the logica yearbook

2006

FILOSOFIA
nakladatelství
Filosofického ústavu
AV ČR
ΦΙΛΟΣΟΦΙΑ

the logica yearbook 2006

THE LOGICA YEARBOOK 2006

Edited by Ondřej Tomala
and Radek Honzík

FILOSOFIA
Prague 2019

Publication of the Logica Yearbook 2006 as well
as the organisation of the Logica 2006 conference
was supported by grant no. 401/04/0117
of the Grant Agency of the Czech Republic.

Published by FILOSOFIA - ΦΙΛΟΣΟΦΙΑ, 2007
Institute of Philosophy
Academy of Sciences of the Czech Republic,
Prague

Edited by Ondřej Tomala and Radek Honzík
Design and typesetting by Martin Pokorný

Printed by PB tisk Příbram, Czech Republic

ISBN 978-80-7007-254-7 (print book)
ISBN 978-80-7007-582-1 (e-book)
DOI 10.47376/filosofia.2007.1

Table of Contents

Preface

When the first of the Logica symposium series took place in 1987 in Liblice in Czechoslovakia, at that time still under communist rule, very few people would dare to predict its future. The study of logic had been neglected for more than 40 years and the discipline itself was in ruins. And yet, Logica not only survived but flourished. It quickly developed from a regional to an international conference with a stable place in the international academic calendar, and has recently celebrated its 20th anniversary. In these twenty years it has also hosted many talks by outstanding logicians, philosophers, mathematicians, linguists and other scholars with an interest in logic. Among them are David Lewis, Jaako Hintikka, Robert Brandom, Barbara Partee, and Nuel Belnap, to mention only few.

The success of the first Logica event also started the long tradition of publishing the related proceedings, which were replaced by Logica Yearbook series in 1997, another volume of which we are pleased to present in this book. It contains a selection of papers presented at Logica 2006, the 20th in the series, which was held for the third time in the magnificent surroundings of the former Franciscan monastery in Hejnice in north-eastern Bohemia. We are very glad that a significant majority of participants of the conference have decided to contribute to this volume.

The original idea of the founding father of Logica, the Head of the Department of Logic of the Institute of Philosophy and Sociology of the Czechoslovak Academy of Sciences, Ivo Zapletal, was to create a platform where logicians, both those of mathematical bent and those interested in philosophical logic and philosophy could meet, present and discuss their results. Over all these years Logica has strived to retain its multidisciplinary flavour. Therefore in this volume you may find papers from the field of philosophical and mathematical logic as well as other areas of analytic philosophy. As a result, and in line with previous editorial policy, we have arranged the papers alphabetically by author, forgoing any attempt to group them by theme or topic.

Naturally Logica 2006 as well as this volume were the result of a joint effort of many people, who deserve our deep thanks. In the first place we would like to thank the main organisers Vladimír Svoboda and Timothy Childers from the Department of Logic, the Institute of Philosophy of the Academy of Sciences of the Czech Republic. The conference would almost certainly not take place without their commitment and dedication. On behalf of the organisers we would also like to thank the Director of the Institute of Philosophy for his support. We would also like to thank the Grant Agency of the Czech Republic whose support of the project no. 401/04/0117 significantly facilitated prepara-

tions of the conference as well as publication of the present Yearbook. We are further indebted to Marie Vučková, Head of the Foreign Relations Department of the Institute for organisational support before and during the symposium and to Martin Pokorný for the layout of this volume. We are grateful for the help of David Göttlich and Petra Ivaničová on organisational matters during and before the conference, especially David's creation and production of the conference documentary movie (see the conference webpage http://www.flu.cas.cz for the result soon). Above all we would like to thank the staff of Hejnice Monastery, especially Father Miloš Raban, for their great hospitality. Special thanks also go to the Bernard Family Brewery of Humpolec, our traditional sponsor of the much appreciated Bernard Open Beer Party, an indispensable part of the conference programme.

Last but not least we would like to thank all conference participants who took the extra effort to prepare their papers for publication and thus made this volume possible. We would also like to thank them for their outstanding cooperation during the editorial process.

Prague, May 2007 Ondřej Tomala and Radek Honzík

Counterpart Semantics for Quantified Modal Logic

Francesco Belardinelli

1. Introduction

In this paper we deal with the semantics for quantified modal logic, *QML* in short, and their philosophical relevance. In the first part we introduce Kripke semantics for the first-order modal language $\mathcal{L}^{=}$ with identity, then we consider some unsatisfactory features of this account from an actualist point of view. In addition, we show that the calculus $Q^E.K + BF$ on free logic, with the Barcan formula, is incomplete for this interpretation. In the second part of the paper we present counterpart semantics, as defined in (Brauner & Ghilardi, 2007; Corsi, 2001). We show that it faithfully formalizes Actualism, encompasses Kripke semantics, and analyses the modal properties of individuals in a more refined way.

Quantified modal logic has always had a strong philosophical appeal, since it first appeared in papers by Barcan Marcus (Barcan, 1946a; 1946b; 1947), Hintikka (Hintikka, 1961; 1969), Prior (Prior, 1956; 1957; 1968) and Kripke (Kripke, 1959; 1963a; 1963b). Besides the topics of propositional modal logic – necessity and possibility, individual knowledge, obligations and permissions, programs and computations – quantified modal logic especially focuses on individuals: we can talk about *actual and possible objects*, the *existence* and the *modal properties* of individuals, as well as *counterfactual situations*. In the philosophy of *QML* we find dramatically relevant issues such as *Actualism/Possibilism*, *realism about possible worlds*, *trans-world identity of individuals*[1]. It is clear that the formal development of quantified modal logic will provide an useful tool to precisely define the concepts above.

2. Kripke semantics

Kripke semantics is widely used to assign a meaning to modal languages; it stems from Leibniz's intuition of defining necessity as truth in every possible world.

[1] See (Chihara, 1998; Loux, 1979; Menzel 2005) for surveys of these subjects.

We start with introducing the first-order modal language $\mathcal{L}^=$, which contains an infinite set of individual variables $x_1, x_2, \ldots$; an infinite set of n-ary predicative constants $P_1^n, P_2^n, \ldots$, for every $n \in \mathbb{N}$; the propositional connectives $\neg$, $\rightarrow$; the universal quantifier $\forall$; the modal operator $\Box$, and the identity symbol $=$. The first-order modal formulas $\phi, \phi', \ldots$ in $\mathcal{L}^=$ are defined as follows:

$$\phi \ ::= \ P^n(y_1,\ldots,y_n) \mid y = y' \mid \neg\phi \mid \phi \rightarrow \phi \mid \Box\phi \mid \forall y \phi$$

The logical constants $\bot$, $\wedge$, $\vee$, $\leftrightarrow$, $\exists$ and $\Diamond$ are defined in the standard way. By $\phi[y_1,\ldots,y_n]$ we mean that the free variables in ϕ are among $y_1,\ldots,y_n$; while $\phi[y/y']$ denotes the formula obtained by substituting some, possibly all, free occurrences of y in ϕ with y', renaming bounded variables if necessary.

Note that no symbol for constants or functors appears in $\mathcal{L}^=$, therefore the only terms in our language are individual variables.

In order to assign a meaning to the formulas in $\mathcal{L}^=$ we extend the Kripke structures for propositional modal logic to the first-order.

Definition 2.1 (Kripke Frame) *A Kripke frame $\mathcal{F}$ - K-frame in short - is a 4-tuple $\langle W,R,D,d \rangle$ s.t.*

- *W is a non-empty set;*
- *R is a relation on W;*
- *for $w,w' \in W$, $D(w)$ is a non-empty set s.t. wRw' implies $D(w) \subseteq D(w')$;*
- *for $w \in W$, $d(w)$ is a possibly empty subset of $D(w)$.*

Intuitively, W is the set of possible worlds and R is the accessibility relation between worlds. Each outer domain $D(w)$ contains the individuals which it makes sense to talk about in w, while each inner domain $d(w)$ is the set of inviduals actually existing in w.

We say that a K-frame $\mathcal{F}$ has *constant* (resp. *increasing*, *decreasing*) inner domains iff wRw' implies $d(w) = d(w')$ (resp. $d(w) \subseteq d(w')$, $d(w) \supseteq d(w')$).

Definition 2.2 (Kripke Model) *A Kripke model $\mathcal{M}$ - K-model in short - is a couple $\langle \mathcal{F}, I \rangle$ where $\mathcal{F}$ is a K-frame and the interpretation I is a function s.t.*

- *for every n-ary predicative constant P^n and $w \in W$, $I(P^n,w)$ is an n-ary relation on $D(w)$;*
- *$I(=,w)$ is the equality relation on $D(w)$.*

Finally, we define the truth conditions for a formula $\phi \in \mathcal{L}^=$ at a world w w.r.t. a w-assignment σ from the variables to the elements in $D(w)$:

$$(\mathcal{M}^\sigma, w) \models P^n(y_1,\ldots,y_n) \quad \text{iff} \quad \langle \sigma(y_1),\ldots,\sigma(y_n) \rangle \in I(P^n, w)$$
$$(\mathcal{M}^\sigma, w) \models y = y' \quad \text{iff} \quad \sigma(y) = \sigma(y')$$

$$
\begin{array}{rll}
(\mathcal{M}^\sigma,w) \vDash \neg\psi & \text{iff} & (\mathcal{M}^\sigma,w) \nvDash \psi \\
(\mathcal{M}^\sigma,w) \vDash \psi \to \psi' & \text{iff} & (\mathcal{M}^\sigma,w) \nvDash \psi \text{ or } (\mathcal{M}^\sigma,w) \vDash \psi' \\
(\mathcal{M}^\sigma,w) \vDash \Box\psi & \text{iff} & \text{for every } w' \in W,\ wRw' \text{ implies } (\mathcal{M}^\sigma,w') \vDash \psi \\
(\mathcal{M}^\sigma,w) \vDash \forall y\psi & \text{iff} & \text{for every } a \in d(w), (\mathcal{M}^{\sigma\binom{y}{a}},w) \vDash \psi
\end{array}
$$

where $\sigma\binom{y}{a}$ is the w-assignment that differs from σ at most on y and assigns element a to y. Note that the clause for $\Box$-formulas is well-defined, as by the increasing outer domain condition σ is a w'-assignment whenever it is a w-assignment.

The truth conditions for the formulas containing the logical constants $\wedge$, $\vee$, $\leftrightarrow$, $\exists$ and $\Diamond$ are defined from those above. Furthermore, a formula $\phi \in \mathcal{L}^=$ is

true at a world w	iff it is satisfied at w by every w-assignment σ;
valid on a model $\mathcal{M}$	iff it is true at every world in $\mathcal{M}$;
valid on a frame $\mathcal{F}$	iff it is valid on every model based on $\mathcal{F}$;
valid on a class C of frames	iff it is valid on every frame in C.

While a w-assignment σ has outer domain $D(w)$ as codomain, the quantifiers range over the inner domain $d(w)$. This means that the classic theory of quantification is not valid on the class of all Kripke frames.

In the next paragraph we highlight the unsatisfactory features of Kripke semantics from an actualist point of view.

3. Actualism

Kripke semantics assumes the increasing outer domain condition: for all $w, w' \in$ W, if wRw' then $D(w) \subseteq D(w')$. This constraint is required for evaluating $\Box$-formulas – otherwise a variable y s.t. $\sigma(y) \in D(w)$ might have no denotation in $D(w')$ – but is it philosophically motivated? In this section we negatively answer this question, on the grounds of problems related to the existence and trans-identity of individuals. Thus, we lay the foundations of a counterpart-theoretic approach to quantified modal logic.

3.1 Increasing outer domains

In section 2 we presented Kripke semantics for the first-order modal language $\mathcal{L}^=$. We remind the evaluation clause for $\Box$-formulas:

$$(\mathcal{M}^\sigma,w) \vDash \Box\phi \quad \text{iff} \quad \text{for every } w', wRw' \text{ implies } (\mathcal{M}^\sigma,w') \vDash \phi$$

The same assignment σ to the variables in $\mathcal{L}^=$ appears in evaluating both $\Box\phi$ and ϕ. This means that $\Box\phi$ is true at a world w for the individuals $a_1,...,a_n$ in the outer domain of w, iff in all the worlds accessible from w, the formula ϕ is true for the same $a_1,...,a_n$. This definition lays down a problem of trans-world existence: in order to evaluate $\Box$-formulas in a K-model, we have to assume that the individuals $a_1,....a_n$ in a world w exist in all the worlds accessible from w. Kripke semantics requires the increasing outer domain condition, which was assumed in def. 2.1.

Nonetheless, there is a number of contexts in which this constraint is not intuitive at all, just consider temporal logics: things now existing probably will not exist in some future time[2]. Even in epistemic and modal logic, we may be willing to think of epistemic states and possible worlds containing fewer individuals than the present one. After all, the actualists deny the existence of all the possible individuals but the actual ones:

> Actualism is the philosophical position that everything there is – everything that can be said to exists in any sense – is actual. Put another way, actualism denies that there is any kind of being beyond actuality; to be is to be actual.[3]

If we accept the actualist account of existence, then we are eventually forced to dropping the increasing outer domain condition.

3.2 Varying domain K-models

In Kripke semantics we have a way to reconcile increasing outer domains and Actualism. It consists in distinguishing for each possible world w an outer domain $D(w)$ of objects, to which it makes sense to ascribe properties and relationships, from an inner domain $d(w)$ of existing individuals, over which the quantifiers range. In this way we obtain the varying domain K-models in section 2, which first appeared in (Kripke, 1963b) as a formal representation of Actualism in the author's intent. This approach has some point, as the varying domain K-models formalize the idea of diverse individuals existing in different instants. Moreover, possibilist principles such as the Barcan formula $\forall x\Box\phi \rightarrow \Box\forall x\phi$, its converse $\Box\forall x\phi \rightarrow \forall x\Box\phi$ and the necessity of existence $\forall x\Box E(x)$ – which are all rejected by actualists – are no longer valid. In conclusion, can actualists be content with the varying domain settings in Kripke semantics?

---

[2] As a roman epigraph states: *Fui non sum, es non estis, nemo immortalis.* This ontological account is known as presentism, for a survey of the eternalism/presentism issue see (Loux, 1998; Lowe, 1998).

