of commitment: social norms and speech acts
- Conclusions

Towards a Definition of Social Commitment

- An explicit social commitment is a social interaction in which an agent acquires the obligation of performing an action, in virtue of another agent's agreement (attribution of obligation).

- Examples of explicit social commitments: promises, contracts, coordinated actions, orders given, etc.

- Features of social commitments not considered here:
  - Implicit social commitments
  - The role of the witness
  - Institutional commitments (contracts)
- Three features: expectations, creation process, and operations on commitments.

Towards a Definition of Social Commitment (I) Expectations

- Empirical expectations
  - The agent's beliefs about what facts will be the case (intentions, actions and states of affairs), and her motivational attitudes towards them.
  - Precondition for the commitment: what is possible, what is attainable, the agent's capabilities.
      - A minimum trust is necessary for commitment (Müller, 2008)
      - Similar arguments can be found in the case of intentions
  - A S-Comm can generate empirical expectations, specially regarding the agent's intentions.
  - Play an important role in the persuasive power of social commitments. They can be strategically created.

Towards a Definition of Social Commitment (II) Expectations

- Normative expectations:
  - Two senses of “normative expectations”:
    - Traditional sense (Sudgen, 1998): *what we think others should do* (→ normative beliefs)
    - Bicchieri's sense (Bicchieri, 2006): *what we think that others believe we should do* → we use this sense
  - They are generated with the act of committing
  - If they are pre-existent, committing may lose its point: is it reasonable to commit oneself to perform an action which is already mandatory?
  - Normative expectations and beliefs are the ground for considering the non fulfillment of a commitment as an *incorrect* action, this is, a *wrong* action.

Towards a Definition of Social Commitment Creation and Attribution

- Commitment creation and attribution
  - Explicit social commitments necessarily begin with the act of committing oneself, although they may not have a clear end, temporally speaking.
  - Committing oneself is not an individual action: it requires the other person's agreement. S-Comms cannot be unilaterally created.
  - S-Comm generates rights and entitlements, with the agreement of both agents: the creator and the attributer, or the debtor and the creditor (Singh, 1999).
  - Without agreement, the agent can feel internally committed, but there is not *social obligation*.

Towards a Definition of Social Commitment
Operations on Commitments

- S-Comms can change their state.
- Possible operations on S-Comms (Singh, 1999):
  - Creation and discharge (fulfillment)
  - Commitments may be cancelled (non-fulfilled) or released (no success or failure)
  - And also delegated (change of debtor) or assigned (change of creditor)
- The main idea is that social commitments describe a complex set of actions and choices, and not only the mere action of promising.
- The state of a S-Comm can be negotiated, in the same sense that its creation is negotiated.

Towards a Definition of Social Commitment
Preliminary definition

- A social commitment is a kind of social interaction involving, at least, two agents: a debtor (x) and a creditor (y), and an action to perform (z).
- The S-Comm is mutually known by x and y.
- The act os S-Committing is a social obligation for z acknowledged by x and y.

Towards a Definition of Social Commitment 4 conditions

If x is socially committed to y to perform the action z, then:

- (C1) y has empirical expectations concerning x's performance of z: z is a possible state and x is able to perform it.
- (C2) y has a normative belief concerning x's performance of z, in virtue of the S-Comm: because he is committed to, x should do z.
- (C3) x has normative expectations concerning her performance of z (and acknowledges the S-Comm as a reason for action): because he is committed to, x believes that y thinks that x should do z.
- (C4) x and y have performed a social action in which the S-Comm has been explicitly created, and its conditions are mutually known by the agents.

Towards a Definition of Social Commitment
4 Conditions

- "Self-defeating" commitments
  - Attributing a commitment while believing either that the goal will not be attained or that the debtor is not honest (C1) may be strategically useful, although they cannot be considered social commitments in the sense above.
- Dishonest commitments
  - Conditions 2 and 3 are not necessarily applied to x, otherwise dishonest commitments would not be commitments at all. Of course, (C2) and (C3) are applied to x in the case of honest commitments.

