HOKKAIDO UNIVERSITY

Title	SOCREAL 2016 : Proceedings of the 4th International Workshop on philosophy and Logic of Social Reality [Full version]
Issue Date	2016
Doc URL	http://hdl.handle.net/2115/65209
Type	conference proceedings
Note	SOCREAL 2016 : Proceedingsof the 4th International Workshop on philosophy and Logic of Social Reality. 28-30 October 2016 Hokkaido University, Sapporo, Japan
File Information	SR2016_1.pdf

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

SOCREAL 2016

4th International Workshop on Philosophy and Logic of Social Reality

28 - 30 October 2016
Hokkaido University, Sapporo, JAPAN

Postproceedings

Edited by Tomoyuki Yamada

Under the Auspices of
Center for Applied Ethics and Philosophy (CAEP)
Graduate School of Letters, Hokkaido University
and
Department of Philosophy
Graduate School of Letters, Hokkaido University

Preface

Since the last years of the 20th century, a number of attempts have been made in order 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 earlier editions of SOCREAL Workshop has been held in March 2007, March 2010, and October 2013. Building upon the success of these editions, its fourth edition was held from 28 October till 30 October 2016.

There were two changes in this edition. Firstly, SOCREAL 2016 was held as one of the satellite workshops of

10th International Conference on Applied Ethics:
The Past, Present and Future of Applied Ethics,
October 28-30, 2016,
Hosted by the Center for Applied Ethics and Philosophy (CAEP),
Hokkaido University, Sapporo, Japan.

Secondly, its longer name was slightly modified in order to make its theoretical focus more evident. It is now called "International Workshop on Philosophy and LOGIC of Social Reality".

SOCREAL 2016 consisted of lectures by the invited speakers and the authors of the selected submitted abstracts as well as lively discussions by participants, and this post-proceedings includes either the (revised) extended abstract or the presentation slides for each lecture presented in SOCREAL 2016. Accepted lectures given at SOCREAL 2016 were selected by peer reviewing of the abstracts by the members of the program committee. We thank all the researchers who submitted their papers for their interest in SOCREAL 2016, the members of program committee for their reviewing work, and all the participants for their lively discussions.

WORKSHOP CO-CHAIRS
Tomoyuki Yamada
Berislav Žarnić

Committees

WORKSHOP CO-CHAIRS
Tomoyuki Yamada
Berislav Žarnić

PROGRAM COMMITTEE

Thomas Ågotnes (University of Bergen, Norway, and Southwest University, China)
Johan van Benthem (University of Amsterdam, The Netherlands, and Stanford University, USA, Tsinghua University, China)
Jose Carmo (Universidade da Madeira, Portugal)
Mamoru Kaneko (Waseda University, Japan)
Fenrong Liu (Tsinghua University, China)
Yuko Murakami (Tohoku University, Japan)
Yasuo Nakayama (Osaka University, Japan)
Mitsuhiro Okada (Keio University, Japan)
Manuel Rebuschi (Nancy University, France)
Katsuhiko Sano (Hokkaido University, Japan)
Taiki Takahashi (Hokkaido University, Japan)
Allard Tamminga (Rijksuniversiteit Groningen, The Netherlands)
Tomoyuki Yamada (Hokkaido University, Japan)
Berislav Žarnić (University of Split, Croatia)

LOCAL ORGANIZING COMMITTEE

Shunzo Majima (Hokkaido University)
Katsuhiko Sano (Hokkaido University, Japan)
Tomoyuki Yamada (Hokkaido University)

The Program and the Table of Contents

DAY 1: 28 OCT Friday

12:00- [Registration]

14:35-14:40 [Opening] Tomoyuki Yamada & Berislav Žarnić

[Session 1] Logic and Society in Networks

DAY 2: 29 OCT Saturday

[Session 2] Logic, Knowledge, and Preferences

13:00 - 14:30 Lunch Break

[Session 3] Action, Norm, and Belief

DAY 3: 30 OCT Sunday

[Session 4] Speech Acts, Gradability, and Value

Reasoning about Local Broadcast Messages in Social Networks (Extended Abstract)

Thomas Ågotnes Jeremy Seligman

July 1, 2016

1 Introduction

Information flow in social networks has been studied in the field of social network analysis [9] for a long time. Phenomena such as information cascades [4, 13] and power laws [7] describe how local influence between people and their friends leads to global collective outcomes that aggregate local behaviour in different ways.

At the same time, the logical dynamics of information-transmitting interaction between agents have been studied on a more detailed level in the field of epistemic logic [5, 12]. This level of analysis is rich and non-trivial and reveals many subtle issues, in particular when information about the information possessed by other agents is communicated. A typical example is communication of the so-called Moore sentence: "p is true and you don't know it". Even if this sentence is true the moment before it is communicated, it becomes false the moment after, contradicting the intuition that communication leads to a monotonic accumulation of correct information.

While research in epistemic logic has focussed on describing the epistemic pre- and post-conditions of different types of information-transmitting events, the logical dynamics of information flow in social networks have mostly been considered only very recently [10, 11, 6]. It turns out that information-transmitting events taking place in social networks exhibit many special interesting logical phenomena.

While social networks have been studied long before the emergence of *electronic* social networks, the latter have given us new types of communication. One example is *local broadcast* messages. These are messages that are sent by an agent and received by all her friends (or "followers") at the same time. Most major electronic social network platforms, such as Facebook, Twitter, Sina Weibo, WeChat and Line, have implemented such broadcast communication in some form or other. Indeed, on some platforms, such as Twitter, this is the *main* form of communication.

In this paper we study the logical dynamics of Twitter-like local broadcast communication. In particular, we identify some interesting logical dynamics phenomena as well as some logical challenges that are particular for this setting. We propose formal logical languages for reasoning about this type of communication, and formalise a number of different communication protocols in this setting.

2 Quantifying over announcements

In standard epistemic logic we use expressions like $K_a p$ to say that agent a knows proposition p. Here we are interested in reasoning about the possible knowledge states that can result from local broadcast messages in networks. What kind of logical operators can we use for such reasoning on an interesting level of abstraction? A natural choice is to introduce operators for *quantifying* over possible messages, making it possible to say which states of affairs an agent can and cannot make come about by sending some local broadcast message. Logics for reasoning about what one or several agents can make come about have been of significant interest recently, the most important frameworks being Coalition Logic [8], Alternating-time Temporal Logic [2] and Seeing-To-It-That (STIT) Logic [3]. What these logics have in common is a logical operator of the form $\langle a \rangle$, such that a formula of the form $\langle a \rangle \phi$ means that agent a can make ϕ come about. Recently [1] these operators have been investigated for the case of *public announcements*: a can make ϕ come about by making some public announcement.

Here, we propose to use similar operators (extending standard epistemic logic) to reason about local broadcast messages in networks. Some examples of formulas are:

1. $\langle a \rangle K_b p$: agent a can make agent b know p. This could be true even if a does not know p; maybe a sends the message q and b knows that $q \to p$. This could also be trivially true if b already knew p.

2. $\neg K_a \langle a \rangle K_b p$: ... but a does not know it

3. $\neg K_b p \wedge \langle a \rangle K_b p$: agent a can make agent b know p, and he didn't already know it. This can only be true if b follows (is a friend of) a.

What are the possible axioms of this logic? The third example above hints to a universally valid formula: if it is true, then b must follow a, in which case a can tell b anything she knows. A candidate axiom is thus:

$$(\neg K_b p \wedge \langle a \rangle K_b p) \to (K_a q \to \langle a \rangle K_b q). \tag{1}$$

Depending on assumptions about exactly how agents update their knowledge states when receiving messages from their friends, there are also other potential axioms, such as $K_b p \to \neg \langle a \rangle \neg K_b p$ (montonicity). However, we can already see

potential problems with attempting to find a *complete* axiomatisation. Consider the following set of formulas:

$$\Gamma = \{\neg K_b p \wedge \langle a \rangle K_b p\} \cup \{K_a \phi \to K_b \phi : \text{ any } \phi\} \tag{2}$$

This says that a can make b know p, and b does not already know it (so he must follow a), but that b knows everything that a knows. Intuitively, Γ is not consistent – it represents a contradiction. It cannot be possible that a can make b know something new if b already knows everything a knows. However, any finite subset of Γ intuitively *is* consistent, because such as subset will not say that b knows everything a knows. This means, in meta-logical terms, that any logic based on these operators is not semantically *compact*. One consequence of this is that such a logic cannot be *strongly* complete. Another is that even for *weak* completeness, the standard canonical model method cannot be used directly (the standard truth lemma would fail).

3 Secret Tweeting and Learning Network Structure

You are a secret agent with a secret S that you would like to transmit to a fellow agent a unobtrusively using a very public network like Twitter. Any information you tweet will be received by your followers on the network. You correctly assume that they will send the message on to their followers (retweet it) if and only if it does not conflict with any information they already possess. With luck, your message will be tweeted through the network until it eventually reachers a. Under what conditions is it possible for you to convey S to a in this way, without other agents in the network learning this information? Clearly, you cannot tweet S itself, but if, for example, a is the only agent who knows that K, then the message "if K then S" may work, if there is a suitable path from you to a. To know whether you can succeed or not and what to tweet, you need to know something about the network and the information already possessed by the other agents. But you can learn something about this with a test tweet. If, for example, you know that you have two followers b and c and only b believes P and then you tweet the message $\neg P$. Then if, after a certain length of time, someone tweets P to you, you know that there is a loop back to you via c.

We also look at how can agents can learn information about the network structure from broadcast messages from their friends, even when the messages does not explicitly contain such information. A simple example is that if you tell me that the party is on Saturday, I know that you must be friends with Ann because I know that only she knew that.

References

[1] Thomas Ågotnes, Philippe Balbiani, Hans van Ditmarsch, and Pablo Seban. Group announcement logic. *Journal of Applied Logic*, 8(1):62–81, 2010.

[2] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49:672–713, 2002.

[3] N. Belnap and M. Perloff. Seeing to it that: a canonical form for agentives. *Theoria*, 54:175–199, 1988.

[4] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, pages 992–1026, 1992.

[5] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning About Knowledge*. The MIT Press, Cambridge, Massachusetts, 1995.

[6] Fenrong Liu, Jeremy Seligman, and Patrick Girard. Logical dynamics of belief change in the community. *Synthese*, 191(11):2403–2431, 2014.

[7] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[8] M. Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002.

[9] John Scott. *Social network analysis*. Sage, 2012.

[10] Jeremy Seligman, Fenrong Liu, and Patrick Girard. Logic in the community. In *Indian Conference on Logic and Its Applications*, pages 178–188. Springer, 2011.

[11] Jeremy Seligman, Fenrong Liu, and Patrick Girard. Facebook and the epistemic logic of friendship. *arXiv preprint arXiv:1310.6440*, 2013.

[12] H.P. van Ditmarsch, W. van der Hoek, and B.P. Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library*. Springer, 2007.

[13] Ivo Welch. Sequential sales, learning, and cascades. *The Journal of finance*, 47(2):695–732, 1992.

Forms of organizing human society inspired by technological structures

Alfons Josef Schuster
School of International Liberal Studies, Waseda University
1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050 Japan
a.schuster@aoni.waseda.jp

Abstract

Fundamentally, any system (simple or complex) consists of objects (natural or artificial) and relationships between these objects. A major goal in systems theory concerns the discovery of general patterns or principles that apply to a wide range of (ideally all) systems. From this perspective, this paper investigates the patterns and other relationships that may emerge between computer networks and organizations of human society. Our investigation emphasizes not only that computer networks reflect human society in various ways, but also that new ways of organizing human society inspired by technological structures may be emerging.

1 Introduction

This paper extends our ongoing work in computer science and artificial intelligence [Schuster, 2007; 2008; Schuster and Yamaguchi, 2011]. More specifically, our work is on the interface between the aforementioned areas and the social science domain. The motivation for our efforts come from observations, in our work [Schuster and Yamaguchi, 2009] and elsewhere [Caldarelli and Catanzaro, 2012; Kadushin, 2012, e.g.], that computer networks and human societies have many properties in common. We find that these commonalities are not only restricted to network topologies. Indeed, they include many of the behaviors (e.g., communication protocols) demonstrated by various types of agents (e.g., humans, software agents, or robots) sharing networked environments.

For instance, the *client-server* model (e.g., for a business website) and the *peer-to-peer* model (e.g., for a file sharing application) represent two ways in which a distributed application may be structured on a network (e.g., the Internet or some other network environment) [Gray, 2011, pp.67–88]. The client-server model is one of the most common network architectures. In this model, one or more computers act as servers to the rest of the network (the clients). On the other hand, in a peer-to-peer network there is no central service provider (server). All participating computers (clients) have equal status. From the viewpoint of our work, we find that these two types of architectures map into the domain of human society without much ado. For instance, a peer review process may involve the evaluation of some body of work between people of similar standing and competence (the peers), while a travel agent (the server) may provide its services to a larger customer base (the clients).

In the client-server and the peer-to-peer example, the clients, servers, and peers, are physical objects (computers). It is necessary to understand that parallel to this physical dimension, there also exists the less tangible (e.g., in the case of wireless technology) communication dimension, in which the various objects exchange data. In this domain, it is possible to distinguish communication protocols such as a *broadcast* or a *token ring* technology. It is easy to understand these protocols in the human domain. Imagine a teacher handing out assignments to his students. One possibility for the teacher is to simply call the name (say X) of a student. In case the student is present, the student might raise a hand and receive the assignment from the teacher. Note that in a broadcast the call goes out to all students at the same time. In a token ring scenario, the teacher goes from one student to the next student asking "Are you student X?". In the affirmative case, the student receives the assignment. In the negative case, the teacher moves on, and repeats the question to the next student. Note that in the token ring case the call goes out individually and sequentially. In reality, the number of communication protocols in computer networks is large, and so is the range of tasks they cover, which includes: preventing data packets from colliding on a network, routing of data, avoiding congestion, preventing errors in incoming and outgoing data, assuring data integrity, or some form of security, naming just a few tasks.

2 Interplay between human society and technological structures

From one point of view, humans design and produce artificial network environments that appear to mirror structures and behaviors underlying the very way human societies organize themselves. From a reversed point, we can observe that the very creations of human enterprise (the computer networks) feed-back a stimulus that encourages the emergence and formation of new behaviors and new organizational patterns in human society and human conduct. For instance, the production of a reliable and robust network (out of unreliable parts) was a major goal of the designers of the early Internet.

In order to achieve these goal, these designers developed and implemented the now well-known design principles of decentralization, end-node verification, dynamic routing, or packet switching [Casad, 2011]. In return, the Internet of today encourages social scientists and researchers in other domains to contemplate a new type of decentralized human society. This new type of society may embrace concepts such as decentralized communication, decentralized law, decentralized energy production, and decentralized finance, all facilitated by the Internet[1]. Clearly, the intensity with which these ideas can be pursued today would have been impossible in the pre-Internet (pre-computing) era.

Of course, whether we are going to witness the migration of modern humanity society into such a form of decentralized society, or whether it may be a process of meandering into an all-embracing infosphere (a reality envisaged by the philosopher Floridi [Floridi, 2010]), is debatable. In this work we are going to comment on this debate in some detail. Our main thrust would be a desire to investigate the bidirectional interplay between computer networks and human society in general. Obviously, this investigation requires us to comment on ideas that are more or less well-known, such as scale-free networks [Barabási and Bonabeau, 2003], or crowdsourcing [Kadushin, 2012]. However, we would also like to use this opportunity to discuss two types of organizational patterns that seem to have some degree of novelty according to our understanding of the field. In this paper, we refer to these novel patterns of organization as *swarm societies*, and *entropic societies*.

3 Swarm societies and entropic societies

In order to understand the forthcoming interpretation of these terms, we first compress the development of computer and network technology into three distinguishable phases.

Although there are various other names worth mentioning, the first phase may include Alan Mathison Turing (1912–1954), John von Neumann (1903–1957), or Claude Elwood Shannon (1916–2001). In simple terms, Turing outlined the limits of what computers can do by formulating the ultimate digital machine, the so-called *Universal Turing Machine*. Von Neumann's contributions include the description of a computer architecture, the so-called *von Neumann architecture*, that remains a fundamental design feature of any modern-day computer. On the other hand, Shannon is often referred to as the father of the mathematical theory of information. Crudely, this theory describes the effective encoding and transfer of data through a communication channel (e.g., a computer network).

The second distinct phase should include names such as William Henry (Bill) Gates III (born 1955), a late Steven Paul (Steve) Jobs (1955–2011), Vinton Gray Cerf (born 1943), or a Sir Timothy (Tim) John Berners-Lee (born 1955). These names stand representatively for many people involved in the realization of powerful computer systems, their underlying hardware and software, as well as for the invention of inventions, the *Internet* and the *World Wide Web* (Web).

Lastly, the third phase of progress revolves around a generation that is now in their early 30s to early 40s and includes the founders of companies like *Google* (founded 1998), *Facebook* (founded 2004), *YouTube* (founded 2005), or *Twitter* (founded 2006). All of these companies pursue their business under the names *Web 2.0* or *social web*. The term social web is an extremely powerful abstraction. Among other things, the social web urges information society to redefine traditional values such as ownership [Heaven, 2013], or friendship [Brent, 2014]. In addition, the term also stands synonymously for the seamless integration, augmentation, and infiltration of computing devices (e.g., tablet computers, smartphones, computer games, or virtual reality glasses), including the services these devices provide, into our information hungry society as a new way of life. The relatively young, interdisciplinary academic field of new media studies might be seen as a response to these developments. The field explores a wide range of issues on the intersection of computing, science, the humanities, and the visual and performing arts [Press and Williams, 2010].