[3] (Menzel 2005), p. 1. See also (Linsky & Zalta, 1994), p. 436.

In (Menzel 2005) Menzel lists two actualist issues, which are not completely satisfied by this solution:

1. In the object-language the quantifiers range only over the individuals in the inner domain, as it is expressed by the evaluation clause for $\forall$-formulas:

   $$(\mathcal{M}^{\sigma},w) \vDash \forall y\phi \quad \text{iff} \quad \text{for every } a \in d(w), (\mathcal{M}^{\sigma\binom{y}{a}},w) \vDash \phi$$

   but in the meta-language of K-frames we deal with two distinct sets, i.e. $D(w)$ and $d(w)$, for each $w \in W$. Thus, the possibilia swept out by the door, come back through the window. Furthermore, since the classic theory of quantification is no longer valid, we are eventually forced to introduce the existence predicate E and free logic to recover a sound first-order calculus. This is a quite ironic consequence for a philosophical account which does not want to discriminate between actual and possible existence.

2. In varying domain K-models it can be the case that some individual a belongs to $D(w)$ but not to $d(w)$, for some $w \in W$, nonetheless properties and relationships are usually ascribed to a even in w. From a certain perspective this is quite intuitive: think about Plato who is considered, at the present time, a great philosopher even if he died in 347 BC. But this characteristic of Kripke semantics conflicts with the fundamental thesis of *Strong Actualism*[4]: if an object a does not exist in a world w, then nothing can be said about a in w. If we accept Strong Actualism, then we must admit truth-value gaps in Kripke semantics, even for modal formulas evaluated on existing objects $a_1,...,a_n$, whenever any a_i does not appear in some accessible world.

We conclude that Kripke models with varying inner domains are not a satisfactory proposal for reconciling increasing outer domains and the actualist account, in particular w.r.t. Strong Actualism. These last remarks seem to deny the very possibility of a formal representation for Actualism in Kripke semantics.

3.3 Trans-world identity

There is a further question, concerning the increasing outer domain condition, which deserves more insight. The definition of satisfaction for $\Box$-formulas is an *a priori* construction, the well-definiteness of which is guaranteed by the recur-

---

[4] See (Prior, 1968) for a brief presentation of Strong Actualism.

sive process. When *a posteriori* we want to check whether a modal statement $\Box\phi$ is true for an individual a, we need a method to recognize the same a across possible worlds. This tantamounts to the well-known problem of *trans-world identity*, the bibliography of which has been enlarging during the last half-century[5]. This issue is not our concern for the moment, we consider only the (negative) solution to the problem given by Lewis in (Lewis, 1979). But before, we list two other unsatisfactory aspects of Kripke semantics.

The necessity of identity $x = y \to \Box(x = y)$ and the necessity of difference $x = y \to \Box(x \neq y)$ hold in every K-model, as consequences of the unrestricted validity of Leibniz's Law $x = y \to (\phi \to \phi[x/y])$. But in temporal logics, for instance, we may wish to talk about fusion and fission of individuals.

The calculi $Q^E.K + BF$ (resp. $Q^E.K + CBF + BF$) on free logic, with the Barcan formula (resp. BF and CBF) are incomplete for Kripke semantics, that is, they both validate the necessity of fictionality $\neg E(x) \to \Box\neg E(x)$, but none of them prove this formula. See (Belardinelli, 2006) for a formal proof of this fact. These incompleteness results extends to modalities stronger than K. Furthermore, in (Ghilardi, 1991) Ghilardi proved that Kripke semantics is incomplete for a wide range of *QML* calculi.

We conclude that Kripke semantics is far from being completely satisfactory from an actualist point of view, and it cannot handle fusion and fission of individuals. Moreover, the incompleteness results reveal confusion in the meaning of formulas. *QML* demands a more perspicuous semantics.

4. Counterpart semantics

In the second part of this paper we introduce the counterpart semantics for *QML*, which is based on Lewis' intuition in (Lewis, 1979) that it is not possible to identify individuals across possible worlds. He even denies that the same individual can exist in different worlds. Lewis substitutes the notion of trans-world identity with a not further explained counterpart relation C, that – he claims – need to be neither transitive, nor symmetric, nor functional, nor injective, nor surjective, nor everywhere defined, but is only reflexive. Now a formula $\Box\phi$ is true at a world w for the individuals $a_1,\dots,a_n$ iff in every world w' accessible from w, ϕ is true not for the same $a_1,\dots,a_n$, but for their counterparts $b_1,\dots,b_n$ in w'.

In (Brauner & Ghilardi, 2007) Ghilardi, Corsi in (Corsi, 2001) and Kracht & Kutz in (Kracht & Kutz, 2001; 2002) present various semantics for quantified modal logic based on counterparts. We start with the definition of counterpart frame.

---

[5] We refer to (Loux, 1979), which contains relevant papers on this subject.

Definition 4.1 (Counterpart Frame) *A counterpart frame $\mathcal{F}$ – c-frame in short – is a 5-tuple $\langle W,R,D,d,C\rangle$ s.t.*

- *W, R, D, d are defined as for K-frames, but D need not to satisfy the increasing outer domain condition;*
- *C is a function assigning a subset of $D(w)\times D(w')$ to every couple $\langle w,w'\rangle \in R$.*

Note that we relax Lewis's Counterpart Theory and allow individuals to exist in more than one world. Interpretations and models are defined as in Kripke semantics, but we run into problems if the truth conditions of formulas are given by means of infinitary assignments. Consider the following clause which appears in (Fitting, 2004):

$$(\mathcal{M}^{\sigma},w) \vDash \Box\phi[y_1,\dots,y_n] \text{ iff for every } w' \in W, \text{ for every } w'\text{-assignment } \tau, \\ wRw' \text{ and } C_{w,w'}(\sigma(y_i),\tau(y_i)) \text{ imply } (\mathcal{M}^{\tau},w') \vDash \phi[y_1,\dots,y_n]$$

By this definition, Aristotle's Law $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$ is no longer valid, see (Corsi, 2001) for a counterexample. For recovering this principle either we have to assume Kracht and Kutz's Counterpart-Existence Property: for $w,w' \in W$, for every $a \in D(w)$ there exists $b \in D(w')$ s.t. $C_{w,w'}(a,b)$; or we adopt *finitary assignments* and *typed languages* as Corsi and Ghilardi do. Kracht and Kutz's condition is rather strong and has no deep philosophical motivation, so we choose the second approach.

First of all, we say that each variable x_i is an n-term, for $n \geq i$. The typed language $\mathcal{L}_T^{=}$ is the set of first-order modal formulas inductively defined as follows:

- if P^k is a k-ary predicative constant and $\vec{t}$ is a k-tuple of n-terms, then $P^k(t_1,\dots,t_k)$ is an n-formula;
- if ϕ, ϕ' are n-formulas, then $\neg\phi$ and $\phi \to \phi'$ are n-formulas;
- if ϕ is an $n+1$-formula, then $\forall x_{n+1}\phi$ is an n-formula;
- if ϕ is a k-formula and $\vec{t}$ is a k-tuple of n-terms, then $(\Box\phi)(t_1,\dots,t_m)$ is an n-formula.

We write $\Box(\psi[t_1,\dots,t_k])$ as a shorthand for $(\Box(\psi[t_1,\dots,t_k]))(x_1,\dots,x_n)$.

For $w \in W$, let a finitary n-assignment $\vec{a}$ be an n-tuple of elements in $D(w)$. The valuation $\vec{a}(x_i)$ of an n-term x_i is tantamount to a_i. Finally, the truth conditions for an n-formula ϕ at a world w w.r.t. a finitary n-assignment $\vec{a}$ are inductively defined as follows:

$$\begin{array}{rcl}
(\mathcal{M}^{\vec{a}},w) \vDash P^k(t_1,\dots,t_k) & \text{iff} & \langle \vec{a}(t_1),\dots,\vec{a}(t_k)\rangle \in I(P^k,w) \\
(\mathcal{M}^{\vec{a}},w) \vDash t = t' & \text{iff} & \vec{a}(t) = \vec{a}(t') \\
(\mathcal{M}^{\vec{a}},w) \vDash \neg\psi & \text{iff} & (\mathcal{M}^{\vec{a}},w) \nvDash \psi
\end{array}$$

$$
\begin{array}{rll}
(\mathcal{M}^{\vec{a}},w) \vDash \psi \to \psi' & \text{iff} & (\mathcal{M}^{\vec{a}},w) \nvDash \psi \text{ or } (\mathcal{M}^{\vec{a}},w) \vDash \psi' \\
(\mathcal{M}^{\vec{a}},w) \vDash (\Box\psi)(t_1,\dots,t_k) & \text{iff} & \text{for every } w' \in W, \text{ for every } b_1,\dots,b_k \in D(w'), \\
 & & wRw', C_{w,w'}(\vec{a}(t_i),b_i) \text{ implies } (\mathcal{M}^{\vec{b}},w') \vDash \psi \\
(\mathcal{M}^{\vec{a}},w) \vDash \forall x_{n+1}\psi & \text{iff} & \text{for every } a^* \in d(w), (\mathcal{M}^{\vec{a}\cdot a^*},w) \vDash \psi
\end{array}
$$

where $\vec{a} \cdot a^*$ is the $n + 1$-assignment $\langle a_1,\dots,a_n,a^*\rangle$.

The truth conditions for the formulas containing the logical constant $\wedge$, $\vee$, $\leftrightarrow$, $\exists$ and $\Diamond$ are defined from the ones above. The definitions of truth and validity go as in Kripke semantics. Note that in counterpart semantics the n-formulas $(\Box\psi)(t_1,\dots,t_k)$ and $\Box(\psi[t_1,\dots,t_k])$ are not equivalent: the former has a *de re* reading, while the latter is *de dicto*. Only the implication from the first to the second one holds, while the coimplication $\Box(\psi[x_1,\dots,x_n]) \leftrightarrow (\Box\psi)(x_1,\dots,x_n)$ holds iff the counterpart relation is everywhere defined. Thus, substitution commutes with the modal operators only in particular cases. In the next paragraph we consider the advantages of counterpart semantics.

5. Counterparts and actualism

In par. 3.2 we focused on three features of varying domain K-models, which are not completely satisfactory from an actualist point of view:

1. the presence of *possibilia* at least in the meta-language of Kripke semantics;
2. the recourse to the existence predicate E and free logic to recover quantification;
3. the violation of the principle of Strong Actualism, according to which something not existing in a world w cannot have properties in w.

We show that counterpart semantics can deal with all these problems and solve them, thus giving Actualism the first adequate formal representation probably. As regards the presence of *possibilia* in the meta-language of semantics, we assume that for every $w \in W$, $D(w) = d(w)$, i.e. the individuals, which it makes sense to talk about in w, are all and only the objects existing in w. By this choice the classic theory of quantification holds, therefore neither the existence predicate E nor free logic are needed.

Pay attention to the different consequences of assuming $D(w) = d(w)$ in Kripke and counterpart semantics. In the former this constraint validates some principles the kripkean reading of which is rejected by actualists, i.e. the converse of BF. Hence, Kripke semantics eventually forces actualists towards varying domain K-models and free logic. In counterpart semantics we have none of this, we can set $D(w) = d(w)$ for every $w \in W$ and reject Possibilism and free logic at once. Clearly CBF is still valid in this framework, but its counterpart-theoretic interpretation no longer clashes with the actualist account, as it only corresponds to the following condition:

$$\text{for } w,w' \in W, \text{ for } a \in d(w) = D(w),\ C_{w,w'}(a, b) \text{ implies } b \in d(w') = D(w')$$

This constraint is actualistically acceptable, as it just says that every counterpart in w' of an existing object exists in w'.

As to the third point, if an individual a does not belong to $D(w')$, we need not to ascribe properties or relationships to a in w' in order to avoid truth-value gaps. In evaluating modal formulas w.r.t. the individual a, we consider the features of a only in the actual world, and of its counterpart(s) in the other accessible worlds. Thus, counterpart semantics soundly formalizes Actualism, as it is free from all the three faults listed above.

Furthermore, counterpart semantics can discriminate formulas which are equivalent in Kripke semantics. In K-frames both BF and the necessity of fictionality $\neg E(x) \to \Box\neg E(x)$ are implied by decreasing inner domains: wRw' implies $d(w') \subseteq d(w)$. On the other hand, in c-frames BF tantamounts to the surjectivity of the counterpart relation:

$$\text{for } w,w' \in W, \text{ for every } b \in d(w') \text{ there exists } a \in d(w) \text{ s.t. } C_{w,w'}(a, b)$$

while $\neg E(x) \to \Box\neg E(x)$ holds iff

$$\text{for } w,w' \in W, \text{ for every } b \in d(w'),\ C_{w,w'}(a, b) \text{ implies } a \in d(w)$$

These are quite different constraints, which collapse into decreasing inner domains only in virtue of the strong assumptions on individuals underlying Kripke semantics. In fact, K-frames can be seen as a limit case of c-frames, where the counterpart relation is everywhere defined and it is identity. In this case both surjectivity and fictional faithfulness reduce to decreasing inner domains. We refer to (Belardinelli, 2006) for a formal proof of this fact.

Finally, in counterpart semantics the necessity of identity and the necessity of difference are not unrestrictedly valid, contrarily to what happens in Kripke semantics, but correspond to precise constraints on the counterpart relation:

a c-frame $\mathcal{F}$ is

functional	iff	wRw', $C_{w,w'}(a, b)$ and $C_{w,w'}(a, b')$ imply $b = b'$
	iff	$\mathcal{F} \vDash (x = y) \to \Box(x = y)$
injective	iff	wRw', $C_{w,w'}(a, b)$ and $C_{w,w'}(a', b)$ imply $a = a'$
	iff	$\mathcal{F} \vDash (x \neq y) \to \Box(x \neq y)$

Nonetheless, Leibniz's Law unrestrictedly holds, without implying either the necessity of identity or the necessity of difference.

6. Conclusions

We conclude that counterpart semantics is a major improvement in comparison to the kripkean framework. The former encompasses the latter, in addition it adequately formalizes the actualist account of existence. In *c*-frames we can discriminate formulas deemed equivalent in Kripke semantics and make further subtle distinctions. Counterpart semantics is philosophically and logically motivated, thus deserves a thorough analysis.