Towards a Definition of Social Commitment 4 conditions

- Commitments as obligations
  - Conditions 2 and 3 relate to the normative aspect of commitments. A social commitment is a social obligation:
    - It is a socially acknowledged reason for action. Even if there are no other motivations or reasons for performing z, committing oneself to perform z is a self-sufficient reason for performing z.
    - It is not necessary that the commitment itself is x's reason for action: we must distinguish between a reason for the agent and the agent's reason (Miller, 2006)

Outline

- Background: Individual, Social and Collective Commitments
- Theoretical framework for social commitments
  - Three features: expectations, creation process, operations on commitments
  - Towards a definition of social commitment
- The specificity of commitment: social norms and speech acts
- Conclusions

The Specificity of Social Commitments
Social Norms

- The act of committing, as a social action, is regulated by social norms (SN).
  - Let's consider the following norm: “commitments ought to be kept”
    - Is this norm *necessary* to explain S-Comm? **Yes**; otherwise, S-Comms would not be social obligations and therefore they would not create any specific rights or entitlements.
    - Is this norm *sufficient* to explain S-Comm? **No**; the content of the commitment is also norm-dependent:
      - Pointless commitment: S-Comm(xyz) = SN O(z) or SN O(C → z)
      - Contradictory commitment: S-Comm(xyz) = SN O(¬z) or SN O(C → ¬z)
      - Von Wright (1951): If ¬(PB) & (OA → B), then ¬(PA)

The Specificity of Social Commitments
Social Norms

- Social norms are impersonal commands, while social obligations refer to specific agents and circumstances
- A social norm may have a wide interpretation:
  - $O(A \rightarrow B)$
  - The relation between A and B is obligatory
- A social obligation (a S-Comm) has only a narrow interpretation:
  - $A \rightarrow RB$; $A \rightarrow OB$
  - A is a self sufficient reason for doing B; A makes B obligatory

The Specificity of Social Commitments Speech acts: persuasion and manipulation

- Persuasion: the paralelism between a S-Comm and an argument:
  - They both try to show a link between premises and conclusion:
    - Conclusion C follows from premises P1 and P2 in virtue of the reasonability* of infering C from P1 and P2.
    - Action z follows from my commitment to perform it in virtue of considering this act of committing a social obligation
  - They both can be accepted or rejected (but not for the same reasons)
    - You should take your unbrella because it's raining → Ok
    - You should eat your vegetables, because otherwise you won't have dessert → O

The Specificity of Social Commitments
Speech acts: persuasion and manipulation

- However:
  - This paralelism refers to:
    - The persuasive power of commitments (when they are used as arguments)
    - The similar structure of conditional S-Comms and normative arguments (A should be the case because of B), but not empirical ones (A is the case because of B)
  - These two aspects are easily mixed up, which can lead us to believe that S-Comms can be set unilaterally.

The Specificity of Social Commitments Speech acts: the Gricean Maxim

- Gricean maxim of quality: “be truthful: do not say what you believe to be false or lacks of supporting evidence”
  - If we consider MQ a SN, then it can be argued that all speech acts are S-Comms (except, maybe, those pointless or explicitly forbidden)
  - But: it is necessary that the speech act becomes a social obligation, a socially acknowledged reason for performing the action uttered in the speech act.
  - This makes the difference between S-Comm and declarations of intentions:
    - Ex: if John says to Ann: “I will go to the zoo tomorrow”, neither John or Ann recognize this speech act as a reason for John to go to the zoo.

The Specificity of Social Commitments Speech acts: Threats and Promises

- Conditional and unconditional S-Comm:
  - Conditional: x will perform z if/unless conditions C are satisfied
  - Unconditional: x will perform z whatever conditions exist
- Promises and threats as different speech acts
  - Are threats S-Comms?
    - If we assume that S-Comm implies a shared goal (or desire), then no, they aren't.
    - If we assume that a threat does not involve normative beliefs (y does not believe x should keep his threat), then no, they aren't.

The Specificity of Social Commitments
Speech acts: Threats and Promises

- Problems with this account:
  - The difference between a threat and a promise is defined by the *desirability* of the action, mainly for agent y.
    - Hard to define (t-inconsistency): the drunk's friend.
  - The difference between conditional threats and conditional promises is not clear.
    - A conditional threat is a kind of promise (of not performing the action if certain conditions are satisfied
- A threat will be a S-Comm if making this threat is acknowledged as a social obligation.
  - This differences between promises and threats as S-Comms and as mere declaration of intentions.

The Specificity of Social Commitments
Speech acts: conclusions

- S-Comms exceed the scope of the speech act analysis.
- Unlike speech acts, S-Comms cannot be unilaterally set or created.
- Setting a S-Comm requires communication between the agents, but cannot be identified with the speech act of committing itself.

Outline

- Background: Individual, Social and Collective Commitments
- Theoretical framework for social commitments
  - Three features: expectations, creation process, operations on commitments
  - Towards a definition of social commitment
- The specificity of commitment: social norms and speech acts
- Conclusions

Conclusions

- It is important to analyze S-Comms as non reducible to individual commitments, in order to explain collective commitments and agency.
- S-Comms are a social interaction which starts with, but is not limited to, the act of committing.

Thank you!