Whatever the responses are, it is important to understand that a crucial feature of this new environment lies in the potential for individual users (who do not need much technical expertise) to contribute and express themselves intelligently and creatively. So-called *content management systems* are among the technologies facilitating this kind of participation. A content management system is like a tool that allows users to decorate an empty room according to their individual tastes. In our context, such a room could be a Facebook or a Twitter account. Initially, such an account is empty in terms of user specific content. Over time, however, its content evolves and (usually) represents a form of virtual home and identity of its user. Let us use an analogy. The early Internet was like a skeleton. The Web was weaving a skin and some cloths onto this skeleton. Eventually, content management systems transformed the Internet into a catwalk where human users metamorphose into attractive virtual models that in some way behave artificially alive.

From a historical perspective, the instruments of technology played by a global orchestra of users seem to constitute a form of reality contemplated by visionaries such as Paul Marie Ghislain Otlet (1868-1944), or the German artist Joseph Beuys (1921-1986). Paul Otlet is often considered to be the father of *information science*. Among other things, he carried the vision that knowledge is going to have a positive impact on humanity and world peace. The so-called *Mundaneum*[2], a global (universal) repository for all the world's knowledge, is a powerful testament of this vision. On the other hand, the conceptual artist Joseph Beuys expressed his vision of a society in which *"jeder Mensch ein Künstler sei"* (where everyone is an artist). Besides this famous phrase, Beuys also coined the mystical term *social sculpture*. Interpretations of this term understand social sculpture foremost as a process of transformation or shaping of society through

[1] Johann Gevers. The four pillars of a decentralized society. https://www.youtube.com/watch?v=8oeiOeDq_Nc. Accessed: 2016-07-14.

[2] The Mundaneum. http://www.mundaneum.org/en. Accessed: 2016-07-16.

the collective creativity of its members. Past perception was that driving this process is the role of the traditional artist. Today, as we saw before, it could be everyone (who has access to the Internet).

3.1 Swarm societies

We investigate swarm societies from the points of centralization and decentralization. In computing, the field of *swarm intelligence* studies the collective behavior of decentralized, self-organized systems (natural or artificial) [Hassanien and Emary, 2015]. Individual agents in such systems, usually, do have limited ability. In unison, however, they often achieve outcomes that seem to be beyond the ability of an individual agent. Take the case of the captivating flight behavior among migrating birds. In a close-up, the trajectory of a single bird looks erratic and unorganized. Yet, from a wide-angle shot perspective, the flock of birds remains together and stable, hence *self-organized*.

Human societies exhibit various forms of de/centralization, too. A human society may have a center in terms of location (e.g., from caves to our solar system), or leadership (e.g., from tribal chiefs to emperors). Likewise, a society may feel decentralized as in being lost in the fastness of space (location), or enjoying a status of equality as in animal farms (leadership). We believe that human societies can experience a form of intellectual de/centralization, too. An example from the domain of literature may support this view. The so-called *literary canon* is an important idea in literature and culture. Simplistically, a literary canon may describe a body of books (e.g., the works of William Shakespeare) that have been traditionally accepted by scholars as the most important and influential in their society or culture. More recently, those researchers who investigated large-scale trends in literary style found that the influence of classic literature (the literary canon) on contemporary writers is declining [Hughesa *et al.*, 2012]. Their research on a large body of literary works (current and past) revealed several trends including that: (i) authors of any given period are similar in style to their contemporaries, (ii) the stylistic influence of the past is decreasing, and (iii) authors writing in the late 20th century are instead strongly influenced by other contemporary writers. From our point of view, therefore, the literary canon represents a centralized form of literature, while the current literary scene appears more decentralized, self-organized, and swarm-like. The literary canon is one example. Other examples, may come from the world of on-line publishing, open-access, blogs, wikis, and various other environments holding an ingrained capacity to shape and impact the intellectual sphere of human societies.

3.2 Entropic societies

Entropy is a fundamental concept in physics. We distinguish two types of entropy: *natural* entropy and *artificial* entropy[3]. An ice cube melting in a glass of water is an instance of natural entropy. The ice cube changes from a state of low entropy (generally associated with a form of higher organization, here ice) to a state of increased entropy (generally associated with a state of lower organization, here water). The process happens without any enforcement, spontaneously and naturally, according to the laws of physics. Artificial entropy is similar to natural entropy in considering situations where systems change from states of high organization (low entropy) to those of low organization (high entropy), and vice versa. Crucially, however, it is human thought and rational thinking that initiates these state changes[4].

Events such as the French Revolution, for instance, represent cases of artificial entropy. The French Revolution transformed a monarchy (high organization, low entropy) in a process of social and political upheaval into an intermittent state of chaos (low organization, high entropy), before converging into a modern democracy (high organization, low entropy again). For us, the crucial step is the intermittent state of chaos with its rapid increase of artificial entropy. More generally, we suggest the concept of artificial entropy as a tool for analyzing situations where systems, triggered by human thought and intent, experience rapid and massive organizational change. The Internet and the social network domain represent such a system in our eyes. For instance, the starting point in the work of audiovisual artist Davide Quagliola are popular paintings (high organization, low entropy) of the *Old Masters* (e.g., Botticelli, or Rubens)[5]. Initially, Quagliola produces high-resolution digital images of these paintings. A computer program then repeatedly fragments these digital images in a series of intermittent images (low organization, high entropy) until, in a process of inspection, a new object of art (high organization, low entropy again) that is aesthetically pleasing to the artist's eye emerges. We hypothesize that similar processes of repetitive destruction and reorganization are an essential part of human existence. Old ideas and concepts not only gradually change, they are also often smashed and utterly destroyed. Far from being useless debris, eventually, these high entropy remnants are the material out of which new, low entropy structures and forms of organization will emerge. We feel that the Internet and its environment are a platform where such processes can thrive and grow deep into the fabric of human society and beyond. This is why we feel that calling such forms of organization entropic societies may not be the least sensible thing to do.

4 Summary

The motivation in this work was to investigate the relationships that may emerge between computer networks and organizations of human society. A range of examples demonstrated these relationships in various ways, including: structural forms, the form of communication protocols, or in pat-

[3]Please note that this distinction is not entirely new. The philosopher Floride speaks about anti-entropic information entities and processes in his work [Floridi, 2010; 2011]. Our motivation here is to pursue and understand (and hopefully extent) this distinction as much as possible.

[4]Understandably, such a distinction should lead to a wider discussion. For the sake of brevity, however, this text omits a detailed exploration of this topic.

[5]Meet Audiovisual Auteur Quayola. `https://www.youtube.com/watch?v=fTw8b9Pwmr4`. Accessed: 2016-07-16.

terns of dynamic behavior. We commented on two specific types of organization, swarm societies and entropic societies. We rationalized that new infrastructures such as the Internet and the social web may be environments in which such societies can flourish.

References

[Barabási and Bonabeau, 2003] A.L. Barabási and E. Bonabeau. Scale-free networks. *Scientific American*, 9(2):50–59, April–June 2003.

[Brent, 2014] L. Brent. Friendship: Friends with many benefits. *New Scientist*, 2970:37–39, 2014.

[Caldarelli and Catanzaro, 2012] G. Caldarelli and M. Catanzaro. *Networks: A Very Short Introduction*. OUP Oxford, 2012.

[Casad, 2011] J. Casad. *Teach Yourself TCP/IP in 24 Hours*. Sams, 5th edition, 5 edition, 2011.

[Floridi, 2010] L. Floridi. *Information: A Very Short Introduction*. OUP Oxford, 2010.

[Floridi, 2011] L. Floridi. *The Philosophy of Information*. OUP Oxford, 2011.

[Gray, 2011] S. Gray. *Information Technology in a Global Society*. CreateSpace Independent Publishing Platform, 2011.

[Hassanien and Emary, 2015] A.E. Hassanien and E. Emary. *Swarm Intelligence: Principles, Advances, and Applications*. CRC Press, 2015.

[Heaven, 2013] D. Heaven. Lost in the cloud: How safe are your online possessions? *New Scientist*, 2910:35–37, 2013.

[Hughesa *et al.*, 2012] J.M Hughesa, N.J. Fotia, D.C. Krakauerb, and D.N. Rockmore. Quantitative patterns of stylistic influence in the evolution of literature. *Proceedings of the National Academy of Sciences of the United States of America*, 109(20):7682–7686, 2012.

[Kadushin, 2012] C. Kadushin. *Understanding Social Networks: Theories, Concepts, and Findings*. Oxford University Press, 2012.

[Press and Williams, 2010] A.L. Press and B.A. Williams. *The New Media Environment: An Introduction*. Wiley-Blackwell, 1 edition, 2010.

[Schuster and Yamaguchi, 2009] A. Schuster and Y. Yamaguchi. The survival of the fittest and the reign of the most robust: in biology and elsewhere. *Minds and Machines*, 19(3):361–389, 2009.

[Schuster and Yamaguchi, 2011] A. Schuster and Y. Yamaguchi. From foundational issues in artificial intelligence to intelligent memristive nano-devices. *International Journal of Machine Learning and Cybernetics*, 2(2):75–87, 2011.

[Schuster, 2007] A. Schuster. Intelligent computing everywhere. In A. Schuster, editor, *Intelligent Computing Everywhere*. Springer Verlag, 2007.

[Schuster, 2008] A. Schuster. Robustness in nature as a design principle for artificial intelligence. In A. Schuster, editor, *Robust Intelligent Systems*. Springer Verlag, 2008.

On interdependence between belief updates and reliability structures: An approach from two-dimensional hybrid logic

Katsuhiko Sano

Philosophy Department, Graduate School of Letters, Hokkaido University, Japan
v-sano@let.hokudai.ac.jp

1 Introduction

We, social agents, update our beliefs with new information based on our reliability structures among agents. When an agent receives a piece of information from a reliable sender, it is likely that she changes her belief. If the information is from an unreliable sender, she might discard the information. Conversely, our reliability structures might be affected by the information that agents provide. An agent might downgrade the sender in her reliability structure when she receives unreliable information from the sender. This paper aims to model this interdependence between belief updates and reliability structures. This paper employs Lewis [5]'s system of spheres to model beliefs as plausibility structures between states and reliability structures as relations between agents. To capture the interdependence between the two structures within a single framework, this paper conceptually follows a semantic idea of epistemic logic of friendship [11, 12], applies Sano [9]'s technique of (dependent) product of hybrid logics, and propose two dynamic operators, one for plausibility structures and the other for reliability structures. We use the following as our running example in this paper.

Example 1. Let a and b be logicians and each of them believes that she solved a difficult open problem called "no.7." Both start communicating with a head editor c of an excellent nice logic journal. After the editor c receives emails from logicians a and b, c believes that a solved the open problem no.7, while c does not believe that b solved the open problem. This is because the logician a is more reliable to the editor c than the logician b.

We proceed as follows. Section 2 introduces Lewis' system of spheres and explains how the notion of relativized belief and reliability are formalized over a system of spheres. Section 3 follows a semantic idea of epistemic logic of friendship [11, 12] to combine the connective for relativized belief and the connective for relativized reliability in a single syntactic framework called two-dimensional hybrid logic [9]. Section 4 adds two dynamic operators which allow us to capture the interdependence between plausibility structures between states and reliability structures as relations between agents. We note that similar dynamic operators were already proposed in one-dimensional hybrid counterfactual logic [10] (for hybrid counterfactual logic, the reader is referred to [8]).

2 Lewis' System of Spheres

Lewis' system of spheres [5] on a fixed domain (a set of states or a set of agents) are used to model the notion of "similarity" on the domain from a chosen element's perspective. Fig. 1 provides a graphical representation of a system of speres. For example, when the fixed domain is a set of states, then a system of spheres may model a fixed agent's plausibility relation between states from a chosen state's perspective, e.g., we may say that a state w' is as least plausible to the fixed agent as a state w'' from a state w's perspective. When the fixed domain is regarded as a set of agents, then a system of spheres may model a reliability relation between agents from a chosen agent's perspective and so we may say that an agent b is at least reliable as an agent c from an agent a's perspective. Formally, we say that (X, σ) is a *system of spheres* if X is a non-empty set and $\sigma : X \to \wp\wp(X)$ is a mapping such that, for every $x \in X$, a set $\sigma(x)$ of *spheres* satisfies the following three conditions:

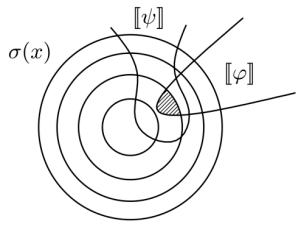


Fig. 1. A system of spheres and the semantics of $\varphi \mathrel{\Box\!\!\rightarrow} \psi$

1. $\sigma(x)$ is closed under taking (arbitrary) unions and (arbitrary) non-empty intersections.
2. $S \subseteq T$ or $T \subseteq S$ for all spheres $S, T \in \sigma(x)$.
3. $\sigma(x)$ satisfies the following *limit assumption* (or *minimality condition*), i.e., if $Y \cap \bigcup \sigma(x) \neq \emptyset$ then there exists a least[1] sphere $S \in \sigma(x)$ such that $Y \cap S \neq \emptyset$.

On the top of the Boolean connectives as well as propositional atoms, Lewis introduced the binary connective $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ as a *counterfactual connective*: "If it were the case that φ, then it would be the case that ψ." Given a system of sphere (X, σ) and a valuation V for propositional atoms, the satisfaction of the connective $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ is defined as follows:

$$(X, \sigma, V), x \models \varphi \mathrel{\Box\!\!\rightarrow} \psi \text{ iff } \mathrm{Min}_{\sigma(x)}([\![\varphi]\!]) \cap [\![\varphi]\!] \subseteq [\![\psi]\!]$$

where $[\![\varphi]\!]$ is the truth set of φ, which is defined as $\{\, x \in X \mid (X, \sigma, V), x \models \varphi \,\}$, and $\mathrm{Min}_{\sigma(x)}([\![\varphi]\!])$ means the least sphere in $\sigma(x)$ which intersects $[\![\varphi]\!]$ (when $[\![\varphi]\!]$ does not intersect any spheres, it is understood as an empty set). The shaded area in Fig. 1 corresponds to the set $\mathrm{Min}_{\sigma(x)}([\![\varphi]\!]) \cap [\![\varphi]\!]$ and it is easy to see that the shaded area is a subset of the truth set $[\![\psi]\!]$ of ψ and so $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ is true at $x \in X$ in Fig. 1.

This binary connective can also formalize the notion of relativized belief of a fixed agent when a given domain is interpreted as states and $\sigma(x)$ is understood as representing a plausibility relation of the fixed agent. Instead of $\varphi \mathrel{\Box\!\!\rightarrow} \psi$, let

[1] The word "least" should be understood in terms of $\subseteq$.

us use the symbol $\mathsf{B}(\varphi, \psi)$ for an agent's relativized belief and read it as "Under the condition of φ, an agent believes that ψ." Then the above semantics tells us that $\mathsf{B}(\varphi, \psi)$ is true at x, if, and only if, ψ holds at the most plausible φ-states with respect to x. When φ is the constant $\top$ for a tautology, $\mathsf{B}(\top, \psi)$ means an *absolute* belief, which is read as "an agent believes that ψ." On the other hand, when a given domain is interpreted as agents, $\Box\!\!\rightarrow$ can formalize the notion of relativized reliability. Instead of the symbol, let us use the symbol $\mathsf{R}(\varphi, \psi)$ whose reading is "All the most x's reliable agents being φ satisfy ψ, " where it is noted that a formula φ should be regarded as a property or a unary predicate for agents.

3 Combining Agent's Belief and Reliability in Two-dimensional Hybrid Logic

To talk about interaction between an agent' belief and a reliability relation among agents, we need to put the connectives $\mathsf{B}(\varphi, \psi)$ and $\mathsf{R}(\varphi, \psi)$ into one syntax and we should provide a uniform semantics for both operators. For this purpose, we follow a semantic strategy of Epistemic Logic of Friendship [11, 12]. In the traditional semantic view (cf. [4]), we evaluate a formula in terms of a given state as:

$$w \models \mathsf{B}_a(a \text{ solved an open problem no.7}).$$

where B_a (read "a believes that –") is a belief operator. In a semantic view of the epistemic logic of friendship, we extract the information of agents from the above formula and push it into the semantic side as:

$$(w, a) \models \mathsf{B}(\text{I solved an open problem no.7}).$$

It is noted that, under this new semantic view, a propositional atom p is read indexically as "I am p" or "I satisfy p."

Now let us see the technical details. We start with combining a system of spheres for plausibility with a system of spheres for reliability. Let W be a possibly infinite set of states and A be a possibly infinite set of agents. For every agent $a \in A$, we prepare a mapping $\sigma_a : W \to \wp\wp(W)$ for a system of spheres for agent a's plausibility. Moreover, for every state $w \in W$, we set up a mapping $\rho_w : A \to \wp\wp(A)$ for a reliability structure at state w. This means that each agent a's reliability may change through states. Our structure combining plausibility and reliability is summarized as a tuple

$$F = (W, A, (\sigma_a)_{a \in A}, (\rho_w)_{w \in W}),$$

which is called a *PR-structure*.