We briefly outline some possible developments: (a) There is no standard formalism for typed modal languages, the one used here has to be improved and made more natural. (b) Counterpart semantics is context-sensitive; contexts are represented by the types of formulas, that make explicit the (finite string of) individuals w.r.t. which formulas are meaningful. This feature is relevant in applications to linguistics, in order to explicitly state the background of a meaningful statement. (c) In typed languages we syntactically discriminate between the *de re* and *de dicto* reading of formulas; this characteristic is useful for epistemic logic.

Francesco Belardinelli
Department of Computing
Imperial College
180 Queen's Gate
London SW7 2AZ
F.Belardinelli@imperial.ac.uk
http://www.doc.ic.ac.uk/~fbelard

References

Barcan, R. C. (1946a). The deduction theorem in a functional calculus of first order based on strict implication. *Journal of Symbolic Logic, 11*, 115–118.

Barcan, R. C. (1946b). A functional calculus of first order based on strict implication. *Journal of Symbolic Logic*, *11*, 1–16.

Barcan, R. C. (1947). The identity of individuals in a strict functional calculus of second order. *Journal of Symbolic Logic*, *12*, 12–15.

Belardinelli, F. (2006). *Quantified Modal Logic and the Ontology of Physical Objects.* Ph.D. Dissertation. Pisa: Scuola Normale Superiore.

Brauner, T., & Ghilardi, S. (2007). First-order modal logic. In P. Blackburn, & J. van Benthem, & F. Wolter (Eds.). *Handbook of Modal Logic* (pp. 549–620). Elsevier.

Chihara, C. S. (1998). *The Worlds of Possibility: modal realism and the semantics of modal logic.* Oxford: Clarendon Press.

Corsi, G. (2001). *Counterparts and possible worlds. A study on quantified modal logic*, volume 21. Bologna: CLUEB.

Fitting, M. (2004). First order intensional logic. *Annals of pure and applied logic*, *127*, 173–191.

Ghilardi, S. (1991). Incompleteness results in Kripke semantics. *Journal of Symbolic Logic*, *56*, 516–538.

Hintikka, J. (1961). Modality and quantification. *Theoria*, *27*, 119–128.

Hintikka, J. (1969). *Models for Modalities.* Dordrecht: Reidel Publishing Company.

Kracht, M., & Kutz, O. (2001). The semantics of modal predicate logic II. Modal individuals revisited. In R. Kahle (Ed.). *Intensionality, Lecture Notes in Logic*, volume 22. Los Angeles: ASL.

Kracht, M., & Kutz, O. (2002). The semantics of modal predicate logic I. Counterpart frames. In F. Wolter, & H. Wansing, & M. de Rijke, & M. Zakharyaschev (Eds.). *Advances in Modal Logic*, volume 3. World Scientific Publishing.

Kripke, S. (1959). A completeness theorem in modal logic. *Journal of Symbolic Logic*, *24*, 1–14.

Kripke, S. (1963a). Semantical analysis of modal logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, *9*, 67–96.

Kripke, S. (1963b). Semantical considerations on modal logic. *Acta Philosophica Fennica*, *16*, 83–94.

Lewis, D. (1979). Counterpart theory and quantified modal logic. In M. Loux (Ed.). *The Possible and the Actual* (pp. 110–128). Ithaca, NY: Cornell University Press.

Linsky, B., & Zalta, E. N. (1994). In defense of the simplest quantified modal logic. *Philosophical Perspectives, 8, Logic and Language*, 431–457.

Loux, M. J. (1979). *The Possible and the Actual.* Ithaca, N.Y.: Cornell University Press.

Loux, M. J. (1998). *Metaphysics: A contemporary introduction.* London: Routledge.

Lowe, E. J. (1998). *The Possibility of Metaphysics.* Oxford: Clarendon Press.

Menzel, C. (2005). Actualism. In E. N. Zalta (Ed.). *The Stanford Encyclopedia of Philosophy.*

Prior, A. N. (1956). Modality and quantification in *S5*. *Journal of Symbolic Logic*, *21*, 60–62.

Prior, A. N. (1957). *Time and Modality.* Oxford: Oxford University Press.

Prior, A. N. (1968). Tense logic for non-permanent existents. In *Papers on Time and Tense.* Oxford: Clarendon Press.

Combining Modal Concepts: Philosophical Applications*

Alexandre Costa-Leite

1. Introduction

Some mathematicians and computer scientists have the tendency to believe that modal logic is just about relational structures (i.e. structures composed by a set and relations on this set. Check for instance (Blackburn & De Rijke &Venema, 2001)). This is just one possible way to understand modal logic and, therefore, it does not imply that modal logic can be reduced to such a conception. Conceiving modal logics as "a tool for working with relational structures" (Blackburn et al., 2001) allows logicians, especially mathematically-oriented logicians, to unify a lot of different objects under the same label. However, for philosophical applications such a definition is not entirely adequate because it is not able to capture single philosophical aspects of concepts. Modal logic cannot be reduced to the study of Kripke semantics; nor can it be reduced to the research on what modalities such as possibility and necessity are. Indeed, one can find many definitions of modal logic in the literature. Some of the most important modal logicians have a lot of different conceptions of modal logic. At the very beginning of Hughes and Cresswell (1996), one finds the following remarks:

> "Modal logic is the logic of necessity and possibility, of 'must be' and 'may be'. By this is meant that it considers not only truth and falsity applied to what is or is not so as things actually stand, but considers what would be so if things were different. If we think of how things are as the actual world then we may think of how things might have been as how things are in an alternative, non-actual but possible, state of affairs – or possible world."

The above conception is clearly not founded in the "relational structures slogan", but in a much more passionate account of modal logic. Such a conception can make someone think therefore that modal logic is not about the real world, but just about fictional worlds, because modal logicians are talking

* Work supported by a grant of the Swiss National Science Foundation.

about *possible worlds* or *possible situations* and such entities are not the real world, although the real world is also a possible world. Nobody knows exactly what possible worlds are or if they really exist. Questions about the ontological status of possible worlds have been studied in the literature for a long time. David Lewis (1986) is one of the most famous philosophers who argues that possible worlds have an existence in the same way that the real world has. Such a conception generates an interesting philosophical discussion. Accepting the actual world as a constant realization of possible worlds (or possible worlds becoming real by updating reality), follows that some possible worlds, those which become real, have an ontological status and then really exist, given that they are the actual world.

Other interesting definition of modal logic is defended by Chagrov and Zakharyaschev (1997):

> "Modal logic is a branch of mathematical logic studying mathematical models of correct reasoning which involves various kinds of necessity-like and possibility-like operators."

It seems that both definitions of modal logic were targets of criticism specially developed by those people working on the "relational structures slogan". Blackburn et al. (2001) state the following:

> "One still encounters with annoying frequency the view that modal logic amounts to rather simple-minded uses of two operators $\Diamond$ and $\Box$. The view has been out of date at least since the late 1960's."

Such a comment attempts to establish a new approach to modal logic. Even if the "relational structures" are so fundamental to modal logic, there is no guarantee that in the future a new understanding of modalities will not change the way actual researchers on modal logic think about their subject.

Modal logic is interesting for philosophers because it is related to the metaphysical status of objects and with the content of an agent's mental states. In this sense, modal logic is the study of different existential dimensions of objects and the relations between such objects. For example, in the case of propositional logics, the objects to be considered are propositions and its different existential dimensions are expressed by modalities. Given a hierarchy of possibility operators, each one would be responsible for a given existential dimension of a given proposition. Modal logic, therefore, is a form of research that is concerned with the different ontological dimensions of objects and shows how to manipulate such dimensions.

In this article, such different dimensions of objects are considered in order to show how to apply combined modal logics in philosophy. Modal logic is helpful because it is a tool to clarify the analysis of philosophical concepts. Combining logics plays an important role in philosophical issues because there are some statements containing non-interdefinable operators which cannot be formalized using a single modal logic (i.e a modal logic with just one modal operator). The philosopher usually constructs and finds complicated propositions containing at the same time different modal notions. One example is that of the verification principle, which can be stated as: "All true propositions can be known" this principle often appears in discussions about realism and anti-realism. In the verification principle, one can find two non-interdefinable modalities: possibility and knowledge. Therefore, a very simple modal logic is not able to formalize such a sentence. (A detailed study to this problem is proposed in Costa-Leite, 2006.) Another example, the one treated in this article, is that of non-skepticism about the world. The statement "All contingent propositions are known" involves two non-interdefinable modalities: contingency and knowledge. It is difficult to find works studying in detail how to combine contingency and knowledge. Therefore, attempts to study non-skepticism fail while formalizing the statement. In this article, the philosophical statement above is studied from the viewpoint of combined logical systems. Indeed, one very simple method called fusion is applied to provide an example of formalization. The sense in which such complex formalisms can help in the understanding and formalization of statements linking metaphysics and epistemology will be explained.

2. Formal tools and philosophical concepts

The general theory of modalities still awaits some basic developments, considering that up to now it is not clear what modalities are and just what properties do modalities possess. There are a lot of different modalities and each modality is a particular way to modify a proposition updating its dimensional content. Given a proposition φ, one can always introduce to it a modality. One could create, for instance, $\Diamond\varphi$ (the possibility of φ) or $K\varphi$ (the knowledge of φ). Such modalities allow the construction of expressions of the form "φ is possible" and "φ is known", for example. The intuition and the study of modalities is important to understand other dimensions and properties of the actual world. Although introducing modalities in a given proposition allows statements of the above form, nothing can be said, from the viewpoint of non-combined modal systems, when multiple modalities are interacting in a proposition. In this paper, the interactions between two different families of modalities, those called *metaphysical* (or alethic) modalities and those called *epistemic* modalities are examined. While studying metaphysical and epistemic modalities there is also an attempt to provide explanations in metaphysics and epistemology, respectively. The study of formal concepts can help in the analysis and understanding of philosophical areas. However, it is important to note how such formal tools and concepts are limited. Van Benthem (2005) argued that:

> "Here is the worst that can happen. Some atlases of philosophical logic even copy philosophical geography (epistemic logic, deontic logic, alethic modal logic), leading to a bad copy of a bad map of reality".

This statement seems to contain the key to discovering what is the exact role of formal methods in philosophy. What Van Benthem is arguing for is that formal tools can help, but cannot give an entire understanding of philosophical areas. And there is no doubt that sometimes a formal approach to some philosophical notion can even be a caricature of how to proceed. Van Benthem's claim is correct. It is not reasonable to think that a formal study of metaphysical concepts will examine entirely, or even replace a realistic and intuitive approach, because many aspects of concepts cannot be formalized inside logical systems. In this sense, it is a mistake to think that a formal approach to a given concept can give a complete account to the whole of a given philosophical area. It seems that there will always be some controversy or problem. His statement is interesting to show in what sense reality and language are ingredients of two different things. Consider the formal and logical modality of possibility. Does this modality correspond to what possibility really is? It is hard to say. Formal tools help in the clarification and partial description of what a concept really is, but it is never able to describe the totality of the concept. One interesting property of formal concepts and tools is that some philosophical revolutions can be reached by a formal approach to concepts. One good example is that of Kripke (1980) who showed that there are some necessary *a posteriori* truths. Such a result shows exactly the right role of logic in philosophy: from one side, logic cannot eliminate all philosophical problems and it cannot give a total description of a given concept. But from the other side, the use of logic really helps to create some models of reality.

3. The problem

Gabbay (1999) pointed out the existence of a very interesting logical problem which reflects directly in philosophical issues. This is called the *fibring problem*. It can be explained in the following way: take a Kripke model $\langle W,R,v \rangle$ for $\Diamond$, a formula φ and the complex modality $\Diamond K$. Given φ, introduce to it the combined modality in order to obtain $\Diamond K\varphi$. Now, to determine the truth-condition of the formula in the Kripke model one has to proceed in the standard way. However, when the truth-condition of the modality $K\varphi$ is examined, the above Kripke model cannot continue the procedure, because it is not able to recognize what K means. In this sense, Gabbay proposed to associate to each world a new model using something called the fibring function in order to be able to analyze the truth-condition of the complete formula. The idea of fibring is that sometimes the models are not sufficiently rich to determine truth-conditions of all propositions. Some many new variations of fibring have been proposed by many researchers in the branch called *combining logics.* A general approach to combined modalities and a great variety of references can be found in Costa-Leite (2004).

The fibring problem appears everywhere in philosophical analysis. In Costa-Leite (2006) tools from combining logics played an important role in studying in detail the exact set to formulate and think about a paradox related to the verification principle. In this work, a new example is provided using combinations of a metaphysical modality and an epistemic modality.

Metaphysical (or alethic) modalities are those related to the general structure of reality. The name *metaphysical* reflects this content. A metaphysical modality is one which is not directly related to the actual world, but with some potential configuration of this world. The name *alethic* suggests that the notion of truth appears in these modalities. The name alethic therefore is not a good one, because one can think that just alethic modalities deal with the notion of truth, what is incorrect. The general name *metaphysical* describes the job: modalities which state potential configurations of reality.

Epistemic modalities are not directly related to the general structure of reality, but rather with the cognitive status that an agent can have with respect to the world. The name *epistemic* suggests, of course, some relation between agents and the world. Epistemic modalities are also related to the concept of truth, especially when it comes to analyzing truth-conditions of epistemic formulas.

The study of metaphysical, deontic, epistemic, temporal and others kinds of modalities has been the target of much research. What has not been studied are conditions where one can find *interactions* of different families of modalities, as in the example above where the combination $\Diamond K\varphi$ appears. Some other examples of interactions are the following: $K\Diamond\varphi$, $K\varphi \rightarrow \nabla\varphi$ (knowledge implies contingency), $K\varphi \rightarrow \Diamond\varphi$ etc. There are a lot of cases. Such statements show propositions where two different families of modalities are interacting in a combined complex statement. The study of interactions between metaphysical and epistemic modalities deserves attention, because it provides a key to the door linking metaphysics and epistemology, and allows therefore a study of philosophical statements involving such concepts. Dana Scott (as cited in Hendricks & Symons 2006) correctly said:

> "Here is what I consider one of the biggest mistakes of all in modal logic: concentration on a system with just one modal operator. The only way to have any philosophically significant results in deontic logic or epistemic logic is to combine these operators with: Tense operators (otherwise how

can you formulate principles of change?); the logical operators (otherwise how can you compare the relative with the absolute?); the operators like historical or physical necessity (otherwise how can you relate the agent to his environment?); and so on and so on. (Dana Scott, 1970)"

3.1 The example

One of the first examples, which is not explained here in detail, can be found in Costa-Leite (2006). There is showed that the right language and logic in which to formulate Fitch's paradox is composed by a fusion of modal languages and modal logics. In this sense, one can add the verification principle $\varphi \to \Diamond K\varphi$ to such a fusion without the collapse of truth and knowledge. The reader is invited to check that article to see how Fitch's paradox can be studied from the viewpoint of combined logics. Fusion of modal logics is a very simple method to combine modal logics. Such method has been studied mainly by Gabbay, but it has been discovered by Kracht & Wolter (1991), and also by Fine & Schurz (1997). The method is briefly explained in the example.

Consider the statement

(ST) "All contingent propositions are known."

There are many possible formalizations of the above sentence, it depends in what logic it is being formalized. First it is clear that a modal logic with just one modal operator cannot do the job. If one has just a metaphysical modal logic, then it is not able to formalize the knowledge operator. In the same way, with just an epistemic logic, it would not be possible to formalize contingency. Therefore, just a combined formalism can realize the task. But what is such combined modal logic?

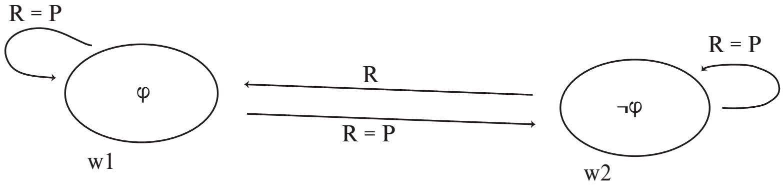
Logics of contingency were proposed by Montgomery & Routley (1966), and the authors presented a lot of systems taking contingency as a primitive operator. Such an approach is followed here (i.e. contingency as a primitive modality). Surely, if contingency is not taken as primitive, but defined using possibility, then the formalization is different. The contingency of a formula φ is represented by $\triangledown\varphi$. One can read such formula as "φ is contingent". Contingency of a formula φ means that φ is possible and it is possible non-φ. Epistemic logics use K to formalize knowledge. Consider now a language containing $\langle \&, \to, v, \sim, \triangledown \rangle$ and a language containing $\langle \&, \to, v, \sim, K \rangle$. A language containing $\triangledown$ and K among its symbols is certainly a logic able to formalize

(ST') "If a proposition is contingent, then it is known."

The language <&,→,v,~,▽,K> is called a fusion of the above structures. (Check Gabbay (1999) for details on fusions of modal logics.)

(ST) and (ST') are equivalent ways to announce the non-skeptical thesis. Such a thesis intends to show that the world is an object of knowledge. The first conclusion of this paper is that to formalize the non-skeptical thesis one needs a fusion of two languages, one for contingency and the other for knowledge.[1] But what is the logic of such a language? What does it semantics looks like?

The answers to the above problems depend of what the reader intend to do, assuming that there is no absolute answer. Using the fused language, the formalization of (ST) or (ST') is: $\triangledown\varphi \rightarrow K\varphi$. The fused axiomatic system generated using such a language determines whether (ST) is valid or not (the same for (ST')). Such axioms certainly contain at least the axioms of a minimal modal logic of contingency and minimal epistemic logic. Fusion of two axiomatic systems A1 and A2 consists in putting together both axiomatic systems in a big set which contains all axioms of both A1 and A2, and all inference rules of both (check Gabbay (1999) for a detailed study on fusions). Surely, from the semantical viewpoint, fusion of two Kripke structures F1 and F2 consists in putting together both accessibility relations. In this sense, if F1 = <W,R> is a structure for contingency, and F2 = <W,P> is a structure for knowledge, the fusion of both is the structure F1⊕F2 = <W,R,P>. The accessibility relations of the fused structure have the same properties of the original accessibility relations. It means for instance that if R is reflexive in F1, then R is also reflexive in the fusion. Let consider here a fusion where the accessibility relation R is reflexive and symmetric, but P is just reflexive. Consider semantically (ST). Assume a fused Kripke model F1⊕F2 = <W,R,P,v>, the formula $\triangledown\varphi \rightarrow K\varphi$ and put $P \subseteq R$. Take $W = \{w1,w2\}$ and the following valuation: $v(\varphi) = \{w1\}$. In such a model, the formula is not valid, and therefore it is not a theorem of the fused logic, given that completeness is preserved by fusions (check Gabbay (1999) for details on completeness preservation). See the picture below:



[1] Classical propositional language can be viewed as a fusion of a language containing just negation and a language containing, for instance, conjunction. It is important to state that it is a fusion because it is now clear what method is used to generate such a language.

In the world w1 and w2, $\nabla\varphi$ holds, but in w1 and w2 $K\varphi$ does not hold. In this sense, the formula is not valid in the model. Surely one could create a modal logic showing that (ST) is valid, but again it depends of what one intends to do. What is important to state is that a complex formula involving two non-interdefinable modalities cannot be analyzed from the semantical viewpoint without a combined modal system able to understand at the same time what each one of the modalities means.

4. Conclusion

Combining modal concepts allows the study of complex statements formulated in natural languages. Such kind of approach provides a formal study on many different philosophical statements. In the example studied in this text, a concept from metaphysics (contingency) is linked to a concept from epistemology (knowledge) using a fusion of Kripke models. In this sense, combining concepts formally generates new insights in the study of the bridges between many different philosophical subjects.[2]

University of Neuchâtel
Switzerland
alexandre.costa-leite@unine.ch
http://luna.unine.ch/alexandre.costa/costaleite.html

References

Blackburn, P., & De Rijke, M., & Venema, Y. (2001). *Modal logic*. Cambridge: Cambridge University Press.

Chagrov, A., & Zakharyaschev, M. (1997). *Modal logic*. Oxford: Clarendon Press.

Costa-Leite, A. (2004). Towards a general theory of the combination of logics. *Travaux de Logique, 17* (pp. 219–230). Neuchâtel: Centre de Recherches Sémiologiques.

Costa-Leite, A. (2006). Fusions of modal logics and Fitch's paradox. *Croatian Journal of Philosophy, 17*, 281–290.

[2] I would like to thank to Casey McGinnis for important comments, and to Ondrej Tomala for useful help.

Gabbay, D. (1999). *Fibring logics*. Oxford: Clarendon. Press.

Hendricks, V., & Symons, J. (2006, January 4). Epistemic logic. In *Stanford Encyclopedia of Philosoph*y. Retrieved December 02, 2006 from http://plato.stanford.edu/archives/spr2006/entries/logic-epistemic/

Hughes, G. E., & Cresswell, M. (1996). *A new introduction to modal logic.* London: Routledge.

Kripke, S. (1980). *Naming and necessity.* Cambridge: Harvard University Press.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell.

Montgomery, H., & Routley, R. (1966). Contingency and non-contingency bases for normal modal logics. *Logique et Analyse, 9*, 318–328.

Van Benthem, J. (2005). Logic in philosophy. In D. Jacquette (Ed). *Handbook of the Philosophy of Logic*. Amsterdam: Elsevier. Retrieved February 18, 2006 from: http://www.illc.uva.nl/Publications/reportlist.php?Year=2005

The Use-Mention Distinction[1]

Marie Duží

1. Introduction

In this paper we are not going to examine the linguistic problem of distinguishing between using and mentioning expressions as the title might suggest. Instead, we are going to logically analyse particular different ways of using expressions. When we use an expression in a communicative act then we communicate its meaning; we are not interested in other meanings the words might have had. Logical analysis presupposes understanding and linguistic competence.

Our analyses comply with the principles of compositionality and referential transparency: When an expression *E* is used to communicate its meaning then *E* expresses the same entity as its meaning and denotes the same entity as its denotation (or 'semantic value') regardless of the embedding context. This means rejecting so-called reference shift across the board. We are going to show that instead of the 'shifts' of reference there are different ways in which *E* may occur relative to a logical-semantic context. Either its meaning is *used* to pick up an entity denoted by *E* (if any) or the meaning itself is just *mentioned* as an entity referred to by another expression *E*' of which *E* is a subexpression. And if the meaning is used, it may occur either with supposition *de dicto* or *de re*.

The underlying Transparent Intensional Logic (TIL) is a hyper-intensional λ-calculus, which means that the terms are not interpreted as denoted functions. Rather, they denote, or 'encode', (algorithmically) *structured procedures* known as *TIL constructions* that are assigned to expressions as their (structured) meanings[2]. Constructions, when being executed, produce functions. The theory also contains the resources to distinguish in a principled manner between functions and their values by distinguishing between constructions occurring *intensionally* and *extensionally.*

---

[1] This work has been supported by the Grant Agency of the Czech Republic (project No 401/04/2073, Transparent Intensional Logic (systematic exposition)), and by the Project No. 1ET101940420 "Logic and Artificial Intelligence for multi-agent systems" pursued within the program "Information Society" of the Czech Academy of Sciences.

[2] The notion of structured meaning and hyperintensionality has been introduced by Cresswell (1975). A similar semantic conception has been applied by Yiannis Moschovakis, see his (1994), (2006).

The examples in, e.g., Gamut (1991, pp. 203, 204) illustrate the problems which may arise from the confusion of different ways of using expressions. To adduce one, consider the following (obviously invalid) argument:

The temperature in Amsterdam equals the temperature in Prague.
The temperature in Amsterdam is increasing.
___
The temperature in Prague is increasing.

There is an essential difference between the way of using the term 'the temperature in Amsterdam' in the first and the second premise. In the first premise the (empirical) *function*, namely the *magnitude TA* denoted by 'temperature in Amsterdam', is *used* to point to its current actual value; the premise claims that this value equals the current value of another magnitude *TP* (denoted by 'the temperature in Prague'). However, the second premise ascribes the property of being increasing to the whole magnitude *TA* regardless its current value: the *function TA* itself is not used (as a pointer to its current value) but only *mentioned*.

Here is another example:

$$\text{(Calc)}\;\frac{\begin{array}{l}\textit{Charles calculates } 2+5.\\ 2+5=7\end{array}}{\text{Charles calculates 7.}}$$

Again, there is a substantial difference between using the term '*2 + 5*' in the first and second premise: whereas in the first premise the meaning of '*2 + 5*' is mentioned, in the second one it is used to identify the number 7. The first premise expresses Charles' relation(-in-intension) to the very procedure of calculating 2+5. Charles is trying to execute the procedure, and the *procedure*, the meaning of the expression '2+5', is *mentioned* here. When evaluating the truth-conditions expressed by the first premise, the procedure of adding numbers 2 and 5 is *not* executed; this is Charles' responsibility.

We are going to solve the apparent paradoxes by means of the TIL fine-grained analysis of premises that neither makes it possible to over-infer (which leads to paradoxes) nor under-infer (which leads to a lack of inferential knowledge).

2. TIL in brief

In this chapter we provide just a brief introductory explanation of the main notions of TIL. For exact definitions see, e.g., Tichý (1988), Materna (1998), Materna (2004).

Constructions are procedures, or instructions, specifying how to arrive at less-structured entities. *Qua* procedures they operate on input objects (of any type, even on constructions of any order) and yield as output (or, in well defined cases fail to yield) objects of any type; in this way constructions construct *partial* functions.

By claiming that constructions are algorithmically structured, we mean the following: a construction *C* – being an instruction – consists of particular steps, i.e., sub-instructions (or, *constituents*) that have to be executed in order to execute *C*. The concrete/abstract objects an instruction operates on are not its constituents, they are just mentioned. Hence objects have to be supplied by another (albeit trivial) construction. The constructions themselves may also be only mentioned: therefore one should not conflate *using* constructions as *constituents* of composed constructions and *mentioning* constructions that enter as input into composed constructions. Mentioning is, in principle, achieved by using atomic constructions. A construction is atomic if it is a procedure that does not contain any other construction as a used constituent but itself. There are two atomic constructions: *variables* and *trivializations.*

Variables are constructions that construct an object dependently on *valuation*: they *v-construct,* where *v* is the parameter of valuations. When X is an object (including constructions) of any type, the *Trivialization* of X, denoted $^{0}\boldsymbol{X}$, constructs X without the mediation of any other construction.

TIL constructions as well as the entities they construct all receive a type. The formal ontology of TIL is bi-dimensional. One dimension is made up of constructions, the other dimension encompasses non-constructions. On the ground level of the type-hierarchy, there are entities unstructured from the algorithmic point of view belonging to a ***type of order 1***. Given a so-called *epistemic (or 'objectual')* ***base*** of ***atomic types*** (o-truth values, ι-individuals, τ-time moments / real numbers, ω-possible worlds), we have the induction rule for forming types of partial functions: where α, $\beta_1,\ldots,\beta_n$ are types of order 1, the set of partial mappings from $\beta_1 \times\ldots\times \beta_n$ to α, denoted $(\boldsymbol{\alpha\, \beta_1\ldots\beta_n})$, is a ***functional type*** of order 1 as well.[3]

Constructions that construct entities of order 1 are ***constructions of order 1***. They belong to a ***type of order 2***, denoted by $*_1$. By using the induction rule, any collection of partial functions, type $(\boldsymbol{\alpha\, \beta_1\ldots\beta_n})$, *involving* $*_1$ in their domain or range is a ***type of order 2***. Constructions belonging to a type $*_2$ that identify entities of order 1 or 2, and partial functions involving such constructions, belong to a ***type of order 3***. And so on *ad infinitum*.