Now let us move to our syntax. To combine two connectives $\mathsf{B}(\varphi, \psi)$ and $\mathsf{R}(\varphi, \psi)$ into one syntax, we employ a two-dimensional hybrid logic framework proposed by the current author in [9]. That is, we introduce state nominals i (syntactic names for states) and agent nominals n (syntactic names for agents)

and satisfaction operators $@_i$ or $@_n$ indexed by a nominal, which allow us to "jump" to a named state by a state nominal or a named agent by an agent nominal, respectively. For an introduction to hybrid logic, the reader is referred to [2, 1]. Now, the set of formulas in our static syntax $\mathcal{L}$ is defined inductively as follows:

$$\varphi ::= p \mid i \mid n \mid \neg\varphi \mid \varphi \wedge \varphi \mid @_i\varphi \mid @_n\varphi \mid \mathsf{B}(\varphi, \varphi) \mid \mathsf{R}(\varphi, \varphi),$$

where p, i, or n is a member of a countably infinite set of *propositional atoms*, or *state nominals*, or *agent nominals*, respectively. The other Boolean connectives than $\neg$ and $\wedge$ are defined as usual. Here are some examples of how to read formulas:

- $\mathsf{B}\varphi := \mathsf{B}(\top, \varphi)$, which is read as "I believe that I am φ."
- $@_n\mathsf{B}p$, which is read as "n believes that she is p."
- $\mathsf{B}@_np$, which is read as "I believe that agent n is p."
- $\langle\mathsf{R}\rangle(\varphi, m)$, where it is read as "$m$ is one of the current agent's most reliable agents being φ" and it is note that $\langle\mathsf{R}\rangle(\varphi, \psi)$ is defined as $\neg\mathsf{R}(\varphi, \neg\psi)$.
- $@_n\langle\mathsf{R}\rangle(\varphi, m)$, which is read as "$m$ is one of n's most reliable agents being φ."

Example 2. Continuing from Example 1. We use **a**, **b** and **c** as agent nominals for logicians a, b and the editor c, respectively. Let us denote "I solved an open problem no.7" by p. Then $@_{\mathbf{a}}\mathsf{B}p$ and $@_{\mathbf{b}}\mathsf{B}p$ mean that each of logicians a and b believes that she solved an open problem no.7. Moreover, $@_{\mathbf{c}}\mathsf{B}@_{\mathbf{a}}\mathsf{B}p$ is read as "c believes that a believes that she solved an open problem no.7" and $@_{\mathbf{c}}\neg\mathsf{B}@_{\mathbf{a}}\mathsf{B}p$ is read as "c does not believes that a believes that she solved an open problem no.7."

Let us provide a semantics based on our PR-structure as follows. Let $F = (W, A, (\sigma_a)_{a\in A}, (\rho_w)_{w\in W})$ be a PR-structure. A valuation V is defined to be a mapping sending a propositional atom p, a state nominal i, and an agent nominal n to a subset of $W \times A$ such that $V(i)$ is a subset of the form $\{ w \} \times A$ and $V(n)$ is a subset of the form $W \times \{ a \}$. Given a valuation V, we denote $V(i) = \{ \underline{i} \} \times A$ and $V(n) = W \times \{ \underline{n} \}$. This definition presupposes that an interpretation of a state nominal i and an agent nominal n is *rigid* for the agents A and the states W, respectively. That is, for example, a syntactic name of an agent does not change through states. A PR-model M is a pair of a PR-frame and a valuation. Then the satisfaction relation $M, (w, a) \models \varphi$ (read "agent a satisfies φ at state w in M" or "(w, a) forces φ in M") is inductively defined as:

$$\begin{array}{ll}
M,(w,a) \models p & \text{iff } (w,a) \in V(p), \\
M,(w,a) \models i & \text{iff } w = \underline{i}, \\
M,(w,a) \models n & \text{iff } a = \underline{n}, \\
M,(w,a) \models \neg\varphi & \text{iff } M,(w,a) \not\models \varphi, \\
M,(w,a) \models \varphi \wedge \psi & \text{iff } M,(w,a) \models \varphi \text{ and } M,(w,a) \models \psi, \\
M,(w,a) \models @_i\varphi & \text{iff } M,(\underline{i},a) \models \varphi, \\
M,(w,a) \models @_n\varphi & \text{iff } M,(w,\underline{n}) \models \varphi, \\
M,(w,a) \models \mathsf{B}(\varphi,\psi) & \text{iff } \mathrm{Min}_{\sigma_a(w)}([\![\varphi]\!]_{(\cdot,a)}) \cap [\![\varphi]\!]_{(\cdot,a)} \subseteq [\![\psi]\!]_{(\cdot,a)}, \\
M,(w,a) \models \mathsf{R}(\varphi,\psi) & \text{iff } \mathrm{Min}_{\rho_w(a)}([\![\varphi]\!]_{(w,\cdot)}) \cap [\![\varphi]\!]_{(w,\cdot)} \subseteq [\![\psi]\!]_{(w,\cdot)},
\end{array}$$

where $[\![\varphi]\!]_{(\cdot,a)} := \{ w \in W \mid M,(w,a) \models \varphi \}$ (the set of states where agent a satisfy φ) and $[\![\varphi]\!]_{(w,\cdot)} := \{ a \in A \mid M,(w,a) \models \varphi \}$ (the set of agents which satisfy φ at state w). Let us say that φ is *PR-valid* if $M,(w,a) \models \varphi$ for every PR-model M and every pair (w,a) in M.

The rough idea of axiomatizing the set of all the PR-valid formulas is that we combine two axiomatizations (one is for the fragment for states and the other is for the fragment for agents) of hybrid counterfactual logic [8] with an inference rule called **CBG**, which characterizes the limit assumption (see [8, Theorem 3]), and then we add to such combination the following five interaction axioms:

$$\begin{aligned} @_i n &\leftrightarrow n, \\ @_n i &\leftrightarrow i, \\ @_i @_n p &\leftrightarrow @_n @_i p, \\ @_n \mathsf{B}(p,q) &\leftrightarrow @_n \mathsf{B}(@_n p, @_n q), \\ @_i \mathsf{R}(p,q) &\leftrightarrow @_i \mathsf{R}(@_i p, @_i q). \end{aligned}$$

Then, we can obtain the following, though the space is limited and so we omit the proof of it (cf. [8, 9]).

Theorem 1. *There is a sound and complete axiomatization of the set of all the PR-valid formulas of the static syntax $\mathcal{L}$.*

4 Adding Dynamisms for Changing Belief and Reliability

This section introduces two dynamic operators for changing an agent's belief and reliability and describes how a reliability change may influence a belief change.

First, we propose a dynamic operator for states, called an *informing action operator* denoted by:

$$[\varphi!_{(m,n)}]\psi,$$

whose reading is "after m informs n that m satisfies φ, ψ holds." So, a sender agent m makes an informing action to a recipient agent n, but we require that, if a sender m is one of the most reliable agents being φ to a recipient n, then the informing action will succeed, and otherwise, the action will fail. Given a PR-model $M = (W, A, (\sigma_a)_{a\in A}, (\rho_w)_{w\in W}, V)$, we define the semantics of the informing action operator as follows:

$$M,(v,b) \models [\varphi!_{(m,n)}]\psi \text{ iff } M^{\varphi!_{(m,n)}},(v,b) \models \psi,$$

where $M^{\varphi!_{(m,n)}} = (W, A, (\sigma'_a)_{a\in A}, (\rho_w)_{w\in W}, V)$ and $\sigma'_a : W \to \wp\wp(W)$ is defined as follows: for all states $w \in W$,

$$\sigma'_a(w) = \begin{cases} \{ S \cap [\![\varphi]\!]_{(\cdot,\underline{m})} \mid S \in \sigma_{\underline{n}}(w) \} & \text{if } a = \underline{n} \text{ and } M,(w,\underline{n}) \models \langle \mathsf{R} \rangle(\varphi, m); \\ \sigma_a(w) & \text{otherwise,} \end{cases}$$

where "$M, (w, \underline{n}) \models \langle \mathsf{R} \rangle(\varphi, m)$" means that $\underline{m}$ is one of the most reliable φ-agents to $\underline{n}$. [2] It is remarked that (W, σ'_a) is still a system of spheres for all $a \in A$. When the informing action succeeds, we *cut* spheres $\sigma_{\underline{n}}(w)$ by the set $[\![\varphi]\!]_{(\cdot,\underline{m})}$, which is the content conveyed by the informing action, i.e., agent m satisfies φ. We may regard this informing action is a generalization of an update by link-cutting by Lewis [6] and Yamada [15].

Example 3. Continuing from Example 2. We may regard an action of each logician's sending an email on the open problem to the editor c as an informing action. There actions are formalized as $[\mathsf{B}p!_{(\mathbf{a},\mathbf{c})}]$ and $[\mathsf{B}p!_{(\mathbf{b},\mathbf{c})}]$, where $\mathsf{B}p$ is read as "I believe that I solved an open problem no.7." Then the editor c's beliefs after these informing actions of Example 1 are described as

$$[\mathsf{B}p!_{(\mathbf{a},\mathbf{c})}][\mathsf{B}p!_{(\mathbf{b},\mathbf{c})}](@_{\mathbf{c}}\mathsf{B}@_{\mathbf{a}}\mathsf{B}p \wedge @_{\mathbf{c}}\neg\mathsf{B}@_{\mathbf{b}}\mathsf{B}p),$$

where the informing action by a succeeds but the action by b fails.

Second, we propose a dynamic operator for agents, which is a generalization of *radical upgrades* in [13] to a system of spheres. Our version of radical upgrade operator is denoted by

$$[i : \varphi \Uparrow^n]\psi,$$

whose reading is "after n *upgrades* all the agents being ψ at state i, ψ holds." The reader may wonder how the corresponding downgrading operation is defined. This is done with the help of the negation symbol as follows:

$$[i : \varphi \Downarrow^n]\psi := [i : \neg\varphi \Uparrow^n]\psi,$$

whose reading is "after n *downgrades* all the agents being ψ at state i, ψ holds." Given a PR-model $M = (W, A, (\sigma_a)_{a\in A}, (\rho_w)_{w\in W}, V)$, the semantics of the upgrade operator is defined as:

$$M, (v, b) \models [i : \varphi \Uparrow^n]\psi \text{ iff } M^{i:\varphi\Uparrow^n}, (v, b) \models \psi,$$

where $M^{i:\varphi\Uparrow^n} = (W, A, (\sigma_a)_{a\in A}, (\rho'_w)_{w\in W}, V)$ and $\rho'_w : A \to \wp\wp(A)$ is defined as follows: for all agents $a \in A$,

$$\rho'_w(a) := \begin{cases} \{\, P \cap [\![\varphi]\!]_{(\underline{i},\cdot)}, (\bigcup \rho_w(a) \cap [\![\varphi]\!]_{(\underline{i},\cdot)}) \cup P \mid P \in \rho_w(a) \,\} & \text{if } a = \underline{n}; \\ \rho_w(a) & \text{if } a \neq \underline{n}. \end{cases}$$

We note that (A, ρ'_w) is still a system of spheres for every $w \in W$.

[2] If we want to assure that the set of the most reliable φ-agents to $\underline{n}$ is non-empty at the current state, then we might modify the semantics of $[\varphi!_{(m,n)}]\psi$ as follows:

$$M, (v, b) \models [\varphi!_{(m,n)}]\psi \text{ iff } \bigcup \rho_v(\underline{n}) \cap [\![\varphi]\!]_{(v,\cdot)} \neq \emptyset \text{ implies } M^{\varphi!_{(m,n)}}, (v, b) \models \psi.$$

We note that $\bigcup \rho_v(\underline{n}) \cap [\![\varphi]\!]_{(v,\cdot)} \neq \emptyset$ iff $M, (v, b) \models @_n\neg\mathsf{R}(\varphi, \bot)$.

Example 4. Continuing from Example 1. Now let us suppose that b's proof of the open problem no.7 is written clearly but a's proof is not so and that the editor c upgrades b with respect to the property: "x's proof of the open problem is written clearly at state i," which is formalized as a radical upgrade $[i : q \Uparrow^{\mathbf{c}}]$ where q means "my proof of the open problem is written clearly." Then b will be more reliable to c than a. Now the informing action $[\mathsf{B}p!_{(\mathbf{b},\mathbf{c})}]$ by b should be successful. Such situation is now formalized as:

$$[\mathsf{B}p!_{(\mathbf{b},\mathbf{c})}]@_{\mathbf{c}}\neg\mathsf{B}@_{\mathbf{b}}\mathsf{B}p \wedge [i : q \Uparrow^{\mathbf{c}}][\mathsf{B}p!_{(\mathbf{b},\mathbf{c})}]@_{\mathbf{c}}\mathsf{B}@_{\mathbf{b}}\mathsf{B}p.$$

This formalization demonstrates how a reliability change may influence a belief change.

Let us denote by $\mathcal{L}^+$ the syntax extending the static syntax $\mathcal{L}$ with dynamic operators $[\varphi!_{(m,n)}]$ and $[i : \varphi \Uparrow^n]$ defined above. Since we can provide *recursion axioms* (cf. [14]) for $[\varphi!_{(m,n)}]$ and $[i : \varphi \Uparrow^n]$ which enable us to eliminate these dynamic operators with keeping logical equivalence, we can also axiomatize all the PR-valid formulas of the extended syntax $\mathcal{L}^+$:

Theorem 2. *There is a sound and complete axiomatization of the set of all the PR-valid formulas of the extended syntax $\mathcal{L}^+$.*

Finally, we would like to illustrate on how a belief change may influence a reliability structure of an agent. The following example is taken from [7].

Example 5. Let us regard a as an owner of a sushi restaurant and b, c as agents and consider the following simple scenario.

- b says to c that fish of sushi restaurant a is not so fresh.
- Then c downgrades sushi restaurant a.

Let us use $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ as agent nominals for a, b and c, respectively. Then the first item can be formalized by an informing action $[@_{\mathbf{a}}\neg\mathsf{Fresh}!_{(\mathbf{b},\mathbf{c})}]$ where "Fresh" formalizes "sushi of my restaurant is not so fresh." To formalize the second item, we need to use the property "c believes that x's sushi is not so fresh" for a radical *downgrade* operation. In order to express such *de re* property, however, we need to employ the downarrow binder $\downarrow n.$ of hybrid logic. The downarrow binder $\downarrow n.$ allows us to name the current agent by an agent nominal n and the semantics of $\downarrow n.$ is given as:

$$M, (w, a) \downarrow n.\varphi \text{ iff } M[n := a], (w, a) \models \varphi,$$

where $M[n := a]$ is the same model as M except that a valuation of $M[x := a]$ sends n to $W \times \{ a \}$. [3] Then the de re property above is formalized as

$$[i :\downarrow n.@_{\mathbf{c}}\mathsf{B}@_n\neg\mathsf{Fresh} \Downarrow^{\mathbf{c}}].$$

[3] To axiomatize the static logic of the expanded syntax $\mathcal{L}$ with the downarrow binder $\downarrow$, it suffices for us to add the following axiom (cf. [3, p.304]) by the technique of *local definability*: $@_m(\downarrow n.\varphi \leftrightarrow \varphi[m/n])$, where the notation $[m/n]$ is the result of substituting the nominal m for all occurences of the nominal n.

After these two updates, let us suppose that a is not one of the most reliable restaurants whose sushi are so fresh. Then this situation is formalized as:

$$[@_{\mathbf{a}}\neg\mathsf{Fresh}!_{(\mathbf{b},\mathbf{c})}][i :\downarrow n.@_{\mathbf{c}}\mathsf{B}@_n\neg\mathsf{Fresh} \Downarrow^{\mathbf{c}}]\neg@_{\mathbf{c}}\langle\mathsf{R}\rangle(\mathsf{Fresh},\mathsf{a}).$$

References

1. C. Areces and B. ten Cate. Hybrid logics. In Patrick Blackburn, Johan van Benthem, and Frank Wolter, editors, *Handbook of Modal Logic*, pages 821–868. Elsevier, 2007.
2. P. Blackburn. Arthur Prior and hybrid logic. *Synthese*, 150(3):329–372, 2006.
3. P. Blackburn and B. ten Cate. Pure extensions, proof rules, and hybrid axiomatics. *Studia Logica*, 84(3):277–322, 2006.
4. J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, Cornell, 1962.
5. D. Lewis. *Counterfactuals*. Blackwell Publishing, 1973. 2nd edition was pubulished in 1986.
6. D. Lewis. A problem about permission. In E. Saarinen, R. Hilpinen, I. Niiniluoto, and M. Hintikka, editors, *Essays in Honour of Jaakko Hintikka*, pages 163–175. Reidel, Dordrecht, 1979.
7. F. Liu. Emergence of preferences from the logical perspective. Plenary Talk at 14th SAET Conference on Current Trends in Economics at Waseda University, August 2014.
8. K. Sano. Hybrid counterfactual logics: David Lewis meets Arthur Prior again. *Journal of Logic, Language and Information*, 18(4):515–539, 2009.
9. K. Sano. Axiomatizing hybrid products: How can we reason many-dimensionally in hybrid logic? *Journal of Applied Logic*, 8(4):459–474, December 2010.
10. K. Sano. Dynamified hybrid counterfactual logic. In *Logic and Interactive RAtionality: Yearbook 2010*, pages 27–48. ILLC, University of Amsterdam, 2011.
11. J. Seligman, F. Liu, and P. Girard. Logic in the community. In *Proceedings of the 4th Indian Conference on Logic and its Applications*, volume 6521 of *Lecture Notes in Computer Science*, pages 178–188, 2011.
12. J. Seligman, F. Liu, and P. Girard. Knowledge, friendship and social announcement. In Johan van Benthem and Fenrong Liu, editors, *Logic Across the University: Foundations and Applications*, volume 47 of *Studies in Logic*, pages 445–469. College Publications, 2013.
13. J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17:157–182, 2007.
14. H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Springer, 2008.
15. T. Yamada. Acts of commanding and changing obligations. In *Proceedings of the Seventh International Workshop on Computational Logic in Multi-Agent Systems*, pages 1–19. Springer-Verlag, 2006.

Harsanyi-Sen Debate and Algebraic Difference Measurement

Satoru SUZUKI, PhD.
bxs05253@nifty.com

Faculty of Arts and Sciences, Komazawa University, JAPAN

SOCREAL 2016
Hokkaido University, Sapporo
29 October 2016

Contents of This Talk

Contents of This Talk

Harsanyi's Aggregation Theorem (1)

- Harsanyi (1953, 1955) develops expected utility theory of von Neumann and Morgenstern (1944) to provide two formalizations of weighted utilitarianism.
- Weymark (1991, 2005) refers to these results as Harsanyi's Aggregation and Impartial Observer Theorems.

Harsanyi's Aggregation Theorem (2)

- Weymark (2005) states the former in the following informal way:

  > *Individual and social preferences on the set of alternatives are assumed to satisfy the axioms of expected utility theory. Furthermore, two alternatives are socially indifferent if every individual is indifferent between them (Pareto Indifference). With these assumptions, if the preferences are represented by expected utility functions, then the social utility function is a linear function of the individual utility functions. Hence, alternatives are socially ranked according to a weighted utilitarian rule.*

Sen's Criticism

- Sen (1976) criticizes Harsanyi's Aggregation Theorem as follows:
- Von Neumann-Morgenstern expected utility theory is an ordinal theory and, therefore, any increasing transform of an expected utility function is a satisfactory representation of an individual's preference relation.
- However, weighted utilitarianism requires a cardinal theory of utility (i.e., theory of utility difference), and so Harsanyi is not justified in giving his theorems utilitarian interpretations.
- Sen's informal discussion of these issues is formalized by Weymark (1991).

Mongin's Theorem (1)

- Mongin (2002) proves a revised version of Harsanyi's Aggregation Theorem that is immune from Sen's criticism by introducing cardinal preference relations.
- He introduces the two sorts of primitive preference relations: (i) cardinal preference relations (i.e., preference intensity relations), (ii) individual and social preference relations.
- Standard ordinal preference relations are defined by cardinal preference relations.
- He gives conditions under which cardinal preference relations can be represented by utility differences, and provides conditions (Conn. 1) and (Conn. 2) that connect individual and social preference relations with cardinal preference relations.
- Under these conditions, he proves the aggregation theorem based on cardinal utilities.

Mongin's Theorem (2)

- Mongin (2002, p.14) says on (Conn. 2) as follows:

  > *(Conn. 2) makes the following demand: y is midway between x and z in cardinal preference terms if and only if y is equivalent in ordinal preference terms to the half-half lottery between x and z.*

- Mongin (2002, p.15) himself, however, questions the validity of (Conn. 2):

  > *More telling reasons for questioning (Conn. 2) come from the normative intuition. It seems to be entirely acceptable for an individual to have a decreasing intensity of preference for money, and nonetheless to be risk-prone, in the ordinary sense of preferring lotteries to their actuarial values; and conversely.*

Harvey's Theorem

- Harvey (1999) proves a revised version of Harsanyi's Aggregation Theorem that has preference intensity relations as primitive that can be represented by utility differences but does not depend on (Conn. 2).
- Harvey's theorem, however, does not justice Harsanyi's insight, which intends to base weighted utilitarianism on individual preferences under risk.
- For Harvey's theorem concerns only with preferences under certainty.

Aim of This Talk

- The aim of this talk is to show that the aggregation theorem that has as primitive individual and social preference intensity relations under risk and does not depend on (Conn. 2) can imply weighted utilitarianism.
- We show it by imposing an algebraic difference measurement structure on preference intensity relations under risk.

Contents of This Talk

Preliminary Concepts (1)

We define some preliminary concepts as follows:

Definition (Outcome, Lottery, etc. (1))

- $O := \{o_1, \ldots, o_n\}$, where $n \geq 2$, is a nonempty finite set of outcomes.
- A lottery $p := (p_1, \ldots, p_n)$ specifies, for any $o_i \in O$, the probability of obtaining o_i.
- The set of all lotteries on O, in symbols $\mathscr{L}(O)$, is the $(n-1)$-dimensional unit simplex, i.e., the set of all $p \in \mathscr{L}(O)$ in which $\sum_{i=1}^{n} p_i = 1$.

Preliminary Concepts (2)

Definition (Outcome, Lottery, etc. (2))

- Letting $e^{(i)} := (e_1^{(i)}, \ldots, e_n^{(i)})$, where $e_j^{(i)} = 1$ if $j = i$ and $e_j^{(i)} = 0$ otherwise, the outcome o_i is equivalent to the lottery $e^{(i)}$ that assigns probability 1 to o_i.
- $\succsim$ is a binary relation on $\mathscr{L}(O)$.
- $p \succ q := q \not\succsim p$ for any $p, q \in \mathscr{L}(O)$.
- $p \sim q := p \succsim q$ and $q \succsim p$ for any $p, q \in \mathscr{L}(O)$.
- For any $p, q \in \mathscr{L}(O)$ and $\alpha \in [0,1]$, the convex combination $\alpha p + (1-\alpha)q \in \mathscr{L}(O)$, because $\alpha p + (1-\alpha)q$ is simply a vector in the $(n-1)$-dimensional unit simplex.

Preliminary Concept (3)

Definition (Weak Order, Continuity, and Independence)

Weak Order $\succsim$ is an weak order (connected and transitive) on $\mathscr{L}(O)$.

Continuity For any $p, q, r \in \mathscr{L}(O)$, if $p \succ q \succ r$, there exists $\alpha \in (0,1)$ such that $\alpha p + (1-\alpha)r \sim q$.

Independence (i) For any $p, q \in \mathscr{L}(O)$ and any $\alpha \in (0,1)$, if $p \succ q$, then $p \succ \alpha p + (1-\alpha)q$. (ii) For any $p, q \in \mathscr{L}(O)$ and any $\alpha \in (0,1)$, if $p \succ q$, then $\alpha p + (1-\alpha)r \succ \alpha q + (1-\alpha)r$. (iii) For any $p, q \in \mathscr{L}(O)$ and any $\alpha \in (0,1)$, if $p \sim q$, then $p \sim \alpha p + (1-\alpha)q$. (iv) For any $p, q \in \mathscr{L}(O)$ and any $\alpha \in (0,1)$, if $p \sim q$, then $\alpha p + (1-\alpha)r \sim \alpha q + (1-\alpha)r$.

Von Neumann-Morgenstern Utility Function

We define a von Neumann-Morgenstern utility function as follows:

Definition (Von Neumann-Morgenstern Utility Function)

A function $U : \mathscr{L}(O) \to \mathbb{R}$ is a von Neumann-Morgenstern utility function if for any $p \in \mathscr{L}(O)$,

$$U(p) = \sum_{i=1}^{n} p_i U(e^{(i)}).$$

Pareto Indifference

- Suppose that there are n individuals.
- Individual $i (= 1, \ldots, n)$ has a preference relation $\succsim_i$ on $\mathscr{L}(O)$.
- There is also a social preference relation $\succsim$ on $\mathscr{L}(O)$.
- The only link between the individual and social preference is provided by Pareto Indifference:

Definition (Pareto Indifference)

For any $p, q \in \mathscr{L}(O)$, if $p \sim_i q$ for any $i (= 1, \ldots, n)$, then $p \sim q$.

Harsanyi's Aggregation Theorem

The following is a formal statement of Harsanyi's Aggregation Theorem:

Theorem (Harsany's Aggregation Theorem)

Suppose that $\succsim_i$ $(i = 1, \ldots, n)$ and $\succsim$ are binary relation on $\mathscr{L}(O)$ that satisfy Weak Order, Continuity, and Independence, and also suppose that $\succsim_i$ and $\succsim$ satisfy Pareto Indifference. Let $U_i : \mathscr{L}(O) \to \mathbb{R}$ be a von Neumann-Morgenstern utility functional representation of $\succsim_i$ $(i = 1, \ldots, n)$ and $U : \mathscr{L}(O) \to \mathbb{R}$ a von Neumann-Morgenstern utility functional representation of $\succsim$. Then there exist $\alpha_i (i = 1, \ldots, n) \in \mathbb{R}$ and $\beta \in \mathbb{R}$ such that for any $p \in \mathscr{L}(O)$,

$$U(p) = \sum_{i=1}^{n} \alpha_i U_i(p) + \beta.$$

Maximum Diversity

- The assumptions of Harsanyi's Aggregation Theorem do not guarantee the unique existence of $(\alpha_1, \ldots, \alpha_n, \beta)$.
- To guarantee this, we suppose that each individual preference relation satisfies the following condition:

Definition (Maximum Diversity)

For any $i (= 1, \ldots, n)$, there exist $p_i, q_i \in \mathscr{L}(O)$ such that $p_i \nsim_i q_i$ and $p_i \sim_j q_i$ for any $j (\neq i)$.

Strict Pareto

- The assumptions of Harsanyi's Aggregation Theorem do not guarantee the positivity of $\alpha_1, \ldots, \alpha_n$.
- The following condition in conjunction with Pareto Indifference can guarantee this:

Definition (Strict Pareto)

For any $p, q \in \mathscr{L}(O)$, if, for some $j (= 1, \ldots, n)$ and for any $i (\neq j)$, $p \succsim_i q$ and $p \succ_j q$, then $p \succ q$.

First Improved Aggregation Theorem

The following is the first improved version of Harsanyi's Aggregation Theorem that guarantees both the unique existence of $(\alpha_1, \ldots, \alpha_n, \beta)$ and the positivity of $\alpha_1, \ldots, \alpha_n$.:

Theorem (First Improved Aggregation Theorem)

Suppose that $\succsim_i$ $(i = 1, \ldots, n)$ and $\succsim$ are binary relation on $\mathscr{L}(O)$ that satisfy Weak Order, Continuity, Independence, and Maximum Diversity, and also suppose that $\succsim_i$ and $\succsim$ satisfy Pareto Indifference and Strict Pareto. Let $U_i : \mathscr{L}(O) \to \mathbb{R}$ be a von Neumann-Morgenstern utility functional representation of $\succsim_i$ $(i = 1, \ldots, n)$ and $U : \mathscr{L}(O) \to \mathbb{R}$ a von Neumann-Morgenstern utility functional representation of $\succsim$. Then there uniquely exist $\alpha_i (i = 1, \ldots, n) \in \mathbb{R}^+$ and $\beta \in \mathbb{R}$ such that for any $p \in \mathscr{L}(O)$,

$$U(p) = \sum_{i=1}^{n} \alpha_i U_i(p) + \beta.$$

Weighted Utilitarianism w.r.t. $\succsim$

- We define Weighted Utilitarianism w.r.t. $\succsim$ as follows:

Definition (Weighted Utilitarianism w.r.t. $\succsim$)

$\succsim$ is said to satisfy Weighted Utilitarianism if there exists an weighted vector $\alpha := (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ for which for any admissible $\mathbf{U} := (U_1, \ldots, U_n)$,

$$(1)\ p \succsim q \text{ iff } \sum_{i=1}^{n} \alpha_i U_i(p) \geq \sum_{i=1}^{n} \alpha_i U_i(q),$$

for any $p, q \in \mathscr{L}(O)$.

- The conclusion Harsanyi intends to draw is that Weighted Utilitarianism follows from Harsanyi's Aggregation Theorem.

Contents of This Talk

Sen-Weymark's Argument (1)

We define increasing and positive affine transforms as follows:

Definition (Increasing and Positive Affine Transforms)

- $f : \mathbb{R} \to \mathbb{R}$ is an increasing transform if for any $x, y \in \mathbb{R}$, $f(x) \geq f(y)$ iff $x \geq y$.
- A utility function that is unique up to an increasing transform is said to be ordinal.
- $f : \mathbb{R} \to \mathbb{R}$ is a positive affine transform if there exist $\alpha \in \mathbb{R}^+$ and $\beta \in \mathbb{R}$ such that $f(x) = \alpha x + \beta$ for any $x \in \mathbb{R}$.
- A utility function that is unique up to a positive affine transform or any smaller set of transforms is said to be cardinal.

Sen-Weymark's Argument (2)

- We define co-cardinality as follows:

Definition (Co-Cardinality)

The n-tuple of positive affine transforms $F := (f_1, \ldots, f_n)$, where $f_i(x) = \gamma x + \delta_i$ for any $x \in \mathbb{R}$ for some $\gamma \in \mathbb{R}^+$ and $\delta_1, \ldots, \delta_n \in \mathbb{R}$, is called co-cardinal.

- According to Weymark (2005), the underlying reason for the problems identified by Sen is that in order for weighted utilitarianism to be meaningful, it must be possible to compare utility differences (gains and losses) both intrapersonally (guaranteed by cardinality) and interpersonally (guaranteed by co-cardinality).

Sen-Weymark's Argument (3)

- The need of difference comparability can be seen most clearly rewriting (1) as follows:

$$(2)\ p \succsim q \quad \text{iff} \quad \sum_{i=1}^{n} \alpha_i (U_i(p) - U_i(q)) \geq 0,$$

  for any $p, q \in \mathscr{L}(O)$.
- The utility difference sum in (2) does not change if the utility profile $\mathbf{U} := (U_1, \ldots, U_n)$ is replaced by the profile $\mathbf{V} = F \circ \mathbf{U} := (f_1 \circ U_1, \ldots, f_n \circ U_n)$ for some co-cardinal n-tuple of transform F.

Sen-Weymark's Argument (4)

- Let $\mathscr{U}^C$ denote the set of such profile of utility functions.
- The profiles in $\mathscr{U}^C$ are the only profiles that preserve the utility difference sum in (2).
- However, nothing in the version of expected utility theory that Harsanyi employed in his theorems rules out the use of non-affine increasing transform of U_i.
- So, because the set of admissible profiles is not always a subset of $\mathscr{U}^C$, Weighted Utilitarianism does not follows from (First Improved) Harsanyi's Aggregation Theorem.

Contents of This Talk

Algebraic Difference Structure

Krantz et al. (1971) define an algebraic difference structure as follows:

Definition (Algebraic Difference Structure)

Suppose $\mathscr{A}$ is a nonempty set and $\succsim^*$ a quaternary relation on $\mathscr{A}$, i.e., a binary relation on $\mathscr{A} \times \mathscr{A}$. $(\mathscr{A} \times \mathscr{A}, \succsim^*)$ is an algebraic difference structure iff, for any a, b, c, d, a', b' and $c' \in \mathscr{A}$, the following five axioms are satisfied:

Weak Order* $\succsim^*$ is a weak order.
Sign Reversal* If $(a,b) \succsim^* (c,d)$, then $(d,c) \succsim^* (a,b)$.
Weak Monotonicity* If $(a,b) \succsim^* (a',b')$ and $(b,c) \succsim^* (b',c')$, then $(a,c) \succsim^* (a',c')$.
Solvability* If $(a,b) \succsim^* (c,d) \succsim^* (a,a)$, then there exist $d', d'' \in \mathscr{A}$ such that $(a,d') \sim^* (c,d) \sim^* (d'',b)$.
Archimedeanity* Every strictly bounded standard sequence is finite.

Interpretation of Solvability*

Remark (Interpretation of Solvability*)

- *In terms of Algebraic Difference Measurement Theorem below, $(a,b) \succsim^* (c,d)$ says that $f^*(a) - f^*(b) \geq f^*(c) - f^*(d)$. $(c,d) \succsim^* (a,a)$ says that $f^*(c) - f^*(d) \geq 0$. Solvability* says that if $f^*(a) - f^*(b) \geq f^*(c) - f^*(d) \geq 0$, then we can solve the equations*
$$f^*(a) - f^*(d') = f^*(c) - f^*(d)$$
and
$$f^*(d'') - f^*(b) = f^*(c) - f^*(d).$$
- *Solvability* is a structural (nonnecessary) condition.*

Interpretation of Archimedeanity*

Remark (Interpretation of Archimedeanity)

Archimedeanity corresponds to the Archimedean property of real numbers that any positive numbers are comparable, i.e., their ratio is finite.*

Algebraic Difference Measurement Theorem

Krantz et al. (1971) prove the following theorem:

Theorem (Algebraic Difference Measurement, Krantz et al.)

If $(\mathscr{A} \times \mathscr{A}, \succsim^*)$ is an algebraic difference structure, then there exists a real-valued function f^* on $\mathscr{A}$ such that, for any $a,b,c,d \in \mathscr{A}$,
$$(a,b) \succsim^* (c,d) \text{ iff } f^*(a) - f^*(b) \geq f^*(c) - f^*(d).$$
Moreover, f^* is unique up to a positive affine transform.

Continuity* (1)

- Suppose that there are n individuals.
- Individual $i(=1,\ldots,n)$ has a preference intensity relation $\succsim_i^*$ on $\mathscr{L}(O)$.
- For any $p,q,r,s \in \mathscr{L}(O)$, $(p,q) \succsim_i^* (r,s)$ is interpreted to mean that i prefers p to q more than i prefers r to s.
- There is also a social preference intensity relation $\succsim^*$ on $\mathscr{L}(O)$.
- For any $p,q,r,s \in \mathscr{L}(O)$, $(p,q) \succsim^* (r,s)$ is interpreted to mean that the society prefers p to q more than it prefers r to s.

Continuity* (2)

We define Continuity* as follows:

Definition (Continuity*)

(i) For any $p,q,r,s \in \mathscr{L}(O)$, if $(p,s) \succ^* (q,s) \succ^* (r,s)$, there exists $\alpha \in (0,1)$ such that $\alpha(p+(1-\alpha)r,s) \sim^* (q,s)$. (ii) For any $p,q,r,s \in \mathscr{L}(O)$, if $(s,p) \succ^* (s,q) \succ^* (s,r)$, there exists $\alpha \in (0,1)$ such that $\alpha(s,p+(1-\alpha)r) \sim^* (s,q)$.

Strict Pareto*, etc.

- Further, we define Strict Pareto* as follows:

Definition (Strict Pareto*)

(i) For any $p,q,r \in \mathscr{L}(O)$, if, for some $j(=1,\ldots,n)$ and for any $i(\neq j)$, $(p,r) \succsim_i^* (q,r)$ and $(p,r) \succ_j^* (q,r)$, then $(p,r) \succ^* (q,r)$. (ii) For any $p,q,r \in \mathscr{L}(O)$, if, for some $j(=1,\ldots,n)$ and for any $i(\neq j)$, $(r,p) \succsim_i^* (r,q)$ and $(r,p) \succ_j^* (r,q)$, then $(r,p) \succ^* (r,q)$.