There are two *compound* constructions, which consist of other constructions: *Composition* and *Closure.*

Composition is the instruction to apply a function *f* to an argument A in order to obtain the value (if any) of *f* at A: if *X v-constructs* a function *f* of a

[3] TIL is an open-ended system. The above epistemic base $\{o, \iota, \tau, \omega\}$ was chosen, because it is apt for natural-language analysis, but the choice of base depends on the area to be analysed.

type (α $\beta_1...\beta_m$), and $Y_1,...,Y_m$ *v-construct* entities $B_1,...,B_m$ of types $\beta_1,...,\beta_m$, respectively, then the *composition* $[X\ Y_1\ ...\ Y_m]$ is a *construction* that *v-constructs* the value (if any, of type α) of the (partial) function f at the argument $\langle B_1, ..., B_m\rangle$. Otherwise the *composition* $[X\ Y_1\ ...\ Y_m]$ does not *v-construct* anything: it is *v-improper.*

Closure is the procedure of constructing a function by abstracting over variables, i.e., the instruction to do so: If $x_1, x_2, ...,x_m$ are pairwise distinct variables that *v-construct* entities of types $\beta_1, \beta_2, ..., \beta_m$, respectively, and Y is a *construction* that *v-constructs* an entity of type α, then $[\lambda x_1...x_m\ Y]$ is a *construction* called *Closure*, which *v-constructs* the following function f of the type (α $\beta_1...\beta_m$), mapping $\beta_1 \times...\times \beta_m$ to α: Let $B_1,...,B_m$ be entities of types $\beta_1,...,\beta_m$, respectively, and let $v(B_1/x_1,...,B_m/x_m)$ be a valuation differing from v at most in associating the variables $x_1,...x_m$ with $B_1,...,B_m$, respectively. Then f associates with the m-tuple $\langle B_1,...,B_m\rangle$ the value $v(B_1/x_1,...,B_m/x_m)$-constructed by Y. If Y is $v(B_1/x_1,...,B_m/x_m)$-improper, then f is *undefined* on $\langle B_1,...,B_m\rangle$.

Finally, higher-order constructions can be used twice over as constituents of composed constructions: If X is a *construction* that *v-constructs* a *construction X'*, then ^2X is a *construction* called *Double Execution.* It *v-constructs* the entity (if any) *v-constructed* by *X'*. Otherwise the *Double Execution* ^2X is *v-improper*.

Functions values of which depend on a modal (type ω) and/or temporal (type τ) parameters receive a spatial status in TIL likewise in any intensional logic:

(α-)intensions are members of a type ($\alpha\omega$), i.e., functions from possible worlds to the arbitrary type α; *(α-)extensions* are members of the type α, where α is not equal to ($\beta\omega$) for any β.

Notational conventions: An object A of a type α is called an α-object, denoted A/α. That a construction C *v*-constructs an α-object is denoted $C \rightarrow_v \alpha$. We write '$\forall x\ A$', '$\exists x\ A$' instead of '$[^0\forall^\alpha\ \lambda x\ A]$', '$[^0\exists^\alpha\ \lambda x\ A]$', respectively, when no confusion can arise. We also often use an infix notation without trivialisation when using constructions of truth-value functions $\wedge$ (conjunction), $\vee$ (disjunction), $\supset$ (implication), $\equiv$ (equivalence) and negation ($\neg$), and when using a construction of an identity.

Intensions are frequently functions of a type (($\alpha\tau$)ω), abbreviated $\alpha_{\tau\omega}$. We use variables $w, w_1, w_2,...$ as *v*-constructing elements of type ω, and $t, t_1, t_2, ...$ as *v*-constructing elements of type τ. If $C \rightarrow \alpha_{\tau\omega}$ *v*-constructs an α-intension, the frequently used composition of the form $[[C\ w]\ t]$, *v*-constructing the intensional descent of the α-intension, will be abbreviated as C_{wt}.

Some important kinds of intensions are:

Propositions of type $o_{\tau\omega}$, *α-properties* of type $(o\alpha)_{\tau\omega}$, *relations-in-intension* of type $(o\beta_1...\beta_m)_{\tau\omega}$. Omitting $_{\tau\omega}$ we get the type ($o\beta_1...\beta_m$) of *relations-in-extension* (to be met mainly in mathematics); α*-roles, offices* are of type $\alpha_{\tau\omega}$, where $\alpha \neq (o\beta)$, frequent are those with type $\iota_{\tau\omega}$. Individual roles correspond to what Church in his (1956) called "*individual concept".*

3. Using / Mentioning constructions

The distinction between using and mentioning constructions is characterised as follows[4]:

Let *D* be a sub-construction of a construction *C*. Then an occurrence of *D is mentioned in C* if it is not necessary to execute the occurrence of *D* in order to execute *C*. Otherwise the occurrence of *D is used in C as a constituent.*

Following the above example of Charles' calculating, the analyses of premises P_1, P_2 are:

Types: *Charles* / ι; *Calc*(ulate) / $(o\ \iota\ *_1)_{\tau\omega}$; + / (τττ); 2, 5, 7 / τ; = / (οττ).

P_1: $\lambda w \lambda t\ [{}^0Calc_{wt}\ {}^0Charles\ {}^0[{}^0\!+\ {}^02\ {}^05]]$ $(/*_2, \rightarrow o_{\tau\omega})$
P_2: $[{}^0\!=\ [{}^0\!+\ {}^02\ {}^05]\ {}^07] \rightarrow o$.

Now it is obvious that the construction $[{}^0\!+\ {}^02\ {}^05]$ ($\rightarrow \tau$) cannot be substituted for the construction ${}^0[{}^0\!+\ {}^02\ {}^05]$ ($\rightarrow *_1$) into the P_1-constituent. Such a substitution would constitute a type-theoretical category mistake. Calculating is not a relation(-in-intension) between an individual and a particular number; rather it is a relation(-in-intension) between an individual and a *construction* of a number. We see no reason to challenge the unrestricted validity of Leibniz's Law of substitution (except for quotational contexts), and TIL has the resources to validate the Law in any sort of context, which we are going to show.

The occurrence of the construction $[{}^0\!+\ {}^02\ {}^05]$ is *mentioned* in the P_1-constituent by the Trivialisaton ${}^0[{}^0\!+\ {}^02\ {}^05]$, whereas it is *used* in P_2. In order to evaluate (for a state of affairs $\langle W,T\rangle$) the truth-conditions specified by P_1, one does not have to execute the computational step $[{}^0\!+\ {}^02\ {}^05]$. P_1 has nothing to do with whether Charles succeeds in executing the step $[{}^0\!+\ {}^02\ {}^05]$.

Note that an occurrence of a construction can be mentioned in *C* indirectly by being a constituent of another sub-construction which is mentioned in *C*. Moreover, a Double Execution may suppress the effect of Trivialisation. For instance, though the construction $[{}^0\!+\ {}^02\ {}^05]$ is mentioned by the construction ${}^0[{}^0\!+\ {}^02\ {}^05]$, the ${}^{20}[{}^0\!+\ {}^02\ {}^05]$ constructs the number 7, and both ${}^0[{}^0\!+\ {}^02\ {}^05]$ and $[{}^0\!+\ {}^02\ {}^05]$ are used in ${}^{20}[{}^0\!+\ {}^02\ {}^05]$ (the former by itself and the latter by the former). Unlike Trivialisation, which is an operation of *mentioning*, Execution and Double Execution are operations of *using*.

Concerning the case of 'indirect mentioning', consider as an example the sentence

> "Charles knows that dividing six by three makes two and dividing six by zero is improper."

[4] For the definition see Duží & Jespersen & Materna (2007, §4.9).

Note that if we wanted to analyse this sentence in any standard logic (including Montague's intensional logic, which lacks constructions or something akin to them) we would not have a tool to analyse this sentence in the logic. We would have to switch into a kind of linguistic metamathematics.

Let *Improper* be the class of constructions of order 1 which are *v*-improper for any valuation *v*. Hence *Improper* / $(o*_1)$ belongs to a type of order 2. When knowing the above fact, Charles is related to a respective *construction* (belonging to $*_2$) of the value **T**. Therefore, *knowing* is here an $(o\,\iota\,*_2)_{\tau\omega}$-object.

Types: 0, 2, 3, 6 / τ, *Div* / $(\tau\tau\tau)$, *Improper* / $(o*_1)$, *Know* / $(o\,\iota\,*_2)_{\tau\omega}$

The analysis of the embedded clause is:

(Em) $[[[{}^0Div\ {}^06\ {}^03] = {}^02] \wedge [{}^0Improper\ {}^0[{}^0Div\ {}^06\ {}^00]]]$.

The construction (Em) constructs **T**. All its sub-constructions occur as constituents except of $[{}^0Div\ {}^06\ {}^00]$ which is mentioned in (Em) by its constituent ${}^0[{}^0Div\ {}^06\ {}^00]$. Consequently, the second occurrences of ${}^0Div$ and ^{06}, and the occurrence of ^{00} are mentioned in (Em) as well.

The analysis of the whole sentence is:

(*C*) $\lambda w \lambda t\ [{}^0Know_{wt}\ {}^0Charles\ {}^{\mathbf{0}}[[[{}^0Div\ {}^06\ {}^03] = {}^02] \wedge$
$[{}^0Improper\ {}^0[{}^0Div\ {}^06\ {}^00]]]]$

Now all the occurrences of the constituents of (Em) are mentioned in (*C*). The context of Charles's knowing is *hyper-intensional* (or *constructional* in TIL jargon), and a hyper-intensional (i.e., higher-order) context is dominant over lower-order functional (intensional / extensional) contexts.

If a variable is mentioned in *C* then it is not free for substitution, it is ^{0}bound. Consider the (true) sentence

(Dv) "There is a number such that dividing any number by it is improper."

The embedded clause can only be construed as expressing the construction

(1) $[{}^0Improper\ {}^0[{}^0Div\ x\ y]]$.

Now we need to abstract and quantify over ^{0}bound variables *x, y*, which is impossible without some 'pre-processing' of (1). The solution goes via substituting (constructions of) the numbers *v*-constructed by *x*, *y* for the occurrences of *x*, *y* into (1). To this end we use the following functions:

Tr_τ / $(*_1\ \tau)$ – the mapping which takes a number and returns its Trivialisation
Sub_1 / $(*_1 *_1 *_1 *_1)$ – the mapping which takes a construction C_1, a variable x, and a construction C_2 to the resulting construction C_3, where C_3 is the result of substituting C_1 for x in C_2.[5]

Note that there is an essential difference between using the construction Trivialisation and the Tr_τ function. Whereas the construction ${}^0x$ binds the variable x and constructs just x, the variable x is free in the composition $[{}^0Tr_\tau\ x]$ which v-constructs the Trivialisation of the number that v assigns to x.

The analysis of the sentence (Dv) is now as follows:

(Dv') $[{}^0\exists_\tau\ \lambda y\ [{}^0\forall_\tau\ \lambda x\ [{}^0Improper\ [{}^0Sub_1\ [{}^0Tr_\tau\ y]\ {}^0y\ [{}^0Sub_1\ [{}^0Tr_\tau\ x]\ {}^0x\ {}^0[{}^0Div\ x\ y]]]]]]$.

Let v assign 0 to y and 6 to x. Then the sub-construction

$$[{}^0Sub_1\ [{}^0Tr_\tau\ y]\ {}^0y\ [{}^0Sub_1\ [{}^0Tr_\tau\ x]\ {}^0x\ {}^0[{}^0Div\ x\ y]]]$$

v-constructs the *Composition* $[{}^0Div\ {}^06\ {}^00]$, which belongs to the class *Improper*. Since this holds for any valuation of x, (Dv') v-constructs **T**. (Recall that the existential quantifier $\exists_\tau$ / $(o(o\tau))$ is the mapping that returns **T** at a class which is non-empty, otherwise **F**.)

Montague and other intensional logics interpret terms of their language as the respective functions, i.e., set-theoretical mappings. However, these mappings are the outputs of executing the respective procedures. Montague does not make it possible to mention the procedures as objects *sui generis*, and to make thus a semantic shift to hyperintensions. Yet we do need a hyperintensional semantics. Notoriously well-known are attitudinal sentences which no intensional semantics can properly handle, because its finest individuation is equivalence.[6] Typical cases of mentioning constructions are sentences expressing hyper-intensional attitudes which are attitudes to the *meaning* of the embedded clause (see § 3.2).

3.1. Constituents occurring with de dicto / de re supposition.

This difference is closely connected with the distinction between 'using and mentioning functions'. By the latter we mean (roughly): When we use a function f (to point to its value) then we apply f to its argument in order to obtain

---

[5] See Tichý (1988, pp. 74, 75)

[6] See Gamut (1991, p.73)

the value (if any) of *f* at the argument; when mentioning *f* we only talk about the whole function *f*.

For the sake of simplicity we will characterise these two ways only for constructions of α-intensions of type $\alpha_{\tau\omega}$. Generalisation for constructions of mathematical functions can be found in Duží et al. (2007, §4.9).

To adduce an example, compare the sentences:

(S_1) "The President of the USA is a Republican".
(S_2) "G.W. Bush became the President of the USA".[7]

When George W. Bush became the President of the USA he certainly did not become himself (or any other individual), and when he once stops being the President he will not stop being himself. George W. Bush began occupying the office of the President of the USA (PresUSA for short), and will soon stop occupying the office (writing in January 2007). Hence (S_2) relates Bush to the office itself, and 'to become' denotes a $(o\ \iota\ \iota_{\tau\omega})_{\tau\omega}$-object. Using time-honoured terminology, we say that 'The President of the USA' is used with *de re* or *de dicto* supposition in (S_1), (S_2), respectively.

The respective analyses of (S_1) and (S_2) are as follows:

Types: *President* (of something) / $(\iota\iota)_{\tau\omega}$; *Republican* / $(o\iota)_{\tau\omega}$;
Become / $(o\ \iota\ \iota_{\tau\omega})_{\tau\omega}$; *Bush*/ι; *USA*/ι.
Synthesis:
(S_1') $\lambda w\lambda t\ [{}^0Republican_{wt}\ \lambda w\lambda t\ [{}^0President_{wt}\ {}^0USA]_{wt}]$
(S_2') $\lambda w\lambda t\ [{}^0Become_{wt}\ {}^0Bush\ \lambda w\lambda t\ [{}^0President_{wt}\ {}^0USA]]$.