- We can define Independence*, Maximum Diversity* and Pareto Indifference* in much the same way as Continuity* and Strict Pareto*.

Second Improved Aggregation Theorem (Main Theorem)

We can prove the following Second Improved Aggregation Theorem (Main Theorem) from Algebraic Difference Measurement Theorem and Definitions (Continuity*) and (Strict Pareto*) and so on:

Theorem (Second Improved Aggregation Theorem (Main Theorem))

Suppose that $\succsim_i^*$ $(i=1,\ldots,n)$ and $\succsim^*$ are quaternary relation on $\mathscr{L}(O)$ that satisfy Weak Order, Continuity*, Independence*, Maximum Diversity*, Sign Reversal*, Weak Monotonicity*, Solvability*, and Archimedeanity* and also suppose that $\succsim_i^*$ and $\succsim^*$ satisfy Pareto Indifference* and Strict Pareto*. Let $U_i^* : \mathscr{L}(O) \to \mathbb{R}$ be a von Neumann-Morgenstern utility functional representation of $\succsim_i^*$ $(i=1,\ldots,n)$ and $U^* : \mathscr{L}(O) \to \mathbb{R}$ a von Neumann-Morgenstern utility functional representation of $\succsim^*$. Then there uniquely exist $\alpha_i (i=1,\ldots,n) \in \mathbb{R}^+$ and $\beta \in \mathbb{R}$ such that for any $p \in \mathscr{L}(O)$,*

$$U^*(p) = \sum_{i=1}^{n} \alpha_i U_i^*(p) + \beta,$$

where, for any $a,b,c,d \in \mathscr{A}$, $(a,b) \succsim^* (c,d)$ iff $U^*(a) - U^*(b) \geq U^*(c) - U^*(d)$. Moreover, Both U_i^* and U^* are unique up to a positive affine transform.

Cardinality and Co-Cardinality

Corollary (Cardinality and Co-Cardinality)

Both U_i^* and U^* are cardinal. $F^* := (f_1^*,\ldots,f_n^*)$, where $f_i^*(U_i^*) = \gamma U_i^* + \delta_i$ for some $\gamma \in \mathbb{R}^+$ and $\delta_i \in \mathbb{R}$, is co-cardinal.

Remark (Intrapersonal and Interpersonal Comparability)

Cardinality of U_i^* and U^* makes it possible to compare utility differences intrapersonally. Co-cardinality makes it possible to compare utility differences interpersonally.

Weighted Utilitarianism w.r.t. $\succsim^*$

We define Weighted Utilitarianism w.r.t. $\succsim^*$ as follows:

Definition (Weighted Utilitarianism w.r.t. $\succsim^*$)

$\succsim^*$ is said to satisfy Weighted Utilitarianism if there exists an weighted vector $\alpha := (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ for which for any admissible $\mathbf{U}^* := (U_1^*, \ldots, U_n^*)$,

$$(p,q) \succsim^* (r,s) \text{ iff } \sum_{i=1}^{n}(\alpha_i U_i^*(p) - \alpha_i U_i^*(q)) \geq \sum_{i=1}^{n}(\alpha_i U_i^*(r) - \alpha_i U_i^*(s)).$$

for any $p, q, r, s \in \mathscr{L}(O)$.

The next corollary as a conclusion of this talk follows from Second Improved Aggregation Theorem and Definition (Weighted Utilitarianism w.r.t. $\succsim^*$):

Corollary (Weighted Utilitarianism)

$\succsim^*$ of Second Improved Aggregation Theorem satisfies Weighted Utilitarianism.

Contents of This Talk

Summary

- In this talk, we have shown that the aggregation theorem that has as primitive individual and social preference intensity relations under risk and does not depend on (Conn. 2) can imply weighted utilitarianism.
- We have shown it by imposing an algebraic difference measurement structure on preference intensity relations under risk.

Thank You for Your Attention!

A Nonmonotonic Sequent Calculus with Downward Conditional

Shuhei SHIMAMURA
shuhei.shimamura0803@gmail.com

Philosophical Motivations: Logical inferentialism claims that the meaning of a bit of logical vocabulary can be explained by specifying its inferential role, namely the rule according to which it should be used in the logical inference. My system[1] can be regarded as a product of such a logical inferentialist project. However, this system is also motivated by commitment to two distinctive philosophical ideas: Semantic inferentialism and logical expressivism, both of which are developed by Robert Brandom. Although these two ideas may not be as familiar to logicians as logical inferentialism, I believe that they shed new and interesting light on the characteristic role that the logical vocabulary plays in our inferential practice. Furthermore, they are the key to understand the two distinctive features of the system presented here, nonmonotonicity and the downward conditional, neither of which has drawn too much attention in literature.

First, semantic inferentialism can be understood as a radical generalization of logical inferentialism. Semantic inferentialists claim that not only the meaning of a bit of logical vocabulary, but also the meaning of a bit of non-logical vocabulary are determined by its inferential role, namely the way it ought to be used, in connection with other relevant expressions, in our everyday inferential practice.[2]

It is crucial that "inferential practice" here is not understood narrowly as the logical one, but more broadly as "the game of giving and asking for reasons" in general. In this use-theoretic approach to the meaning, an inference that contributes to determine the meaning of a given expression is called a *material* inference (in contrast to a *formal* inference), since the correctness of that inference does not depend on its logical form, but rather on the way relevant expressions appear in it. For example, "If a match is struck, then it lights" is an instance of such material inferences, since this inference is correct not because it has a certain logical form, but because "a match," "is struck," and "lights" are arranged in this particular manner in it. As exemplified in this instance, material inferences are often defeasible, and therefore nonmonotonic. Thus, for semantic inferentialists, who do not limit attention to the formal inference, monotonicity is no longer a feature of inference that can be assumed across the board.

In a word, for semantic inferentialists, the meaning of a given expression, whether it is

[1] This paper owes a great deal to collaborative work of the research group of Prof. Robert Brandom.

[2] Strictly speaking, this is what Brandom (1994) calls *strong* inferentialism.

logical or not, is determined by the role that it plays in the inferential practice broadly understood. At this stage, however, one may wonder on what basis semantic inferentialists count some bits of vocabulary as logical but others as non-logical, and what is the relationship between those logical and non-logical bits of vocabulary. Logical expressivism is an answer to such a demarcation problem concerning logicality.

According to logical expressivism, there are two essential criteria to demarcate the logical vocabulary from the non-logical one. First, the ability to use logical vocabulary can be *algorithmically elaborated* from the ability to use non-logical vocabulary. In other words, if one already knows how to use a set of non-logical vocabulary in the underlying material inferential practice, one need not acquire any extra ability in order to come to know how to use a set of logical vocabulary. Second, the distinctive role that bits of logical vocabulary play is to *express* or *explicate* those material-inferential rules that we implicitly follow when we use bits of non-logical vocabulary.[3] To put differently, logical vocabulary enables us to explicitly talk about, within the object language, what we implicitly follow.

These background philosophical commitments put three substantial constraints on the logical system pursued in this paper. First, a *material* consequence relation that (according to inferentialists) makes *non-logical* sentences mean what they mean should *not* be assumed to be *monotonic*. Second, the consequence relation between *logically complex* sentences must be determined by systematically extending this underlying nonmonotonic material consequence relation. The connective rules conducting this work demonstrates how the ability to use non-logical vocabulary can be algorithmically elaborated into the ability to use logical one. Third, for the inferential roles of logical connectives thus fixed to count as *explicating* the underlying material consequence relation, such extension of the underlying nonmonotonic material consequence relation must be (at least) conservative. Consequently, the extended consequence relation must also be nonmonotonic. Thus, logical inferentialist expressivists need a nonmonotonic logical system.

For logical expressivists, the paradigm of the logical (i.e., explicating) vocabulary is the conditional, since it enables us to *codify* (or "*talk about*"), within the object language, material consequences when a new proposition is added to a premise set (i.e. $\Gamma - \{A\} \mid\sim A \rightarrow B$ *iff* $\Gamma, A \mid\sim B$[4]). In this paper, I enrich the current inventory of logical vocabulary with a new type of conditional——what I call *the downward conditional* ($^{-}\!\rightarrow$). The expressive role of this new conditional is the mirror image of that of the regular conditional: The

[3] See Brandom (2008).

[4] One might wonder why I do not slightly strengthen this biconditional to make it equivalent to Deduction Theorem: $\Gamma \mid\sim A \rightarrow B$ *iff* $\Gamma, A \mid\sim B$. Somewhat surprisingly, however, in the presence of the downward conditional Deduction Theorem (especially its left-to-right direction) is not compatible with conservativeness of the system.

downward conditional enables us to *codify* ("*talk about*") material consequences when a proposition is *subtracted from* (as opposed to *added to*) a premise set (i.e., $\Gamma, A \mid\sim A^{-}\rightarrow B$ *iff* $\Gamma - \{A\} \mid\sim B$).[5][6]

<u>*Merits*</u>: It may seem that the downward conditional is a rather exotic logical connective from the standard truth-conditional semanticist viewpoint. However, from the logical expressivist inferentialist viewpoint, it is a natural counterpart of the regular conditional. Furthermore, there are at least two merits of having the downward conditional in our system. First, the downward conditional massively increases the expressive power of the system. Suppose Γ is given as a premise set. The regular conditional only allows us to talk about, within the object language, consequences of a premise set *(finitely) bigger* than Γ. With the downward conditional in hand, however, we become able to talk about consequences of an *arbitrary* premise set (finitely) reachable from Γ. In other words, the (logically extended) consequences of each premise set Γ (i.e., Con(Γ)) becomes *monadological* in that it encodes all the consequences of its (finite) neighborhood (i.e., Con(Γ') for any Γ' that is *finitely reachable* from Γ, where Γ' is *finitely reachable* from Γ *iff* both $\Gamma - \Gamma'$ and $\Gamma' - \Gamma$ are finite).

Apart from this logical expressivist inferentialist viewpoint, however, the downward conditional can also have wider appeal of its own. This unorthodox conditional enables us to formalize a special class of inferences that are made by *setting aside*, in contrast to *obtaining* or *assuming*, a certain set of knowledge. Inferences of this class play an essential role in several philosophical arguments. Cartesian Skepticism (Descartes 1641/2008) and The Veil of Ignorance (Rawls 1971) are two of the most prominent examples. A more casual instance can be found in everyday mind-reading practice including the false belief task, in which one is required to infer from the viewpoint of another who lacks a piece of knowledge that one already has. Note that such downward inferences cannot be easily assimilated to those inferences already expressible by the regular logical vocabulary. After all, setting aside a premise p is different from either assuming $\neg p$ or assuming $p \vee \neg p$.

<u>*The Logical System*</u>: As an inferentialist, I start with a material consequence relation, $\mid\sim_0$, holding for a base language, L_0. It is stipulated that $\mid\sim_0$ is reflexive (i.e., $\Gamma_0, p \mid\sim_0 p$ for any Γ_0

[5] The idea of the downward conditional and the related idea of Monadologicality (to be explained below) are originally suggested by Bob Brandom. My contributions mainly consist in the technical results I am presenting here.

[6] Similarly, one might wonder why I do not claim a slightly stronger biconditional: $\Gamma, A \mid\sim A^{-}\rightarrow B$ *iff* $\Gamma \mid\sim B$. The reason I avoid this is the same as stated in note 4. Such a stronger biconditional is incompatible with conservativeness.

$\subseteq L_0$ and $p \in L_0$), but Weakening does not hold due to nonmonotonicity of $|\sim_0$. My sequent calculus maps such $|\sim_0$ over L_0 onto $|\sim$ over the logically complex language, L. $|\sim$ preserves reflexivity (i.e., $\Gamma, A |\sim A$ for any $\Gamma \subseteq L$ and $A \in L$). $|\sim$ is also a conservative extension of $|\sim_0$ (i.e., $\Gamma_0 |\sim p$ *iff* $\Gamma_0 |\sim_0 p$ for any $\Gamma_0 \subseteq L_0$ and $p \in L_0$,), and therefore nonmonotonic. The logically extended language L includes the conditional ($\rightarrow$) and the downward conditional ($^{-}\!\rightarrow$), along with the other regular connectives. These paired conditionals jointly explicate, within the object language, an arbitrary consequence of an arbitrary premise set that is (finitely) reachable from a give premise set (i.e., for any $\Gamma \subseteq L$ and $A_1, \ldots, A_m, A_{m+1}, \ldots, A_n \in L$, $\Gamma |\sim (A_1 {}^{-}\!\rightarrow \ldots {}^{-}\!\rightarrow (A_{m+1} \rightarrow \ldots \rightarrow (A_n \rightarrow B) \ldots) \ldots)$ *iff* $\Gamma - \{A_1, \ldots, A_m\}, A_{m+1}, \ldots, A_n |\sim B$, where $A_1, \ldots, A_m \in \Gamma$ and $A_{m+1}, \ldots, A_n \notin \Gamma$).

References

Brandom, R. B. (1994), *Making it Explicit*, Cambridge, MA: Harvard University Press.

Brandom, R. B. (2008), *Between Saying and Doing: Towards an Analytic Pragmatism*, Oxford: Oxford University Press.

Descartes, R. (1641/2008), *Meditations on First Philosophy: With Selections from the Objections and Replies*, trans. Michael Moriarty, Oxford: Oxford University Press.

Rawls, J. (1971), *A Theory of Justice*, Cambridge, MA: Harvard University Press.

Reinterpretation of deontic logic in the light of logical pragmatics

Berislav Žarnić

Research Centre for Logic, Epistemology, and Philosophy of Science
Faculty of Humanities and Social Sciences, University of Split, Croatia

Among many famous philosophers, like G.W. Leibniz (1646–1716) or A. Meinong (1853–1920), who dealt with formal problems of normativity one figure stands out as the proper founder of the research field. It was Georg Henrik von Wright (1916–2003) who first gave the name 'deontic logic' and systematic logical foundations to the formal study of normativity (von Wright, 1951) and whose fruitful contributions to the field spanned the interval of half a century; an overview of his last position is given in (von Wright, 1999). Many hard philosophical problems and paradoxes have aroused during the development of deontic logic. These facts lead Von Wright towards critical and even sceptical view on the very possibility of deontic logic, but, as the quote below vividly shows, the way out of paradoxes lies on the side of logical pragmatics, i.e., in the reinterpretation of deontic logic, as the study of the *use* of language in "rational norm-giving activity".

> Deontic logic, one could also say, is neither a logic of norms nor a logic of norm-propositions but a study of conditions which must be satisfied in rational norm-giving activity. It is strict *logic* because the conditions which it lays down are derived from *logical* relations between states in the ideal worlds which normative codes envisage. (von Wright, 1993, 111)

Von Wright's reinterpretation of deontic logic developed gradually and has introduced a number of important conceptual distinctions and theses, among which the following stand out: the distinction between prescriptive and descriptive use of deontic sentences; the thesis that relation between permission and absence of prohibition is not conceptual but normative in character, and this normative relation is one among other "perfection properties" of the normative system, the norm-giver (in the norm-giving activity by which the normative system is produced) ought to achieve perfection properties of the system. Some of these theses are summarized in Figure 1. In spite of Von Wright's highest authority in deontic logic, a complete formal explication for his reinterpretation has not been given as yet. Here

by the 'formal explication' is meant a model in terms of which definenda are given.

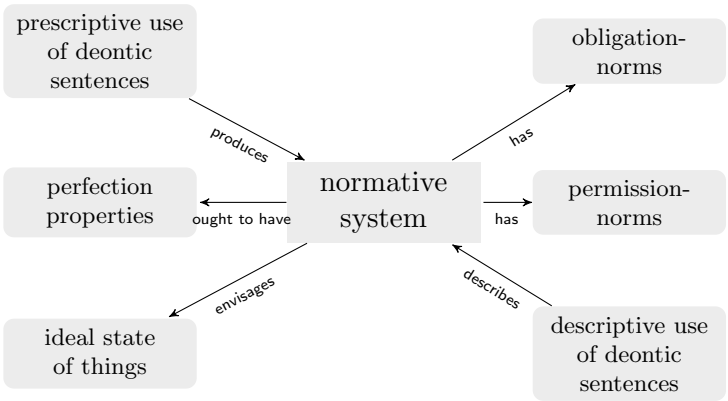


Figure 1: A map depicting a part of the "conceptual space" of norm-giving activity. One the same deontic sentence can be used prescriptively and descriptively. In prescriptive use the norm-giver ought to achieve perfection properties. In descriptive use an observer describes a real and possibly imperfect system.

The "one set model" for deontic concepts has been adopted by many authors thanks to its simplicity. Namely, 'it is obligatory that φ', $\mathrm{O}\varphi$, is modelled as 'sentence φ belongs to the set $\mathcal{N}$ of norms', $\ulcorner\varphi\urcorner \in \mathcal{N}$. The other definienda are obtained in the similar way: permission, $\mathrm{P}\varphi$, is modelled as non-membership of the contradictory content, $\ulcorner\neg\varphi\urcorner \notin \mathcal{N}$; and prohibition, $\mathrm{F}\varphi$, is just an obligation with contradictory content, $\mathrm{O}\neg\varphi$. In this model the relation between obligation and permission is conceptual: $\neg\mathrm{P}\neg\varphi$ is modelled as $\neg\ulcorner\varphi\urcorner \notin \mathcal{N}$, which is equivalent to $\ulcorner\varphi\urcorner \in \mathcal{N}$, the translation of $\mathrm{O}\varphi$.

In Von Wright's reinterpretation of deontic logic the distinction between real and perfect normative systems plays an important role. It is impossible to give a description of an imperfect normative system with a mismatch between obligations and permissions within the "one set model". For example, a system $\mathcal{N}$ based on $\mathrm{O}\varphi$ and $\mathrm{P}\neg\varphi$ is an impossible object since it requires $\ulcorner\varphi\urcorner \in \mathcal{N}$ and $\ulcorner\varphi\urcorner \notin \mathcal{N}$. Therefore, a more complex model must be introduced in order to enable the description of the difference between real and perfect systems.