The proposition constructed by (S_1') takes the value **T** in those *w, t* in which the individual that occupies PresUSA belongs to the class of individuals that instantiate the property Republican, and **F** if the individual does not belong to the class. It might seem that in such states of affairs *w, t* where there is no President of the USA the proposition should be false. However, if it were so, the proposition that the President of the USA is not a Republican would have to be true[8], which would in turn entail that there is a President of the USA. Therefore, in those *w, t* where PresUSA is vacant, the proposition has *no truth-value.*[9]

On the other hand, the proposition denoted by (S_2) remains true (false) even in those states of affairs *w, t* where there is no President of the USA. Actu-

[7] Cf. similar examples in Gamut (1991, §§ 6.4.1 – 6.4.3)
[8] See Strawson (1950)
[9] Remember that our logic is a logic of *partial* functions. Once a constituent ($\lambda w\lambda t\ [{}^0\mathrm{President}_{wt}\ {}^0\mathrm{USA}]_{wt}$ in our case) of a compound construction is *v*-improper, the whole Composition is *v*-improper, and the function (here, a proposition) constructed by the respective Closure is undefined at its argument.

ally, its truth-value does not depend on the occupancy of PresUSA at *w, t*. In particular we cannot substitute a construction of the current occupant of the office. If we could do this, we could deduce that Bush became himself.

To characterise the *de dicto / de re* distinction, we quote Tichý:

> Semantically, the difference between the *de dicto* and *de re* amounts to this. Suppose D is a constituent of an application C, D constructs office *D* and C office *C*. If D occurs in C with supposition *de re*, then the occupancy of *C* in a world *W* and at time *T* depends only on the occupancy of *D* in *W* at *T*: it is irrelevant what (if anything) occupies *D* in worlds other than *W* or at times other than *T*. But if D occurs with supposition *de dicto*, the occupancy of *C* in *W* at *T* depends on the occupancy of *D* in all worlds at all times. (1988, p. 216.)

The *de dicto* context is dominant over the *de re* context. Consider another sentence:

(S_3) "If the President of the USA is a Democrat then Charles believes that the President of the USA is Bill Clinton."

An adequate analysis of the consequent has to respect the fact that Charles can believe that the President of the USA is Bill Clinton even if the President is actually George W. Bush, and even if the President does not exist. The proposition that the President of the USA is Bill Clinton is mentioned in this clause, the context of Charles' believing is *intensional* (Believe[10] / $(o\ \iota\ o_{\tau\omega})_{\tau\omega}$):

(S_{3con}) $\lambda w \lambda t\ [{}^0Believe_{wt}\ {}^0Charles\ \lambda w \lambda t\ [\lambda w \lambda t\ [{}^0President_{wt}\ {}^0USA]_{wt} = {}^0Clinton]]$.

The construction expressed by the embedded clause, namely

(S_{3emb}) $\lambda w \lambda t\ [\lambda w \lambda t\ [{}^0President_{wt}\ {}^0USA]_{wt} = {}^0Clinton]$

occurs *de dicto* in (S_{3con}), as well as $\lambda w \lambda t\ [{}^0President_{wt}\ {}^0USA]$. Quoting again from Tichý:

> [I]n general, a *de re* constituent of D is a *de re* constituent of any application in which D appears as a *de re* constituent; a *de re* constituent of D is a *de dicto* constituent of any application in which D appears as a *de dicto* constituent. A *de dicto* constituent is a *de dicto* constituent of any applica-

[10] We conceive believing as a relation-in-intension of an individual to a proposition here. See, however, §3.2.

tion in which D appears as a (*de re* or *de dicto*) constituent. Briefly, *de dicto* is the dominant one of the two suppositions (1988, p. 217).

The sentence (S_3) expresses the construction:

(S_3') $\lambda w \lambda t\ [^0\supset \lambda w \lambda t\ [^0Democrat_{wt}\ \lambda w \lambda t\ [^0President_{wt}\ ^0USA]_{wt}]_{wt}$
$[\lambda w \lambda t\ [^0Believe_{wt}\ ^0Charles\ [\lambda w \lambda t\ [\lambda w \lambda t\ [^0President_{wt}\ ^0USA]_{wt}$
$= {}^0Clinton]]]]_{wt}]$.

Due to the fact that the construction expressed by the antecedent occurs intensionally descended[11] with respect to *w, t*, the first occurrence of $\lambda w \lambda t\ [^0President_{wt}\ ^0USA]$ is with *de re* supposition in (S_3'). But though the construction (S_{3con}) is subjected to the intensional descent, the second occurrence of $\lambda w \lambda t\ [^0President_{wt}\ ^0USA]$ is *de dicto* in (S_{3con}) as well as in (S_3'), because the construction (S_{3emb}) is not subjected to the intensional descent.

Generalising a bit, let *S* be a construction of an α-intension of a form $\lambda w \lambda t$ *C,* and let *D* be a constituent of *C.* We will say that *D* occurs in the *intensional context of C* if the occurrence of *C* is used with *de dicto* supposition in *S*, otherwise *D* occurs in the *extensional context of C.*

Referring for details to, e.g., Duží (2003, 2004), we now recapitulate the two *de re* principles:

Rule of substitution *of congruent constructions de re.* Let $C \rightarrow \alpha_{\tau\omega}$, $D \rightarrow \alpha_{\tau\omega}$ be *v*-congruent constructions, i.e. $C_{wt} = D_{wt}$, and let *S(D/C)* be a construction that arises from *S* by substituting *D* for the *de re* occurrences of *C* in *S.* Then $S_{wt} = S(D/C)_{wt}$.

The rationale behind the Principle is that what is predicated of the occupant of *C* at $\langle W, T \rangle$ is what is predicated of the occupant of *D* at $\langle W, T \rangle$ on condition of co-occupation of *C* and *D* at $\langle W, T \rangle$.

Principle of the existential presupposition. If a construction *C* of an α-office C occurs with *de re* supposition in a hyper-proposition *P*, then *P* has a *presupposition* that C exists (is occupied); $Exist\ /\ (o\alpha_{\tau\omega})_{\tau\omega}$: $\lambda w \lambda t\ [^0Exist_{wt}\ C]$.

Thus the following arguments are valid ($P \rightarrow (o\alpha)_{\tau\omega}$):

$$\frac{\lambda w \lambda t\ [P_{wt}\ C_{wt}]}{\lambda w \lambda t\ [^0Exist_{wt}\ C]} \qquad \frac{\lambda w \lambda t\ \neg[P_{wt}\ C_{wt}]}{\lambda w \lambda t\ [^0Exist_{wt}\ C]}$$

Since the property of existence $Exist\ /\ (o\ \alpha_{\tau\omega})_{\tau\omega}$ (or rather occupancy of an α-office) can be defined by means of the existential quantifier ($x \rightarrow \alpha$, $r \rightarrow \alpha_{\tau\omega}$, $=_\alpha\ /\ (o\alpha\alpha)$)

$$\lambda w \lambda t\ \lambda r\ [^0\exists\ \lambda x\ [^0{=}_\alpha\ x\ r_{wt}]],$$

[11] i.e., τω-extensionally

the conclusion can be equivalently expressed by the construction

$$\lambda w\lambda t\ [{}^0\exists\ \lambda x\ [{}^0{=}_\alpha\ x\ C_{wt}]].$$

3.2. Attitude reports

This section provides a 'taxonomy' and schematic TIL analysis of attitude reports. Further details and discussion can be found, e.g., in Duží (2004) and in Duží & Jespersen & Müller (2005).

a) '*Taxonomy*': B stands for 'believing', 'knowing', etc.; Ch / ι is an agent; $a \rightarrow \iota_{\tau\omega}$ is a subject of the attitude; $P \rightarrow (o\iota)_{\tau\omega}$ is a construction of the property ascribed to a.

I. *Implicit (propositional) attitudes:* $B \rightarrow (o\iota o_{\tau\omega})_{\tau\omega}$

a) De dicto: Ch Bs that *a* is P.

b) De re:

i) *a* is B-ed by Ch to be a P. — *passive variant*

ii) Ch B-s of *a* that *he* (namely *a*) is a P. — *active variant with anaphoric reference he*

II. *Explicit (hyper-propositional) attitudes*: $B^* \rightarrow (o\iota{*}_n)_{\tau\omega}$

a) De dicto: Ch B*s that *a* is P.

b) De re:

i) *a* is B*-ed by Ch to be a P. — *passive variant*

ii) Ch B*-s of *a* that *he* (namely a) is a P. — *active variant with anaphoric reference he*

b) *Analytic schemes.*

***Ad* I)** *Implicit (propositional) attitudes*

I.a) ***de dicto:*** $\lambda w\lambda t\ [B_{wt}\ {}^0Ch\ \lambda w\lambda t\ [P_{wt}\ a_{wt}]]$

I.b i) ***de re passive variant***:

First we specify the coarse-grained logical form of the sentence:
$\lambda w\lambda t\ [{}^0BCP_{wt}\ a_{wt}]$, where BCP / $(o\iota)_{\tau\omega}$ is the property of being B-ed by Ch to be a P. Second, we have to refine the coarse-grained form by defining the property BCP $(x \rightarrow \iota)$:

$${}^0BCP = \lambda w\lambda t\ [\lambda x\ [B_{wt}\ {}^0Ch\ \lambda w\lambda t\ [P_{wt}\ x]]]$$

Third, the logical form[12] of **I.b i**) is obtained by replacing the left-hand side Trivialisation by the right-hand side definition of the property:

$$\lambda w\lambda t\ [[\lambda w\lambda t\ [\lambda x\ [B_{wt}\ {}^0Ch\ \lambda w\lambda t\ [P_{wt}\ x]]]]_{wt}\ a_{wt}],$$

which can be β-reduced to:

$$\lambda w\lambda t\ [\lambda x\ [B_{wt}\ {}^0Ch\ \lambda w\lambda t\ [P_{wt}\ x]]\ a_{wt}].$$

[12] For the definition and details on the notion of logical form, see Duží & Materna (2005), and Duží et al. (2007).

Further 'syntactic' β-reduction would not be valid, because we would substitute the *de re* occurrence of a_{wt} for x into the intensional context (*de dicto*) of $\lambda w \lambda t\ [P_{wt}\ x]$, which is not an equivalent transformation due to partiality (even in case of a substitution that prevents the collision of variables by their renaming).

I.b ii) ***de re active variant:***

First, a coarse-grained analysis:

$$\lambda w \lambda t\ [{}^0B\text{-}of_{wt}\ {}^0Ch\ a_{wt}\ {}^2p],$$

where $B\text{-}of\ /\ (o\iota\iota o_{\tau\omega})_{\tau\omega}$ is an intension relating an individual to another individual and a proposition ${}^2p \to o_{\tau\omega}$.

Second, we have to define the construction of the relation *B-of.* Schematically:

B-of (*x-who, y-whom, that-he=whom is a P*);

$x,y,he \to \iota$, $Sub_1\ /\ (*_1 *_1 *_1 *_1)$, $Tr_\iota\ /\ (*_1 \iota)$:

$${}^0B\text{-}of = \lambda w \lambda t\ \lambda xy\ p\ [{}^0B_{wt}\ x\ {}^2[{}^0Sub_1\ [{}^0Tr_\iota\ y]\ {}^0he\ p]].$$

The Double Execution is necessary here in order to descend from hyperintensional context of the propositional construction (the result of applying the Sub_1 function) to the intensional context of the proposition to which the individual *v*-constructed by y is related.

Third, the analysis of **II.b ii)** is obtained by substituting ${}^0Ch$ for x, a_{wt} for y and ${}^0[\lambda w \lambda t\ [P_{wt}\ he]]$ for p:

$$\lambda w \lambda t\ [{}^0B_{wt}\ {}^0Ch\ {}^2[{}^0Sub_1\ [{}^0Tr_\iota\ a_{wt}]\ {}^0he\ {}^0[\lambda w \lambda t\ [P_{wt}\ he]]]].$$

Note that the substitution of a_{wt} for y is valid here, because the variable y occurs in the extensional context of the above definition.

***Ad* II)** *Explicit (hyper-propositional) attitudes*

II. a) ***de dicto:*** $\lambda w \lambda t\ [B^*_{wt}\ {}^0Ch\ {}^0[\lambda w \lambda t\ [P_{wt}\ a_{wt}]]]$

II. b i) ***de re passive variant:***

First, a course-grained analysis rendering the logical form is $\lambda w \lambda t\ [{}^0B^*CP_{wt}\ a_{wt}]$, where $B^*CP\ /\ (o\iota)_{\tau\omega}$ is the property of being B*ed by Ch to be a P. Next, we have to refine the analysis by defining the property ($x \to \iota$):

$${}^0B^*CP = \lambda w \lambda t\ [\lambda x\ [B^*_{wt}\ {}^0Ch\ [{}^0Sub_1\ [{}^0Tr_\iota\ x]\ {}^0x\ {}^0[\lambda w \lambda t\ [P_{wt}\ x]]]]].$$

Now we have to use the Sub_1 and Tr_ι functions, because the variable x occurs in the hyper-intensional context of ${}^0[\lambda w \lambda t\ [P_{wt}\ x]]$, and it is thus not free for λ-binding. However, we don't need the Double Execution of the result of applying the Sub_1 function, because the agent *Ch* is related directly to the hyper-proposition.

Second, by substituting the above definition of the property, we obtain a fine-grained analysis:

$$\lambda w \lambda t\ [\lambda w \lambda t\ [\lambda x\ [B^*_{wt}\ {}^0Ch\ [{}^0Sub_1\ [{}^0Tr_\iota\ x]\ {}^0x\ {}^0[\lambda w \lambda t\ [P_{wt}\ x]]]]]_{wt}\ a_{wt}],$$

which can be β-reduced to:

$$\lambda w \lambda t\ [\lambda x\ [B^*_{wt}\ {}^0Ch\ [{}^0Sub_1\ [{}^0Tr_\iota\ x]\ {}^0x\ {}^0[\lambda w \lambda t\ [P_{wt}\ x]]]]\ a_{wt}].$$

Further 'syntactic' β-reduction *is* now an equivalent transformation. However, its performing results in the analysis of the active variant *ad* II.b. ii):

$$\lambda w \lambda t\ [B^*_{wt}\ {}^0Ch\ [{}^0Sub_1\ [{}^0Tr_\iota\ a_{wt}]\ {}^0x\ {}^0[\lambda w \lambda t\ [P_{wt}\ x]]]].$$

II.b ii) ***de re active variant*:**

First, a coarse-grained analysis:

$\lambda w \lambda t\ [B^*of_{wt}\ {}^0Ch\ a_{wt}\ p]$; $B\text{-}of\ /\ (o\iota\iota *_n)_{\tau\omega}$, $p \to *_n$, ${}^2p \to o_{\tau\omega}$.