A "two sets model" has been introduced in (Hansen, 2014), but the solution provided is not fully adequate for the formal explication of Von Wright's reinterpretation of deontic logic. Let $\mathcal{T}$ be the set of explicitly promulgated norms, and let function n deliver contradictory propositions as follows: $n(\varphi) = \neg\varphi$ and $n(\neg\varphi) = \varphi$. Relying on the two sets model introduced in (Žarnić, 2015), the thesis of this paper is that the appropriate model for the formal explication of Von Wright's reinterpretation is given by the pair $\langle\mathcal{N}, \overline{\mathcal{N}}\rangle$, where set $\mathcal{N} = \{\ulcorner\varphi\urcorner \mid \mathrm{O}\varphi \in \mathcal{T}\}$ has contents of obligation-

norms (with prohibitions treated as obligations with contradictory content), while set $\overline{\mathcal{N}} = \{\ulcorner n(\varphi) \urcorner \mid \mathrm{P}\varphi \in \mathcal{T}\}$ has contradictory contents of permission norms, and is called 'permission norm counter-set'.

The acceptability of this model can be checked against Von Wright's hypothesis that standard deontic logic depicts some perfection properties of a normative system. If the two sets model is adequate, then the translation of a valid formula of standard deontic logic will typically result in the description of a perfection property of the normative system. Given that there are two sets, there are three kinds of perfection properties: perfection properties for the each set and the perfection properties of the relation between the sets. Therefore, the three translation functions are needed; they have been introduced in (Žarnić, 2016). Consider as an example axiom schema (D), characteristic of deontic logic: $\mathrm{O}\varphi \to \mathrm{P}\varphi$. Translations show that the three perfection properties characterized by (D) axiom schema are: (i) consistency of obligation norm set, $\ulcorner\varphi\urcorner \in \mathcal{N} \to \ulcorner\neg\varphi\urcorner \notin \mathcal{N}$; (ii) completeness of permission norm counter set, $\ulcorner\varphi\urcorner \in \overline{\mathcal{N}} \vee \ulcorner\neg\varphi\urcorner \in \overline{\mathcal{N}}$; (iii) the relational property defined by $\ulcorner\varphi\urcorner \in \mathcal{N} \to \ulcorner\neg\varphi\urcorner \in \overline{\mathcal{N}}$. The translations can be applied in the philosophical analysis. Suppose that we want to add the "free choice permission" axiom: $\mathrm{P}(\varphi \vee \psi) \to (\mathrm{P}\varphi \wedge \mathrm{P}\psi)$. If the supposed axiom is sound, then it describes perfection properties. The application of the translation function which delivers claims on properties of the obligation norm set, under indubitable assumption that closure under equivalence is its perfection property, gives the following property: $(\ulcorner\varphi\urcorner \in \mathcal{N} \vee \ulcorner\psi\urcorner \in \mathcal{N}) \to \ulcorner\varphi \wedge \psi\urcorner \in \mathcal{N}$. The possession of this property makes inconsistent any non-empty set, and so the supposed axiom must be rejected. The philosophical consequence is that the free choice permission does not belong to the statics of normativity but rather describes the effects of an retractive act.

Although no explicit mention of pragmatics turn can be found in Von Wright's later works on deontic logic, this characterization is appropriate since the use and users of language and language-constructions are taken into the picture. The textual core of a normative system can come into existence thanks to the prescriptive use of language, what has been explicated here as the production of set $\mathcal{T}$ of deontic sentences. The logical properties of real normative systems can be described using the language of the "logic of norm-propositions", what has been explicated here by translation to claims on membership in $\mathcal{N}$ and $\overline{\mathcal{N}}$. Some logical properties are "perfection-properties" of a normative-system, such as the consistency of obligation norm set and the completeness of permission norm counter-set. The absence of a certain perfection-property does not deprive a normative system of its normative force. For example, textual core $\mathcal{T} = \{\mathrm{O}\varphi, \mathrm{P}\neg\varphi\}$ in spite of its imperfection still manages to define the normative system $\langle\{\ulcorner\varphi\urcorner\}, \{\ulcorner\varphi\urcorner\}\rangle$.

In the prescriptive use of language the norm-giver ought to achieve some perfection properties of the normative system. Thus, there are two

types of *oughts*: the ought resulting from the norm-giving activity, and the **ought** to which the norm-giving activity is subordinated. According to Von Wright, deontic logic is a study of logical perfection properties; properties the achievement of which fulfils rationality conditions of norm-giving activity. The approach can be generalized so to include other roles, such as the role of norm-recipient, and other norm-related activities, such as normative reasoning.

Logic has sometimes been understood as the ethics of thinking. Von Wright's reinterpretation of deontic logic prompts us to understood logic also as the ethics of language use. In understanding deontic logic the perspectives of different social roles of should be taken into account as well as the purpose of norm giving activity. In this way deontic logic ceases to be a "zero-actor logic" and becomes the logic of language use which requires the presence of "users". This fact redefines deontic logic as a research which necessarily includes the stance of logical pragmatics.[1]

References

Hansen, J. (2014). Reasoning about permission and obligation. In O. S. Hansson (Ed.), *David Makinson on Classical Methods for Non-Classical Problems*, pp. 287–333. Dordrecht: Springer Netherlands.

von Wright, G. H. (1951). Deontic logic. *Mind 60*(237), 1–15.

von Wright, G. H. (1993). *The Tree of Knowledge and Other Essays*, Chapter A Pilgrim's Progress, pp. 103–113. Leiden: Brill.

von Wright, G. H. (1999). Deontic logic — as I see it. In P. McNamara and H. Prakken (Eds.), *Norms, Logics and Information Systems: New Studies in Deontic Logic and Computer Science*, pp. 15–25. Amsterdam: IOS Press.

Žarnić, B. (2015). A social pragmatic view on the concept of normative consistency. *European Journal of Analytic Philosophy 11*(2), 56–78.

Žarnić, B. (2016). Deontic logic as a study of conditions of rationality in norm-related activities. In O. Roy, A. Tamminga, and M. Willer (Eds.), *Deontic Logic and Normative Systems. 13th International Conference, DEON 2016*, pp. 272–287. London: College Publications.

[1]The author acknowledges support from Croatian Science Foundation (HRZZ) within the research project: LogiCCom–Logic, Concepts, and Communication, and from the organizers of the SOCREAL2016–4th International Workshop on Philosophy and Logic of Social Reality.

Justification of Actions and Shared Belief Revisions

Yasuo Nakayama
Graduate School of Human Sciences, Osaka University

When is an action of a person justified in a group? This is a central question in this presentation. Nakayama (2010, 2011, 2016) proposed *Logic for Normative Systems* (LNS), Nakayama (2014) defined *Dynamic Normative Logic* (DNL), and Nakayama (2015) constructed *Logic of Belief Structures* (LBS). In this presentation, we combine both approaches in order to analyze social justification of actions. When we get more certain information than before and this new information contradicts a part of our current belief state, we would revise our belief state in order to regain a consistent state. This change of a belief state might force us to change the accepted normative state. In this way, a shared belief revision might cause a change in social justifications. For instance, an allegedly justified action might turn out as an unjustified one after a belief revision.

1. Logic for Extended Normative Systems

There are several forms of belief change. Nakayama (2014) discussed belief updates and normative updates and Nakayama (2015) analyzed belief revisions. In this paper, we investigate how belief revisions influence on normative states. For this purpose, we combine LNS and LBS and define *Logic for Extended Normative Systems* (LENS).

We start with definitions of belief structures and revisions of belief structures in Nakayama (2015).

Definition 1. In this presentation, $cons(X)$ means that X is consistent. We use *iff* as an abbreviation of *if and only if*.

(1a) [Belief structure BS] $BS = \langle ST, > \rangle$ is a *belief structure*, when the following three conditions are satisfied:

[1] $ST = \{T_i : 1 \leq i \leq n \ \& \ T_i$ is a consistent set of FO-sentences (i.e., First-Order sentences)$\}$,

[2] $>$ is a total order on ST and $T_1 > \ldots > T_n$, and

[3] for all $T_i \in ST$ and $T_j \in ST$, $T_i \cap T_j = \varnothing$.

(1b) [k first fragment of BS] $top(BS, k) = \cup\{T_i : 1 \leq i \leq k$ and $T_i \in ST\}$. In other

words, *k first fragment of BS* is the union of the first k elements of BS. We can also define $top(BS, k)$ recursively as follows:

[1] $top(BS, 1) = T_1$.

[2] $top(BS, k) = top(BS, k-1) \cup T_k$.

(1c) [Consistent maximum of BS] $top(BS, k)$ is the *consistent maximum of BS* (abbreviated as $cons\text{-}max(BS)$) iff ($cons(top(BS, k))$ & *not* $cons(top(BS, k+1))$). We call k the *consistent maximum number of BS* (abbreviated as $cmn(BS)$), when $top(BS, k) = cons\text{-}max(BS)$.

(1d) [Deductive closure] $Cn(T) = \{p : p$ deductively follows from $T\}$.

(1e) [Belief set for BS] We call $Cn(cons\text{-}max(BS))$ the *belief set for BS*.

Definition 2. Let H be a consistent set of FO-sentences. Let BS be a belief structure with $T_1 > \ldots > T_n$.

(2a) We define $ext(H, BS)$ as the belief structure with $H > T_1 > \ldots > T_n$. In other words, the extended belief structure of BS by H is the belief structure that can be obtained from BS by adding H as the most reliable element.

(2b) [Belief structure revision] $bsR(BS,H) = Cn(cons\text{-}max(ext(H, BS)))$.

(2c) [Belief structure expansion] $bsEX(BS, H) = Cn(cons\text{-}max(BS) \cup H)$.

Now, we define LENS. LENS admits not only *simple* normative systems, but also *structured* and *revised* normative systems.

Definition 3. Let BB and OB be consistent sets of FO-sentences. Let $NS = \langle BB, OB \rangle$. NS is called a *simple normative system*.

(3a) $\mathbf{B}_{NS}\, p$ iff $p \in Cn(BB)$.

(3b) $\mathbf{O}_{NS}\, p$ iff $cons(BB \cup OB)$ **&** $p \in Cn(BB \cup OB)$ **& *not*** $(p \in Cn(BB))$.

(3c) $\mathbf{F}_{NS}\, p$ iff $\mathbf{O}_{NS} \neg p$.

(3d) $\mathbf{P}_{NS}\, p$ iff $cons(BB \cup OB \cup \{p\})$ **& *not*** $(p \in Cn(BB))$.

(3e) When $BSt = \langle ST, > \rangle$ is a belief structure, we set $NS = \langle cons\text{-}max(BSt), OB \rangle$. NS is called a *structured normative system*.

(3f) When $BSt = \langle ST, > \rangle$ is a belief structure and $bsR(BSt,H)$ is a belief revision of BSt by H, we set $NS = \langle bsR(BSt,H), OB \rangle$. NS is called a *revised normative system*.

We read LNS-formulas as follows:

$\mathbf{B}_{NS}\, p$: It is believed in NS that p.

$\mathbf{O}_{NS}\, p$: It is obligated in NS that p.

$\mathbf{F}_{NS}\, p$: It is forbidden in *NS* that *p*.
$\mathbf{P}_{NS}\, p$: It is permitted in *NS* that *p*.

According to Nakayama (2009), social norms are based on shared beliefs. To emphasize this position, we could read $\mathbf{O}_{NS}\, p$ as "It is believed in *NS* that it is obligated that *p* ".

There are some statements that hold in LNS (Nakayama 2014, 2016). In this paper, we use ***not***, ***&***, ***or***, $\Rightarrow$, and $\Leftrightarrow$ as logical connectives in the meta-language.

Proposition 1.
(4a) $(\mathbf{O}_{NS}\,(p \rightarrow q)\ \mathbf{\&}\ \mathbf{B}_{NS}\, p) \Rightarrow \mathbf{O}_{NS}\, q$.
(4b) $(\mathbf{O}_{NS}\,(p \rightarrow q)\ \mathbf{\&}\ \mathbf{O}_{NS}\, p) \Rightarrow \mathbf{O}_{NS}\, q$.
(4c) $(\mathbf{O}_{NS}\ \forall x_1 \ldots \forall x_n\ (P(x_1, \ldots, x_n) \rightarrow Q(x_1, \ldots, x_n))\ \mathbf{\&}\ \mathbf{B}_{NS}\, P(a_1, \ldots, a_n)\ \mathbf{\&}\ \textbf{\textit{not}}\ \mathbf{B}_{NS}\, Q(a_1, \ldots, a_n)) \Rightarrow \mathbf{O}_{NS}\, Q(a_1, \ldots, a_n)$.

Now, we propose a sufficient condition for *justification of actions*.

Characterization 1. Let *G* be a group of people. Let $B(G,t)$ and $O(G,t)$ be sets of FO-sentences. Let $NS(G,t) = \langle B(G,t)\ O(G,t)\rangle$. Let *p* be a FO-sentence. We characterize justification of actions as follows: $\mathbf{O}_{NS(G,t)}\, p \Rightarrow$ [an action expressed by *p* is *justified* in *G* at *t*].

According to this characterization, it is justified to perform an action that is obligated in the given society. For instance, it is an obligation in the Great Britain (*GB*) to shake hands with introduced people. So, when Jack meats John for the first time, Jack's shaking hands with John is justified in *GB*.

2. Social Justifications of actions and Belief revisions

We can ask an agent for reasons of his/her actions. As Alvarez (2016) pointed out, philosophers often distinguish two kinds of reasons for actions, namely *normative reasons* and *motivating reasons*: "there is a single notion of a reason that is used to answer different questions: the question whether there is a reason for someone to do something (normative) and the question what someone's reason for acting is (motivating)." (Alvarez 2016: Section 1)

Because a normative reason justifies for someone to act in a certain way, it is sometimes called *justifying* reason (Alvarez 2016: Section 2). Characterization 1 is in

accordance with this description. We can use LENS in order to explicitly describe justifying situations. To illustrate this possibility, let us consider two stories.

[Story 1] Mr. *A* was accused of committing a theft and was arrested. At this time, we thought *A* should be punished. Later, it turned out that *A* was innocent. Then, we thought *A* should be released. This simple story describes a change of our normative states caused by our belief revision.

In this case, we can define a normative system $NS(G,1)$ with the following property:

$B(G,1) = T_1 = \{\forall x\ (innocent(x) \rightarrow \neg theft(x)),\ \forall x\ (release(G, x) \rightarrow \neg punish(G, x))\}$.
$O(G,1) = \{\forall x\ (theft(\mathrm{x}) \rightarrow punish(G, x)),\ \forall x\ (innocent(x) \rightarrow release(G, x))\}$.
$NS(G,1) = \langle B(G,1), O(G,1)\rangle$.

Based on Definition 3 and Proposition 1, we can conclude the following sentences:

$\mathbf{B}_{NS(G,1)}\ \forall x\ (innocent(x) \rightarrow \neg theft(x))$.
$\mathbf{B}_{NS(G,1)}\ \forall x\ (release(G, x) \rightarrow \neg punish(G, x))$.
$\mathbf{O}_{NS(G,1)}\ \forall x\ (theft(\mathrm{x}) \rightarrow punish(G, x))\ \mathbf{\&}\ \mathbf{O}_{NS(G,1)} \forall x\ (innocent(x) \rightarrow release(G, x))$.

In order to describe the initial situation of the story, we assume the following belief structure *BSt1* and $NS(G,2)$: $T_2 = \{theft(\mathrm{A})\}$, $BSt1 = \langle\{T_1, T_2\}, >\rangle$ with $T_1 > T_2$, and $NS(G,2) = \langle cons\text{-}max(BSt1), O(G,1)\rangle$. Because $\mathbf{B}_{NS(G,2)}\ theft(A)$ holds, it follows from (4c): $(\textbf{\textit{not}}\ \mathbf{B}_{NS(G,2)}\ punish(G, A)) \Rightarrow \mathbf{O}_{NS(G,2)}\ punish(G, A)$, which means: we should punish *A*, when we think that we have not yet done. Note that $cons\text{-}max(BSt1) = T_1 \cup T_2$. So, it is justified in *G* at 2 to punish *A*, when it has not been already done.

In order to describe the situation after the belief change, we assume $T_3 = \{innocent(A)\}$. In this situation, we have to revise the belief structure *BSt1* by T_3. Thus, the revised normative system after this belief revision is the following: $NS(G,3) = \langle bsR(BSt1, T_3), O(G,1)\rangle$. Because $\mathbf{B}_{NS(G,3)}\ innocent(A)$ holds, it follows from (4c): $(\textbf{\textit{not}}\ \mathbf{B}_{NS(G,3)}\ release\ (G, A)) \Rightarrow \mathbf{O}_{NS(G,3)}\ release\ (G, A))$, which means: we should release *A*, when we think that we have not yet done. Note that $bsR(BSt1, T_3) = T_1 \cup T_3$. So, it is justified in *G* at 3 to release *A*, when it has not been already done.

Let us take another example that is related with a political decision.

[Story 2] The British Prime Minister, Tony Blair, wrongly believed in 2002 that Iraq

possessed weapons of mass destruction. Therefore, he decided to join a military operation organized by the USA against Saddam Hussein. The so called 'Operation Iraqi Freedom' lasted from 20 March to 1 May 2003. After the end of the war, it turned out that there was no evidence that supported the claim of weapons with mass destruction.