Second, we have to define B^*of (*x-who, y-whom, that-he=whom is a P*):

${}^0B\text{-}of = \lambda w \lambda t\ \lambda xy\ p\ [{}^0B_{wt}\ x\ [{}^0Sub_1\ [{}^0Tr_\iota\ y]\ {}^0he\ p]]$.

Third, the analysis of **II.b ii)** is obtained by substituting ${}^0Ch$ for x, a_{wt} for y, and ${}^0[\lambda w \lambda t\ [P_{wt}\ he]]$ for p, which is correct even in case of a_{wt} being v-improper, because y occurs in the extensional context of the above definition:

$$\lambda w \lambda t\ [B^*_{wt}\ {}^0Ch\ [Sub_1\ [{}^0Tr_\iota\ a_{wt}]\ {}^0he\ {}^0[\lambda w \lambda t\ [P_{wt}\ he]]]].$$

***Remark*:**

It is easy to prove that *de re* and *de dicto* attitudes are logically independent, neither the *de re* case is entailed by the respective *de dicto* variant*,* nor vice versa.

However, if a is a *rigid designator* of an individual, $a \to \iota$ and a is *v-proper* for any v, then in case I. the *de dicto* and *de re* attitudes are logically equivalent, whereas in case II. it is not so: in this case only the active and passive variant of the *de re* attitude are logically equivalent.

Case I. *Implicit propositional attitudes*:

$$\lambda w \lambda t\ [B_{wt}\ {}^0Ch\ \lambda w \lambda t\ [P_{wt}\ a]] =$$
$$\lambda w \lambda t\ [\lambda x\ [B_{wt}\ {}^0Ch\ \lambda w \lambda t\ [P_{wt}\ x]]\ a] =$$
$$\lambda w \lambda t\ [B_{wt}\ {}^0Ch\ {}^2[{}^0Sub_1\ [{}^0Tr_\iota\ a]\ {}^0x\ {}^0[\lambda w \lambda t\ [P_{wt}\ x]]]]$$

Case II. *Explicit hyper-propositional attitudes:*

$$\lambda w \lambda t\ [B^*_{wt}\ {}^0Ch\ {}^0[\lambda w \lambda t\ [P_{wt}\ a]]] \neq$$
$$\lambda w \lambda t\ [\lambda x\ [B^*_{wt}\ {}^0Ch\ [{}^0Sub_1\ [{}^0Tr_\iota\ x]\ {}^0x\ {}^0[\lambda w \lambda t\ [P_{wt}\ x]]]\ a] =$$
$$\lambda w \lambda t\ [B^*_{wt}\ {}^0Ch\ [{}^0Sub_1\ [{}^0Tr_\iota\ a]\ {}^0x\ {}^0[\lambda w \lambda t\ [P_{wt}\ x]]]].$$

For, the hyperpropositions to which Ch is related are not identical:

$${}^0[\lambda w \lambda t\ [P_{wt}\ a]]] \neq [{}^0Sub_1\ [{}^0Tr_\iota\ a]\ {}^0x\ {}^0[\lambda w \lambda t\ [P_{wt}\ x]]];$$

they are only equivalent on the assumption of a being proper:

$$[\lambda w \lambda t\ [P_{wt}\ a]]] = {}^2[{}^0Sub_1\ [{}^0Tr_\iota\ a]\ {}^0x\ {}^0[\lambda w \lambda t\ [P_{wt}\ x]]].$$

The reason for the above non-identity consists in the fact that while $[{}^0Tr_\iota\ a]$ v-constructs the Trivialisation of the individual v-constructed by a*,* a itself may be substituted for by a composed construction v-constructing the same individual.

3.3. Scheme of TIL rules of inference

Summarising, we are going to provide a scheme of valid TIL rules of inference. First, we define identity, equivalency and *v*-congruency of constructions.
Let $C, D \,/\, *_n$, $C, D \rightarrow \alpha$ be constructions, $=_\beta \,/\, (o\beta\beta)$ the identity of β-entities. Now we use the following notational abbreviations:
C(y) – a construction with a free variable *y*
C(D/y) – the result of a collisionless substitution of *D* for *y* in *C*;
C(D'/D) – the result of the collisionless replacement of *D* by *D'* in *C*;
v-Improper(*A*)/*v-Proper*(*A*) – the construction *A* is/is not *v*-improper for a valuation *v*;
Improper(*A*)/ *Proper*(*A*) – the construction *A* is *v*-improper/*v*-proper for all valuations *v*.
Definition: *C, D* are *v-congruent* iff either *C* and *D* *v*-construct the same α-entity, $C =_\alpha D$, or both *C* and *D* are *v*-improper; *C, D* are *equivalent* iff *C, D* are *v*-congruent for all valuations *v*; *C, D* are *identical* iff $^0C =_{*n} {}^0D$.
Claim **(β-*reduction* ‘*by value*’):**
The Composition $(1 \le i \le m,\ x_i \rightarrow_v \beta_i,\ D_i \rightarrow_v \beta_i,\ Y \rightarrow_v \alpha)$
(Ap) $[[\lambda x_1 ... x_m\ Y]\ D_1 ... D_m]$
is equivalent to the (computationally) reduced construction
(Apβ) $^2[^0Sub_n\ [^0Tr_{\beta 1}\ D_1]\ ^0x_1\ [^0Sub_n\ [^0Tr_{\beta 2}\ D_2]\ ^0x_2 ...$
$[^0Sub_n\ [^0Tr_{\beta m}\ D_m]\ ^0x_m\ ^0Y]...]]$.
Proof:
a) According to the definition of Closure, Composition and Double Execution, the construction (Ap) is *v*-improper iff for some i $(1 \le i \le m)$ D_i is *v*-improper. Then (Apβ) is *v*-improper as well.
b) Let D_i be *v*-proper for all *i*, $1 \le i \le m$, and let D_i *v*-construct entities a_i, respectively. Then according to the definition of Composition and Closure, (Ap) is either $v(a_i/x_i)$-improper if *Y* is $v(a_i/x_i)$-improper, or (Ap) *v*-constructs what is $v(a_i/x_i)$-constructed by *Y*. In other words, (Ap) is $v(a_i/x_i)$-congruent with *Y*. Now, the result of applying the respective substitutions in (Apβ) *m*-times is the construction that is also $v(a_i/x_i)$-congruent with *Y*. Therefore, (Apβ) and (Ap) are $v(a_i/x_i)$-congruent.

Since (Ap), (Apβ) are thus *v*-congruent for any valuation *v*, they are equivalent.
As a consequence, we can now formulate particular rules in more details.
Types: $y \rightarrow_v \beta$, $D \rightarrow_v \beta$, $C(y) \rightarrow_v \alpha$, $\lambda y\ C(y) \rightarrow_v (\alpha\beta)$, $[[\lambda y\ C(y)]\ D] \rightarrow_v \alpha$.
a) ***Closure***: *Proper*($[\lambda y\ C(y)]$)
b) ***Compositionality*** and β-***rule***:

$$Comp \qquad \frac{v\text{-}Improper(D)}{v\text{-}Improper([\lambda y\ C(y)\ D])}$$

$$\textit{Subst} \quad \frac{v\text{-}Improper(D)}{v\text{-}Improper({}^2[{}^0Sub\ [{}^0Tr\ D]\ {}^0y\ {}^0C(y)])}$$

$$\beta\textit{-rule} \quad \frac{v\text{-}Proper(D)}{{}^2[{}^0Sub\ [{}^0Tr\ D]\ {}^0y\ {}^0C(y)] = [[\lambda y\ C(y)]\ D] = C(D/y)}$$

c) *Rules of valid* ***Substitution (Leibniz's law).***

i) Let $C \to (\alpha\ \beta_1...\beta_m)$, $m \geq 0$, be a constituent of D and let C occur in $D\,(\beta_1...\beta_m)$-extensionally[13]. Let $D_1 \to \beta_1,..., D_m \to \beta_m$, and let $[C\,D_1 ... D_m]$, $[C'\,D_1 ... D_m]$ be v-congruent. Then $D([C'\,D_1 ... D_m]\ /[C\,D_1 ... D_m])$ is v-congruent with D.

ii) Let $C \to (\alpha\ \beta_1...\beta_m)$, $m \geq 0$, be a constituent of D and let C occur in $D\,(\alpha\beta_1...\beta_m)$-intensionally. Then if C and C' are equivalent, D and $D(C'/C)$ are equivalent as well.

iii) Let the occurrence of C be mentioned* in D and let ${}^0C = {}^0C'$. Then D and $D({}^0C'/{}^0C)$ are equivalent.

Note that the *de re* rule of existential presupposition is a special case of the *Comp* rule, and the *de re* rule of substitution of congruent constructions is a special case of i).

4. Conclusion

Logic should help to find the objective structures underlying expressions of a language, and it should be now clear how 'value gaps' can be accommodated *via* improper constructions and partial functions, and it's also very clear why we must accept improperness and partiality: when modelling entities the empirical expressions talk about by intensions, functions from possible worlds, these functions have to be partial, for there are intensions we talk about that do not have a value in particular w at time t. TIL handling partiality is determined by the above principles that turn on the same conception of language. A piece of language serves to point to a logical construction beyond itself, its sense. Our semantics runs smoothly even with partial functions and improper constructions that are used / mentioned in (hyper-)intensional contexts.[14]

[13] It means that C occurs in the Composition $[C\ D_1'...D_m']$ for some $D_1' \to \beta_1,...,D_m' \to \beta_m$ and the Composition does not occur in another intensional context of D.

[14] I am grateful to Jan Kuchyňka from Masaryk University of Brno and Bjorn Jespersen from Delft University of Technology for their valuable comments.

Doc. RNDr. Marie Duží, CSc.
VSB-Technical University Ostrava
17. listopadu 15
708 33 Ostrava, Czech Republic
marie.duzi@vsb.cz

References

Church, A. (1956). *Introduction to Mathematical Logic I.* Princeton: Princeton University Press.

Cresswell, M. J. (1975). Hyperintensional Logic. *Studia Logica, 34*, 25–38.

Duží, M. (2003). Do we have to deal with partiality? In *Miscellanea Logica,* vol. 5 (pp. 45–76). edited by K.Bendová & P. Jirků. Praha: Karolinum. Retrievable at *http://logika.ff.cuni.cz/newpages/papers/misclogV.pdf#duzi.partiality*

Duží, M. (2004). Intensional logic and the irreducible contrast between *de dicto* and *de re*. *ProFil, 5*, 1–34. Retrievable at *http://profil.muni.cz/01_2004/duzi_de_dicto_de_re.pdf*

Duží, M., & Jespersen, B., & Materna, P. (2007): *Transparent Intensional Logic (Foundations and Applications).* Manuscript.

Duží, M. & Jespersen, B. & Müller, J. (2005). Epistemic Closure and Inferable knowledge. In L. Běhounek, M. Bílková (Eds.). *The Logica Yearbook 2004* (pp. 125–140). Praha: Filosofia.

Duží, M., & Materna, P. (2005). Logical Form. In G. Sica (Ed.). *Essays on the Foundations of Mathematics and Logic, Vol. 1* (pp. 115–153). Monza: Polimetrica International Scientific Publisher.

Gamut, L. T. F. (1991). *Logic, Language and Meaning.* Volume II. Intensional Logic and Logical Grammar. Chicago/London: The University of Chicago Press.

Materna, P. (1998). *Concepts and Objects.* Acta Philosophica Fennica, Vol. 63. Helsinki.

Materna, P. (2004). *Conceptual Systems.* Berlin: Logos Verlag.

Materna, P., & Duží, M. (2005). Parmenides Principle. *Philosophia, philosophical quarterly of Israel, vol. 32*(1–4). 155–180.

Moschovakis, Y. N. (1994). Sense and Denotation as Algorithm and Value. In J. Väänänen & J. Oikkonen (Eds.). *Lecture Notes in Logic, 2*, 210–249.

Moschovakis, Y. N. (2006). A logical calculus of meaning and synonymy. *Linguistic and Philosophy, 29,* 27–89.

Strawson, P. F. (1950). On Referring. *Mind, 59*, 320–344.

Tichý, P. (1988). *The Foundations of Frege's Logic*. Berlin, New York: DeGruyter.

Tichý, P. (2004). *Pavel Tichý's Collected Papers in Logic and Philosophy.* Svoboda, V., & Jespersen, B., & Cheyne, C. (Eds.). Prague: Filosofia and Dunedin: University of Otago Press.

Fuzzy Type Theory as a Tool for Linguistic Analysis

Antonín Dvořák and Vilém Novák

1. Introduction

In this contribution we will present *Fuzzy type theory* (FTT) and some of its applications, which can be interesting for philosophers and linguists. Fuzzy type theory is a logical system originally proposed in (Novák, 2005a). It is a generalization of the classical simple type theory developed particularly by (Church, 1940) and (Henkin, 1950). Type theory is a basis of various systems of *intensional logics*, (see eg. (Fitting, 2006)) which proved to be very useful in the analysis of natural languages. However, most of these systems do not incorporate the *vagueness* phenomenon (cf. (Dvořák & Novák, 2005)). The latter has been most successfully treated by *fuzzy logic* which is now a well-developed formal system (Hájek, 1998; Novák, Perfilieva, & Močkoř, 1999) with numerous applications in mathematics, computer science, industry etc. It turned out that for the successful applications of ideas of fuzzy logic in linguistics, higher-order logical system is a necessity. Fuzzy type theory is an extension both of classical type theory as well as first-order fuzzy logic.

In this contribution we briefly present basic building blocks of fuzzy type theory. Its syntax is traditional, i.e., formulas can be either provable or non-provable. However, the semantics of FTT is non-classical, i.e., a constituent of frame for its language is a set of multiple truth values and so, formulas of type o (truth value) attain more than two truth degrees. The completeness theorem with respect to Henkin-style general models holds in FTT. We also discuss the importance of *fuzzy equality*. In classical type theory, we can start with a logical constant Q denoting the identity relation, and define all logical connectives, quantifiers etc. by means of Q. Similar construction is used in fuzzy type theory too. Fuzzy equality is then used in the definition of the important concept of the *extensionality* of functions.