We assume an appropriate normative system for the British people $NS(GB,1)$ with $NS(GB,1) = \langle B(GB,1), O(GB,1)\rangle$. Then, it holds $\mathbf{O}_{NS(GB,1))}$ $(attack(GB, Iraq) \wedge disarm(GB, Iraq))$, because $\mathbf{B}_{NS(GB,1)}$ $equipped(Iraq, ms\text{-}weapons)$. [This means: Because Iraq is equipped with weapons of mass destruction, we should attack Iraq and we should disarm Iraq.] Thus, from Characterization 1, we obtain that the attack of Iraq is justified in *GB* at this time.

However, after the belief revision $br1(B(GB, 2)) = bsR(B(GB,1), \neg\, equipped(Iraq, ms\text{-}weapons))$, we obtain: ***not*** $\mathbf{O}_{NS(GB,2)}$ $(attack(GB, Iraq) \wedge disarm(GB, Iraq))$, where $NS(GB,2) = \langle B(GB,2), O(GB,1)\rangle$. Thus, later, the attack of Iraq was seen as not justified in *GB*.

3. Ontological Status of Normative Reasons

John Searle (1995) distinguished *institutional* and *brute* (physical) *facts*. Influenced by this idea, Nakayama (2009) proposed a distinction among *physical*, *introspective*, and *social facts*. An introspective fact for a person *A* is a fact based on a mental state of *A*. A social fact for a group *G* is a fact that is based on a shared mental state of *G*. Social norms are norms based on social facts and social agreements.

We suggest interpreting normative reasons as follows: A normative reason with respect to a group *G* is based on a normative system for *G*. An obligation in a normative system for *G* is epistemologically objective in the sense that its validity is independent of mental states of a single member of *G*. Furthermore, any member of *G* is involved in valid obligations in *G*, so that he/she should perform actions which are accepted as an obligation in *G*. Thus, an obligation in a normative system for *G* can be both objective and *motivating* for members of *G*.

References

Alvarez. M. (2016) "Reasons for Action: Justification, Motivation, Explanation," *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), E. N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2016/entries/reasons-just-vs-expl/>.

Nakayama, Y. (2009) *The Construction of Contemporary Nominalism: Its Application to Philosophy of History*, Tokyo: Shunjusha.

Nakayama, Y. (2010) "Logical Framework for Normative Systems," In: S*OCREAL 2010: Proceedings of the 2nd International Workshop on Philosophy and Ethics of Social Reality*, Hokkaido University, Sapporo, 19-24.

Nakayama, Y. (2011) *Norms and Games: An Introduction to the Philosophy of Society*, in Japanese. Tokyo: Keiso shobo.

Nakayama, Y. (2014) "Speech Acts, Normative Systems, and Local Information Update," In: Y. I. Nakano, *et al.* (eds.) *New Frontiers in Artificial Intelligence (JSAI-isAI 2013 Workshops, Kanagawa, Japan, Selected Papers from LENLS10, JURISIN2013, MiMI2013, AAA2013, DDS13)*, LNAI 8417, Heidelberg: Springer Verlag, pp. 98-114.

Nakayama, Y. (2015) "Formal Analysis of Epistemic Modalities and Conditionals based on Logic of Belief Structures," In: T. Murata, K. Mineshima, D. Bekki, (eds.) *New Frontiers in Artificial Intelligence* (Lecture Notes in Computer Science Vol.9067), Heidelberg: Springer Verlag, pp. 37-52.

Nakayama, Y. (2016) "Chapter 12 Norms and Games as Integrating Components of Social Organizations," In: H. Ishiguro, et al. (eds.) *Cognitive Neuroscience Robotics B*, Heidelberg: Springer Verlag, pp. 253-271.

Searle, J. (1995) *The Construction of Social Reality*, New York: The Free Press.

Should We Abandon the Idea that Norms Have Truth Values?

Takahiro Sawasaki

Graduate School of Letters
Phd Student 1st

2016/10/29

Goals

- To clarify what deontic logicians consider to be norms when they claim that norms do not have any truth values
- To show that we should give up on the idea that norms have truth values

Outline

1. Background Information

- When we utter sentences such as "it is obligatory that ····· ··" or "it is permitted that ·····", we are using *normative sentences*.
- Deontic logic is a logic which deals with the putatively logical relations between normative sentences.

- **SDL** which is the "standard" system of deontic logic has many serious problems.
- One of the most foundamental problems in deontic logic is Jørgensen's dilemma.
- Although Jørgensen's dilemma was originally put forth against the logic of the imperative, we can rewrite this dilemma as follows, so as to make it a dilemma in deontic logic as well.

- Jørgensen's dilemma in deontic logic:
  a dilemma that each of the following propositions seems probable, but not all of them can be true at once.
  1. There are some valid inferences in light of our intuitions which contain normative sentences.
  2. If an inference is valid, then sentences in it have truth values.
  3. All normative sentences express norms.
  4. Norms do not have any truth values.

- Deontic logicians who try to avoid Jørgensen's dilemma paticularly tend to deny proposition 3.
- It seems easy to deny proposition 3 if one realizes the following fact: a normative sentence is sometimes used to express a norm, and sometimes used to express a proposition which describes the existence of the norm.
- The former is often called the prescriptive use of normative sentences and the latter is called the descriptive use of normative sentences.

- They regard proposition 4, i.e., the idea that norms do not have any truth values, as the most philosophically sound idea among them.
- However proposition 4 has recently become a rather dogmatic belief in deontic logic. Many deontic logicians who support this idea rarely provide any convincing arguments for it. Moreover they do not even do so much as to roughly define the concept of norm.

- Therefore, what deontic logicians consider to be norms, and whether it is actually correct to say that norms do not have any truth values is something we must argue for the sake of the foundation of deontic logic.

2. What are Norms?

2-1. The Base of the Concept of Norm

The Base of the Concept of Norm[1]

1. Norms are being expressed by sentences like "it is obligatory that ·····".
   (E.g. Stenius (1963, 250), Alchourrón and Bulygin (1981, 96), Alchourrón and Bulygin (1984, 453-4), Alchourrón and Bulygin (1993, 274), Alchourrón and Martino (1990, 52), Hilpinen and McNamara (2013, 21), Hilpinen (2006, 234).)

The Base of the Concept of Norm[2]

- Therefore norms themselves are not linguistic entities.
- Although literature exists which refers to normative sentences with the word of norm, it is in the minor.

The Base of the Concept of Norm[3]

2. Norms are that which directs something to someone. They do not describes anything.
   (E.g. Alchourrón and Bulygin (1981, 96), Alchourrón and Bulygin (1984, 453-4), Alchourrón and Bulygin (1993, 285), Alchourrón and Martino (1990, 52), Hansen et al. (2007, 3), Makinson (1999, 29-30).)

The Base of the Concept of Norm[4]

- For many deontic logicians, this characteristic has been the strongest reason supporting the idea that norms do not have any truth values.
- So, in this context, to say that a *normative sentence* directs something to someone is not correct, but to say that a *norm* directs something to someone is correct.

The Base of the Concept of Norm[5]

- The base of the concept of norm which deontic logicians have in mind, consists of characteristics 1 and 2.
- However, more detailed characterizations of norm vary greatly depending on the individual. Thus in this presentation, I will classify norms into what I call hyletic norms and expressive norms.
- I owe this use of terminology to Alchourrón and Bulygin's papers.

Hyletic Norm[1]

- Now, in addition to characteristics 1 and 2, what other characteristics do hyletic norms have?

Hyletic Norm[2]

3. Norms are abstract entities similar to propositions. (E.g. Alchourrón and Bulygin (1981, 96), Alchourrón and Bulygin (1984, 453-4), Alchourrón and Bulygin (1993, 274), Alchourrón and Martino (1990, 52).)

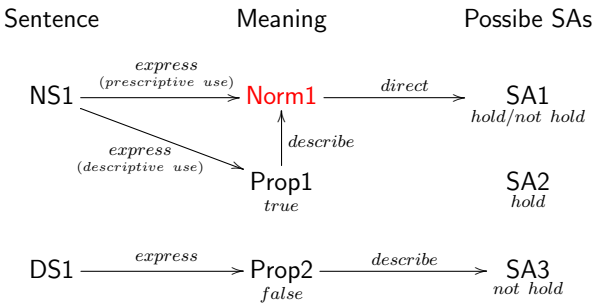
Hyletic Norm[3]

- Hyletic norms are identified with the meaning of normative sentences. The relation between both is just parallel to the relation between propositions and the meaning of declarative sentences.
- Since hyletic norms are abstract entities, they are not temporal entities that appear or disappear.

Hyletic Norm[4]

- Therefore, hyletic norms are abstract entities similar to propositions which direct something to someone; not something which describes anything.
- When we sum up the characteristics of the hyletic norm and the concepts relating to it, we will get the following schema. (NS means "normative sentence", DS means "declarative sentence", and SA means "state of affair")

Hyletic Norm[5]



Expressive Norm[1]

- In opposition to hyletic norms, what other characteristics do expressive norms have besides characteristics 1 and 2?

Expressive Norm[2]

4. Norms are neither the reference nor the meaning of the corresponding normative sentences in Fregean sense. (E.g. von Wright (1963, 94), Alchourrón and Bulygin (1984, 453-4).)
5. Norms are temporal entities which are made or eliminated by the use of their corresponding normative sentences. (E.g. Alchourrón and Bulygin (1993, 277-8), Hilpinen and McNamara (2013, 21).)

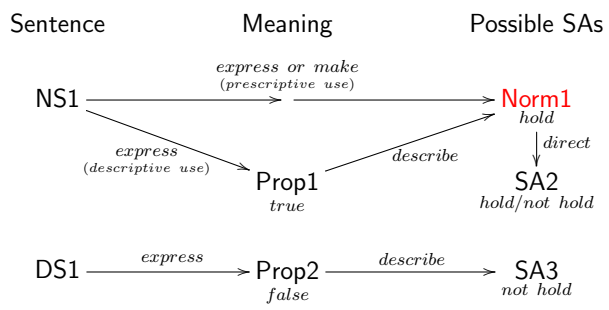
Expressive Norm[3]

- From characteristics 4 and 5, we can say that an expressive norm is a fact, i.e. a state of affair that holds, in the world rather than something located on the side of the language.
- Cf. "The existence of a norm is a fact."(von Wright (1963, 106)).)

Expressive Norm[4]

- Therefore, expressive norms are temporal entities which direct something to someone, and which are expressed or made/eliminated by the use of their corresponding normative sentences.
- Again, when we sum up the characteristics of the expressive norm and the concepts relating to it, we will get the following schema.

Expressive Norm[5]



Summary[1]

- Similarities between both norms :
  - they are expressed by normative sentences; they themselves are not linguistic entities;
  - they do direct something to someone, not describe it.
- Differences between both norms :
  - hyletic norms are the meaning of the corresponding normative sentences, but expressive norms are not;
  - expressive norms are temporal entites, but hyletic norms are not.

Summary[2]

- In my opinion, the framework of expressive norms is closer to the truth on the nature of norms than of the hyletic norms, because the former can capture the aspect of temporality which many norms have.
- However, it seems that many deontic logicians have both of these frameworks in mind, so I will consider whether norms have truth values within each framework below.

3. Do Norms Have Truth Values?

Reasons We Think that Norms Have no Truth Values[1]

- Almost all of the literature, including I have listed here, support the idea that norms do not have any truth values as an indubitable truth.
- The reason for supporting the idea is simple.

Reasons We Think that Norms Have no Truth Values[2]

- The reason is that norms direct something to someone. They do not describe anything.
- As an exception, it seems that von Wright (1963) supports this idea for another reason.

Reasons We Think that Norms Have no Truth Values[3]

- Aside from the reason I have just suggested, another reason for supporting this idea comes from those who point out that norms cannot meaningfully combine with truth functional connectives. (E.g. Reichenbach (1980, 336-44).)
- However, since it seems to me that this fact is the consequence of the idea that norm do not have any truth values, I will ignore it in this presentation.

Reasons We Think that Norms Have no Truth Values[4]

- Now, is it actually correct to say that norms do not have any truth values?

  ⇓

  In my opinion, regardless of whether we understand the concept of norm as a hyletic norm or as an expressive norm, the idea that norms do not have any truth values is actually correct.

Reasons We Think that Norms Have no Truth Values[5]

- When norms are understood as hyletic norms :
  1. norms do not reflect how the world is, because norms are the meaning of normative sentences and do not describe anything;
  2. only that which corresponds to or reflects the world can have truth values;
  3. Therefore, (hyletic) norms do not have any truth values.

Reasons We Think that Norms Have no Truth Values[6]

- Of course this argument presupporses the correspondence theory of truth as a theory of truth.
- However even if we adopt other theories of truth, we still are not able to justify the opposing idea that norms have truth values.

Reasons We Think that Norms Have no Truth Values[7]

- For example, by adopting the coherence theory of truth, some might claim that a norm is true iff it is a part of a consistent normative system.
- However, she then merely restate that norms have truth values in an indirect way, unless she can define the concept of consistency without using truth values.

Reasons We Think that Norms Have no Truth Values[8]

- Moreover, if we can define the concept of consistency without truth values, then whether norms have truth values becomes an extremely trivial problem.
- The reason is that we can just define the logical relations between normative sentences by using the concept of consistency. Indeed, our initial purpose was to elucidate these relations.

Reasons We Think that Norms Have no Truth Values[9]

- When norms are understood as expressive norms :
  1. norms are a kind of fact, i.e. a kind of state of affairs which hold;
  2. Facts themselves cannot be called true or false. For example, the fact that there is a laptop in front of me cannot be called true or false;
  3. Therefore, (expressive) norms do not have any truth values.

Reasons We Think that Norms Have no Truth Values[10]

- Note that characteristic 2, i.e. the characteristic that norms direct something to someone but do not describe anything, has no role in this argument.
- Indeed, even if norms were those which describe something, then we can arrive at the conclusion that norms do not have any truth values when we have expressive norms in mind.

Reasons We Think that Norms Have no Truth Values[11]

- You might wonder if my argument on expressive norms somehow made a mistake, because many deontic logicians have regarded characteristic 2 as a reason for supporting the idea that norms do not have any truth values.
- I think, however, those who have made a mistake here are deontic logicians who adopt the framework of expressive norms, as they have merely confused the ontological status of expressive norms with that of hyletic norms.

Reasons We Think that Norms Have no Truth Values[12]

- One interesting fact is that in *Norm and Action* von Wright also seems to accept for the same reason as mine the idea that norms do not have any truth values. (Cf. von Wright (1963, 103-4).)
- Therefore I believe that norms do not have any truth values for these reasons.

Problems in the Idea that Norms Have Truth Values[1]

- However few deontic logicians have adopted the idea that norms do have truth values.
- I will pick up and critisize Walter (1996), and Kalinowski (1990) as literature which adopt or seem to adopt this idea.

Problems in the Idea that Norms Have Truth Values[2]

- Walter's argument is as follows (Walter (1996, 170)):
- The imperatives or "ought-sentence" are capable of being true or false by introducing the concepts of "the world of "is"" and "the world of "ought"" and assuming the correspondences of declarative sentences to the world of "is" and of ought-sentences to the world of "ought".
- So, norms have truth values (if we do not mind this metaphysically extravaganant assumption on norms).

Problems in the Idea that Norms Have Truth Values[3]

- Walter (1996)'s defects :
  - He seems to confuse norms with propositions which describe the existences of norms.
  - In either the framework of hyletic norms or expressive norms, such propositions are capable of being true or false. However, What is said to be incapable of being true or false are norms of which existences are described by propositions.

Problems in the Idea that Norms Have Truth Values[4]

- Walter (1996)'s defects (continued) :
  - Perhaps he assigns truth values to some kind of sentences, not to norms themselves in the first place.
  - Then of course this criticism would just be the product of my misunderstanding.

Problems in the Idea that Norms Have Truth Values[5]

- Walter (1996)'s defects (continued) :
  - However, even if his argument should be understood as referring to normative sentences, his argument would not be help us to reconstruct deontic logic.
  - The reason is that, in his framework, we do not know how declarative sentences and normative sentences work together.
  - For exmaple, when p is true in the world of "is" and $\mathcal{O}p$ is false in the world of "ought", is "$p \to \mathcal{O}p$" false? If so, why?

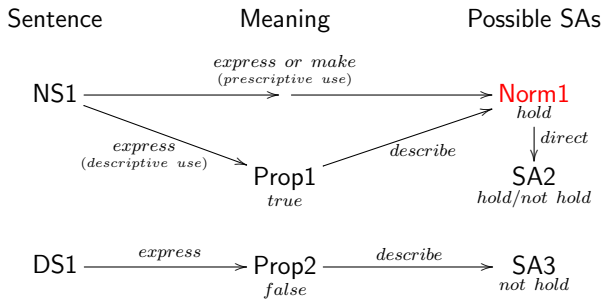
Problems in the Idea that Norms Have Truth Values[6]

- Kalinowski's argument is as follows (Kalinowski (1990, 137-143)):
- If there is a strict parallel for semantic and syntactic level between propositions and norms, then all norms do designate the states of affairs which are normative and real, and thus have truth values.
- Such a parallel exists.
- Therefore, they have truth values.

Problems in the Idea that Norms Have Truth Values[8]

- Kalinowski (1990)' defects :
  - If we adopt the framework of expressive norms, then Kalinowski's argument turns out to be based on the misunderstanding of the concept of norms.
  - Since expressive norms are a kind of fact, norms themselves are neither true nor false.
  - Recall the framework of expressive norms.

Expressive Norm (Revisited)



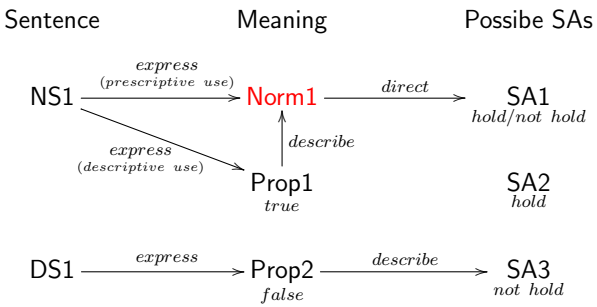
Problems in the Idea that Norms Have Truth Values[9]

- Kalinowski (1990)' defects (continued) :
  - Moreover, suppose we adopt the framework of hyletic norms and, as he argues, norms do designate states of affairs which are normative and real.
  - Then we cannot conceptually distinguish norms from propositions which describe the existence of the norms.

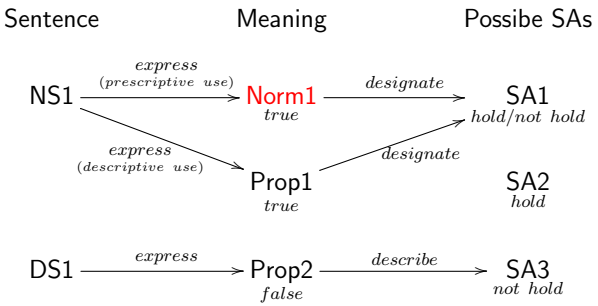
Problems in the Idea that Norms Have Truth Values[10]

- Kalinowski (1990)' defects (continued) :
  - This means two things.
  - The first is that we can use propositions which describe the existences of norms for designating or describing those states of affairs. (In this case rather those states of affairs can be called "norms".)
  - The second is that, in this case, norms have no need to exist at all, regardless of whether or not norms designate anything.

Hyletic Norm (Revisited)



Hyletic Norm (Kalinowski ver.)



Summary[1]

- The idea that norms do not have any truth values is correct.
- A few deontic logicians' idea that norms have truth values is just the product of their misunderstanding the concept of norms which many deontic logicians have in mind.

Summary[2]

- Moreover, it follows that whether or not norms have truth values is thoroughly a problem of the semantics of normative sentences but not metaphysical or epistemological problems on normative sentences.
- I think that because of their misunderstanding the point, Walter (1996), Kalinowski (1990) were misled to the idea that norms have truth values.

4. Conclusion

- Regardless of whether we use the concepts of hyletic norms or expressive norms, we should abandon the idea that norms have truth values.
- Moreover we can see something important:
  1. in deontic logic, when expressions such as $\mathcal{O}p$ are interpreted as norms, such as the expression $\mathcal{O}p \wedge \mathcal{O}q$, they become merely meaningless.
  2. If so, we must reconstruct the system of **SDL** to accomodate this finding.

References

[1] Alchourrón, Carlos E and Eugenio Bulygin (1981) "The Expressive Conception of Norms," in *New Studies in Deontic Logic*: Springer, pp. 95–124.

[2] ——— (1984) "Pragmatic Foundations for a Logic of Norms," *Rechtstheorie*, Vol. 15, No. 1984, pp. 453–464.

[3] ——— (1993) "On the Logic of Normative Systems," in Stachowiak, Herbert ed. *Pragmatik : Handbuch pragmatischen Denkens*: F. Meiner, pp. 273-293.

[4] Alchourrón, Carlos E and Antonio A Martino (1990) "Logic without Truth," *Ratio juris*, Vol. 3, No. 1, pp. 46–67.

[5] Hansen, Jörg, Gabriella Pigozzi, and Leendert van der

Torre (2007) "Ten Philosophical Problems in Deontic Logic," in Boella, Guido, Leon van der Torre, and Harko Verhagen eds. *Normative Multi-agent Systems*, Dagstuhl, Germany: Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

[6] Hilpinen, Risto (2006) "Norms, Normative Utterances, and Normative Propositions," *Análisis Filosófico*, Vol. 26, No. 2, pp. 229–241.

[7] Hilpinen, Risto and Paul McNamara (2013) "Deontic Logic: A Historical Survey and Introduction," in Gabbay, Dov, Jeff Horty, Xavier Parent, Ron van der Meyden, and Leon van der Torre eds. *Handbook of*

Deontic Logic and Normative Systems: College Publication, pp. 3-136.

[8] Kalinowski, Georges (1990) "Ontic and Deontic," *International journal for the Semiotics of Law*, Vol. 3, No. 2, pp. 133–146, translated by E. Landowski and B.S. Jackson from (1989)"Ontique et deóntique",*Rivista Internazionale di Filosofia del Diritto* LXVI, pp. 437–449, with the kind permission of its Director, Sergio Cotta, and Editor, A. Dal Brollo.

[9] Makinson, David (1999) "On a Fundamental Problem of Deontic Logic," *Norms, Logics and Information Systems. New Studies on Deontic Logic and Computer Science*, pp. 29–54.

[10] Reichenbach, Hans (1980) *Elements of Symbolic Logic*: Dover Publications.

[11] Stenius, Erik (1963) "The Principles of a Logic of Normative Systems," *Acta philosophica Fennica*, No. 16, pp. 247-260.

[12] Walter, Robert (1996) "Jörgensen's Dilemma and How to Face it," *Ratio juris*, Vol. 9, No. 2, pp. 168–171.

[13] von Wright, Georg Henrik (1963) *Norm and Action*: Routledge & Kegan Paul PLC.

Thank you for your attention

Acts of Conceding in Dynamified Multi-Agent Epistemic Deontic Logic

Tomoyuki Yamada*

yamada@let.hokudai.ac.jp

1 Introduction

This talk presents some results from an ongoing research on the logic of acts of conceding. The research aims to capture the characteristic effects of acts of conceding in terms of the commitments they generate. The kind of commitments involved here are sometimes called propositional commitments.

The notion of propositional commitments is introduced into the study of dialogue by Hamblin [1], and further studied by Walton and Krabbe [4] with reference to a "persuasion dialogue". They distinguish commitments incurred by making concessions and commitments called assertions.[1] Following Yamada [8], we refer to the commitments of the first type as "c-commitments" and those of the second type as "a-commitments", reserving the term "assertion" for acts of asserting. We say that an agent i is a-committed (or c-committed) to a prosposition φ if i has an a-commitment (or a c-commitment) to φ.

According to Walton and Krabbe ([4], p.8), propositional commitments constitute a special case of commitment to a course of action. The main difference between c-commitments and a-commitments lies in the fact that an agent who is a-committd to a proposition is obliged to defend it if the other party in the dialogue requires her to justify it, while an agent who is c-committd to a proposition is only obliged to allow the other party to use it in the arguments ([4], p.186).

Followin Walton and Krabbe, the characteristic effects of acts of asserting and conceding are studied in terms of these two kinds of propositional commitments in Yamada [8] by developing a dynamic logic of propositional commitments DMPCL. DMPCL incorporates the view that both a-commitments and c-commitments are closed under logical consequence. It enables us to reason to what propositions agents who

*This work is supported by the Grant-in-Aid for Scientific Research on Innovative Areas: Prediction and Decision Making (23120002, MEXT Japan).

[1]They also recognize the third type of commitment called "participants' dark-side commitments" ([4], pp.186-7), but since the commitments of this kind are "hidden or veiled commitments" that are supposed to be fixed, they can be safely ignored for our purpose here.

make assertions and/or concessions are a-committed and/or c-committed. The notion of a-commitments and that of c-commitments, however, are treated as primitives in DMPCL, and thus what these commitments amount to is not analysed in DMPCL.

More recently, Yamada ([9]) have tried to state to what course of action agent who makes an assertion commit herself in epistemic and deontic terms, following Walton and Krabbe's view of propositional commitments as constituting "a special case of commitment to a course of action", and inspired by Williamson's knowledge accounts of assertions in [5] and [6]. This is done by extending the logic DMEDL, Dynamified Multi-Agent Epistemic Deontic Logic, which deals with the effects of acts of commanding, promising, and requesting. In the resulting logic, DMEDL^{+}, an agent who has asserted φ is considered as committing herself to giving enough grouds for the addressee to learn that φ. Though it includes some simplification, it seems that it is on the right track.

Acts of conceding are also briefly discussed in Yamada ([9]). An agent who has conceded φ is considered as committing herself to avoiding asserting $\neg\varphi$. I now find this analysis unsatisfactory. The purpose of this talk is to reconsider it, and to examine what kinds of things need to be done in order to revise it.

References

[1] Charles Leonard Hamblin. *Fallacies*. Methuen, London, 1970.

[2] John MacFarlane. The assessment sensitivity of knowledge attributions. In Tamar Szabó Gendler and John Hawthorne, editors, *Oxford Studies in Epistemology*, volume 1, pages 197–233. Clarendon Press, Oxford, UK, 2005.

[3] John Rogers Searle. *Expression and Meaning*. Cambridge University Press, Cambridge, 1979.

[4] Douglas N. Walton and Erik C. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany, 1995.

[5] Timothy Williamson. Knowing and asserting. *Philosophical Review*, 105:489–523, 1996.

[6] Timothy Williamson. *Knowledge and its Limit*. Oxford University Press, Oxford, UK., 2000.

[7] Tomoyuki Yamada. Acts of requesting in dynamic logic of knowledge and obligation. *European Journal of Analytic Philosophy*, 7(2):59–82, 2011.

[8] Tomoyuki Yamada. Dynamic logic of propositional commitments. In Majda Trobok, Nenad Miščvić, and Berislav Žarnić, editors, *Between Logic and Reality: Modeling Inference, Action, and Understanding*, pages 183–200. Springer-Verlag, Berlin / Heidelberg / New York, 2012.

[9] Tomoyuki Yamada. Assertions and commitments. *Philosophical Forum*, 47(3):475–493, 2016.

A challenge to delineationism from value theory

Richard Dietz, Tokyo Denki University

Summary: This talk brings to bear an idea from value theory to the natural language semantics of gradable adjectives. Delineationists about gradability claim that the meaning of comparatives is *reducible* to the meaning of embedded gradable adjectives (Kamp; Klein; van Benthem; van Rooij; Burnett). On the other hand, it has been argued that in so-called cases of *parity*, pairs of items may be comparable with respect to a covering value even if they determinately fail to instantiate the trichotomy of being either better, worse, or equal (Parfit; Griffin; Chang; Gert; Rabinowicz). I will argue that delineationism fails to supply sufficient means of accommodating cases of parity.

Delineation-based approaches to gradability (henceforth referred to as a *delineationism*) start from the working hypothesis that the meaning of explicit comparatives (*A*er) is a function of the meaning of the positive form of the relevant embedded adjective (*A*). More specifically, they subscribe (in one way or another) to the following kind of reductionist account schema:

Existential Reducibility (ER): x is *A*er than y $\Leftrightarrow_{\mathrm{df}}$ there is a context c such that x is an *A* relative to c, while y fails to be an *A* relative to c.[1]

Delineationism comes in two main varieties: *Kamp-style* accounts, which are formulated in terms of (a) set-theoretic inclusion for admissible valuations in a domain, with admissible valuations going proxy for contexts (Lewis 1970; Kamp 1975); and *Klein-style* accounts, which are formulated in terms of (b) structural constraints on local valuations, with comparison classes going proxy for contexts---very much in formal analogy to revealed choice models of binary of preference relations (Klein 1980; van Benthem 1982; van Rooij 2011; Burnett 2016).

Either type of account has met with objections: Kamp-style accounts face a problem with accommodating biased comparisons (i.e., comparisons of the form x is *A*er than y) between pairs of objects which are indistinguishable in terms of the embedded adjective (*A*). Klein-style accounts, on the other hand, support not only (ER) but moreover the following biconditional

Pairwise Reducibility (PR): x is *A*er than y $\Leftrightarrow_{\mathrm{df}}$ x is an *A* relative to the comparison class $\{x,y\}$, while y fails to be an *A* relative to the same class,

which in the right-to-left direction, seems open to clear counterexamples, e.g., for "tall" (Kennedy 2007). The objection that I am going to raise is orthogonal to the previously presented ones. Neither will I target Kamp's idea of adding in-admissible interpretations that represent counterfactual revisions of a given

[1] For brevity and simplicity, I omit here various details regarding modeltheoretic frameworks for gradability in the delineationist spirit---e.g., I omit here the further complication that typically, context-relative valuations are taken to be partial (i.e. involving potential vagueness). These omissions do not affect the generality of my point though.

model, nor will I target ER (or even PR) in the left-to-right direction, that is, the claim that binary comparative relations must be fully revealed in pairwise choices. That is, even if it is assumed, for the sake of argument, that a Kamp-/Klein-style types of delineationism can be effectively defended against the said challenges, the case for PR (and hence also for ER) in the right-to-left can be effectively rebutted---or so I will argue.

For either case, my argument is essentially the same, appealing to an idea from value theory. On the orthodox picture of value relations, we have a trichotomy of cases: either *x* is overall better than *y*, or *x* is overall worse than *y*, or *x* and *y* are overall equally good (with regard to a relevant covering value). According to this, there is no fourth case of a genuine value relation. On that account, it is suggested that cases where it seems hard to categorise a pair of items in terms of this trichotomy are: either (a) in fact determinately comparable without it being epistemically determinate as to which trichotomy-case holds (epistemicism), or (b) only determinately in the weak sense that it is semantically determinate that the trichotomy holds without it being semantically determinate which case in the trichotomy holds (supervaluationism), or (c) they are simply incomparable. Authors like Parfit (1984) and Griffin (1989), and most famously, Chang (2002) have argued that this picture is too simple and that hard comparison cases may give rise to a fourth type of case where neither trichotomy case determinately holds (or even where each case determinately fails to hold) and yet, where the items are determinately comparable---in short, they are *on a par*. The invoked cases in point for the possibility of parity are typically cases involving evaluative adjectives such as "beautiful" or "clever" that involve more than one contributing value or relevant dimension (e.g., for "clever", it may be at least the skill to manipulate numbers, or the skill to manipulate people---which are independent dimensions).

It is not the aim of my talk to urge further considerations in favour of the thesis that parity is possible. Rather I start from this as a working hypothesis. More specifically, my argument essentially rests on the following weaker claim that is pivotal to the thesis that parity is possible (and which is denied by opponents of this thesis):

Strong comparability (SC): There may be hard comparison cases, *x* and *y*, where it is neither determinate that *x* is better than *y*, nor is it determinate that *x* is worse than *y*, nor is it determinate that *x* is equally good as *y* (with regard to the covering value), and yet *x* and *y* are strongly comparable, in the sense that it is permissible to chose *x* against *y* (or *y* against *x*) as the overall better option.

With (SC) in place, one can give plausible counterexamples to ER in the right-to-left direction. In a nutshell, the general challenge is this: Genuine cases of parity should allow for resolutions in either direction (Rabinowicz 2004). That is, if *x* and *y* are on par (with regard to a covering value), then it is both permissible to chose *x* against *y* and permissible to chose *y* against *x*. Specifically, depending on the context, the same contributing values may be aggregated in different ways such that it may be rational to chose *x* against *y* relative to one context while it may be rationally to chose *y* against *x* in another context (Chang 2002).

It is easy to see that examples of the said type are in conflict with either type of delineationism, since on both accounts, categorical distinctions are irreversible (if in some context, *x* is an *A*, while *y* is not, then there is no context in which *x* and *y* swap sides).

The case against ER in the right-left direction generalises to PR in the right-left direction, which plays a crucial role for Klein-style delineationism.

By parity of reasoning, one can in fact challenge all meaning postulates that are constitutive of standard versions of Klein-style delineationism (specifically, apart from the irreversibility constraint, one can equally challenge a pair of constraints which say that if there is some categorical distinction (of the relevant type) in a comparison class, some distinction will crop up in any contraction and any expansion of this class as well).

Time permitting, I will reply to some possible objections to my objection.

Conclusion: Delineationists about gradability contend (a) that context-dependent categorical distinctions in terms of gradable adjectives reveal associated context-independent orderings, and conversely, (b) that such orderings are fully revealed by associated categorical distinctions. As has been noted in the previous literature, there is reason for casting doubt on this reductionist biconditional in the (b)-direction. The point of my talk is to highlight that the reductionist biconditional is also problematic in the converse (a)-direction. Comparisons that are expressible in terms of multi-dimensional evaluative adjectives give rise to notorious hard cases, which can be reasonably described as cases of parity. On this account, one may reasonably resolve choice problems depending on the context, even if the associated ordering fails to give any guidance as to how to resolve it. Put differently, how to resolve categorisation problems involving multidimensional evaluative adjectives is a meaningful question that is only partially constrained by semantics and to the most part up to substantive theoretical decisions on our part.

References

Benthem, J. van [1982], "Later than late: on the logical origin of the temporal order", *Pacific Philosophical Quarterly* 25:1-36.

Burnett, H. [2016], *Gradability in Natural Language: Logical and Grammatical Foundations*, Oxford: OUP, forthcoming.

Chang, R. [2002], *Making Comparisons Count*, London: Routledge.

Gert, J. [2004], "Value and parity", *Ethics* 114: 492-520.

Griffin, J. [1986], *Well Being*, Oxford: Clarendon.

Kamp, H. [1975], "Two theories of adjectives", in E. Keenan (ed.), *Formal Semantics of Natural Language*, Cambridge: CUP, pp. 123-55.

Kennedy, C. [2007], "Vagueness and grammar: the semantics of relative and absolute adjectives", *Linguistics and Philosophy* 30:1-45.

Klein, E. [1980], "The semantics of positive and comparative adjectives", *Linguistics and Philosophy* 4:1-45.

Lewis, D. [1970], "General semantics", *Synthese* 22:18-67.

Parfit, D. [1984], *Reasons and Persons*, Oxford: OUP.

Rabinowicz, W. [2004], "Modelling parity and incomparability", in W. Rabinowicz and T. Rönnow-Rasmussen (eds), *Patterns of Value: Essays on Formal Axiology and Value Analysis*, vol. 2 (LPR: Lund Philosophy Reports), Department of Philosophy: Lund.

Rooij, R. van [2011], "Revealed preference and satisficing behavior", *Synthese* 179, Suppl. 1: 1-12