We will also present some applications of fuzzy type theory. We cannot go into details, but show how some important notions can be expressed using formal means of fuzzy type theory. We mainly concentrate on the analysis of the so called IF-THEN rules and linguistic expressions which occur in them. Finally we mention a specific deduction method called *perception-based logi-*

cal deduction that is a deduction over a set of linguistically characterized fuzzy IF-THEN rules.

Recently, the use of FTT in the study of generalized quantifiers has been proposed, see (Novák, 2006). It provides a unified treatment of the so-called *intermediate quantifiers*, e.g. *a few, a great deal of, most, many*, etc.

There is a connection between fuzzy type theory and *fuzzy class theory* (FCT) developed in (Běhounek & Cintula, 2005). The goals of both theories, however, are basically different. The main goal of FCT is to establish precise grounds for fuzzy mathematics while the main goal of FTT is to develop a powerful formal system for modeling of the semantics of (parts of) natural language. Note, however, that FTT can serve well for both goals.

2. Fuzzy type theory

For full treatment of FTT we refer particularly to (Novák, 2005a). Here we present some important building blocks of it. The purpose is to provide an overall idea.

Structure of truth values. The structure of truth values is generally supposed to form one of the following: a complete IMTL$_\Delta$-algebra (see Esteva & Godo, 2001), standard Łukasiewicz$_\Delta$ algebra, ŁΠ-algebra or BL-algebra. The most important for applications in linguistics is Łukasiewicz$_\Delta$ algebra

$$\mathcal{L} = \langle [0, 1], \vee, \wedge, \otimes, \oplus, \Delta, \rightarrow, 0, 1 \rangle$$

where

$$\begin{array}{ll} \wedge = \text{minimum}, & \vee = \text{maximum}, \\ a \otimes b = 0 \vee (a + b - 1), & a \rightarrow b = 1 \wedge (1 - a + b), \\ \neg a = a \rightarrow 0 = 1 - a, & a \oplus b = 1 \wedge |a - b|, \\ \Delta(a) = \begin{cases} 1 & \text{if } a = 1, \\ 0 & \text{otherwise.} \end{cases} & \end{array}$$

Fuzzy equality. Important concept in FTT is that of a *fuzzy equality*. This is a fuzzy relation

$$\doteq : M \times M \rightarrow L$$

which fulfils the following properties:

(i) reflexivity $\quad [m \doteq m] = 1,$

(ii) symmetry $[m \doteq m'] = [m' \doteq m]$,
(iii) $\otimes$-transitivity $[m \doteq m'] \otimes [m' \doteq m''] \leq [m \doteq m'']$

for all $m, m', m'' \in M$ where $[m \doteq m']$ denotes a *truth value* of $m \doteq m'$.

A special case of fuzzy equality on the algebra of truth values is *biresiduation* $a \leftrightarrow b = (a \to b) \wedge (b \to a)$, $a, b \in L$. This operation is a natural interpretation of many-valued equivalence. Example of a fuzzy equality on $M = \mathbb{R}$ with respect to standard Łukasiewicz algebra is

$$[m \doteq n] = 1 - (1 \wedge |m - n|), \qquad m, n \in \mathbb{R}.$$

Let $F : M_\alpha \to M_\beta$ be a function and $=_\alpha$, $=_\beta$ be fuzzy equalities in the respective domains M_α and M_β. Then F is extensional w.r.t $=_\alpha$ and $=_\beta$ if there is a natural number $q \geq 1$ such that

$$[m =_\alpha m']^q \leq [F(m) =_\beta F(m')], \qquad m, m' \in M_\alpha$$

where the power is taken with respect to $\otimes$. If $q = 1$, we say that F is *strongly extensional*. It is *weakly extensional* if

$$[m =_\alpha m'] = 1 \quad \text{implies that} \quad [F(m) =_\beta F(m')] = 1.$$

This is equivalent to the condition

$$\Delta[m =_\alpha m'] \leq [F(m) =_\beta F(m')].$$

It is easy to prove that each fuzzy equality $=_\alpha$ (as a binary function) is strongly extensional w.r.t. itself and $\leftrightarrow$. The $\leftrightarrow$ is a fuzzy equality on L strongly extensional w.r.t. itself; $\wedge$ is strongly, and Δ is weakly extensional w.r.t. $\leftrightarrow$.

Basic syntactical elements. The *Types* is a set of types constructed iteratively from the atomic types ε (elements) and o (truth values). $Form_\alpha$ denotes a set of formulas of type $\alpha \in Types$ which is the smallest set satisfying:

(i) Variables $x_\alpha \in Form_\alpha$ and constants $c_\alpha \in Form_\alpha$,
(ii) if $B \in Form_{\beta\alpha}$ and $A \in Form_\alpha$ then $(BA) \in Form_\beta$ *(application),*
(iii) if $A \in Form_\beta$ then $\lambda x_\alpha A \in Form_{\beta\alpha}$ *(abstraction).*

If $A \in Form_\alpha$ is a formula of type $\alpha \in Types$ then we write A_α. Note that variables, constants and the above defined sequences are formulas (alternatively, they are often called lambda-terms in type theory).

Formulas of type o (truth value) can be joined by the following connectives: $\equiv$ (equivalence), $\vee$ (disjunction), $\wedge$ (conjunction), & (strong conjunction), ∇ (strong disjunction), $\Rightarrow$ (implication). General ($\forall$) and existential ($\exists$) quantifiers are defined as special formulas. For the details about their definition and semantics – see (Novák, 2005a).

If $A \in Form_{o\alpha}$ then A represents a fuzzy set of elements. It can also be understood as a first-order property of elements of the type α. Similarly, $A_{(o\alpha)\alpha}$ is a fuzzy relation (between elements of type α).

Logical axioms. Because of lack of space, we will present logical axioms of (Łukasiewicz) FTT without more detailed explanation:

$$
\begin{array}{rl}
(\mathrm{FT}_I1) & \Delta(x_\alpha \equiv y_\alpha) \Rightarrow (f_{\beta\alpha}\, x_\alpha \equiv f_{\beta\alpha}\, y_\alpha) \\
(\mathrm{FT}_I2_1) & (\forall x_\alpha)(f_{\beta\alpha}\, x_\alpha \equiv g_{\beta\alpha}\, x_\alpha) \Rightarrow (f_{\beta\alpha} \equiv g_{\beta\alpha}) \\
(\mathrm{FT}_I2_2) & (f_{\beta\alpha} \equiv g_{\beta\alpha}) \Rightarrow (f_{\beta\alpha}\, x_\alpha \equiv g_{\beta\alpha}\, x_\alpha) \\
(\mathrm{FT}_I3) & (\lambda x_\alpha B_\beta)A_\alpha \equiv C_\beta \quad (\textit{lambda conversion}) \\
(\mathrm{FT}_I4) & (x_\varepsilon \equiv y_\varepsilon) \Rightarrow ((y_\varepsilon \equiv z_\varepsilon) \Rightarrow (x_\varepsilon \equiv z_\varepsilon)) \\
(\mathrm{FT}_I5) & (x_o \equiv y_o) \equiv ((x_o \Rightarrow y_o) \wedge (y_o \Rightarrow x_o)) \\
(\mathrm{FT}_I6) & (A_o \equiv \top) \equiv A_o \\
(\mathrm{FT}_I7) & A_o \Rightarrow (B_o \Rightarrow A_o) \\
(\mathrm{FT}_I8) & (A_o \Rightarrow B_o) \Rightarrow ((B_o \Rightarrow C_o) \Rightarrow (A_o \Rightarrow C_o)) \\
(\mathrm{FT}_I9) & (\neg B_o \Rightarrow \neg A_o) \equiv (A_o \Rightarrow B_o) \\
(\mathrm{FT}_I10) & A_o \vee B_o \equiv B_o \vee A_o \\
(\mathrm{FT}_I11) & A_o \wedge B_o \equiv B_o \wedge A_o \\
(\mathrm{FT}_I12) & A_o \wedge B_o \Rightarrow A_o \\
(\mathrm{FT}_I13) & (A_o \wedge B_o) \wedge C_o \equiv A_o \wedge (B_o \wedge C_o) \\
(\mathrm{FT}_I14) & (g_{oo}(\Delta x_o) \wedge g_{oo}(\neg\Delta x_o)) \equiv (\forall y_o) g_{oo}(\Delta y_o) \\
(\mathrm{FT}_I15) & \Delta(A_o \wedge B_o) \equiv \Delta A_o \wedge \Delta B_o \\
(\mathrm{FT}_I16) & (\forall x_\alpha)(A_o \Rightarrow \mathrm{B}_o) \Rightarrow (A_o \Rightarrow (\forall x_\alpha)B_o) \\
 & \qquad\qquad\qquad\qquad \text{where } x_\alpha \text{ is not free in } A_o. \\
(\mathrm{FT}_I17) & \iota_{\alpha(o\alpha)}(\mathbf{E}_{(o\alpha)\alpha}\, y_\alpha) \equiv y_\alpha, \qquad \alpha = o, \varepsilon
\end{array}
$$

The $\iota_{\alpha(o\alpha)}$ is a description operator which assigns to a fuzzy set an element from its kernel (i.e., its interpretation is a the *defuzzification operation*). $\mathbf{E}_{(o\alpha)\alpha}$ is a special constant which represents fuzzy equality.

Inference rules and provability. FTT has two inference rules:

(R) *Let $A_\alpha \equiv A'_\alpha$ and $B \in Form_o$. Then, infer B' where B' comes from B by replacing one occurrence of A_α, which is not preceded by λ, by A'_α.*

(N) *Let $A_o \in Form_o$. Then infer ΔA_o from A_o.*

The provability is classical. A theory T is a set of formulas of type o. A formula ΔA_o is crisp, i.e., its interpretation is either 0 or 1. There are formulas which are not crisp.

Semantics. Semantics of FTT is a generalization of the semantics of classical type theory. Let D be a set of objects and L be a set of truth values. A *basic frame* is a system of sets $(M_\alpha)_{\alpha\in\text{Types}}$ where $M_\varepsilon = D$ is a set of objects, $M_o = L$ is a set of truth values and if $\gamma = \beta\alpha$ then $M_\gamma \subseteq M_\beta^{M_\alpha}$. A *frame* is a system

$$\mathcal{M} = \langle (M_\alpha, =_\alpha)_{\alpha\in\text{Types}}, \mathcal{L}\rangle$$

where $\mathcal{L}$ is the algebra of truth values and $=_\alpha$ is a fuzzy equality on M_α and for $\alpha \neq o, \varepsilon$, each function $F \in M_\alpha$ is weakly extensional.

A *general model* is a frame such that every formula A_α, $\alpha \in$ *Types*, has interpretation in it (i.e. there is an element in the corresponding set M_α of the frame that interprets A_α).

Because of lack of space, we will omit precise definition of interpretation of formulas. The reader may find it in (Novák, 2005a). Let us remark only that the formula of the form $\lambda x_\alpha A_\beta$ is in $\mathcal{M}$ interpreted as a function assigning to every $m \in M_\alpha$ an element from M_β that is obtained as an interpretation of A_β in which all occurrences of x_α are replaced by the corresponding m. For example, interpretation of $\lambda x_\alpha A_o$ is a fuzzy set on M_α determined by the property represented by the formula A_o.

Completeness holds with respect to Henkin general models.

Theorem 1 ((Novák, 2005a))
$T \vdash A_o$ iff $T \models A_o$ holds for every theory T and a formula A_o.

3. Evaluating linguistic expressions

Evaluating linguistic expressions (or, simply, evaluating expressions) are expressions of natural language, for example, *small, medium, big, about twenty five, roughly one hundred, very short, more or less deep, not very tall, roughly warm or medium hot, quite roughly strong, roughly medium size*, and many others. They form a small but very important part of natural language and they are present in its everyday use any time. The reason is that people very often need to evaluate phenomena around them. Moreover, they often make important decisions based on them, learn how to control, and many other activities. Therefore, it seems to be very important to study them.

All the details about formal theory of evaluating linguistic expressions can be found in (Novák, (to appear) 2006). As usual, we distinguish intension (a property), and extension in a given *context* of use (i.e., a *possible world*; see Fitting, 2006).[1]

Natural language expressions are, in general, names of *intensions*. Mathematical representation of an intension is a function defined on a set of contexts which assigns to each context a fuzzy set of elements. Intension leads to different truth values in various contexts but is invariant with respect to them.

Extension of a natural language expression is a class of elements (i.e., a fuzzy set) determined by the intension, that fall into meaning of the former *in the given context*. It depends on the particular context of use and changes whenever the context is changed. For example, the expression "high" is a name of an intension being a property of some feature of objects, i.e. of their height. Its meaning can be, e.g., 30 cm when a beetle needs to climb a straw, 30 m for an electrical pylon, but 4 km or more for a mountain.

The global characteristics of the meaning of *pure evaluating expressions*[2] are the following:

(i) Extensions are classes of elements taken from nonempty, linearly ordered and bounded scale which represents *context* of use of the evaluating expressions. In each context, three distinguished limit points can be determined: *left bound*, *right bound*, and a *central point*.
(ii) Each of the above limit points is a starting point of some horizon running towards the next limit point in the sense of the ordering and vanishing beyond. Thus, three horizons can be distinguished on each scale, namely *left, right* and *middle* one. Each horizon is determined by a reasoning analogous to that leading to the sorites paradox (Dvořák & Novák, 2005).
(iii) Extension of any evaluating expression is delineated by a specific horizon resulting from a shift of the horizon due to item (ii). The modification corresponds to a linguistic hedge and is "small for big truth values" and "big for small ones".
(iv) Each scale is vaguely partitioned by the fundamental evaluating trichotomy consisting of a pair of antonyms, and a middle member (typically, "small, medium, big"). Any element of the scale is contained in extensions of at most two neighboring expressions from this trichotomy.

A formal logical theory of evaluating linguistic expressions T^{Ev} in FTT is

[1] We follow the possible world semantics. In the theory of evaluating linguistic expressions, however, it is more convenient to replace the general term "possible world" by a more apt term "context".
[2] Pure evaluating expression has the structure ⟨linguistic hedge⟩ ⟨atomic evaluating expression⟩, where linguistic hedges are e.g. *very*, *more or less* and atomic evaluating expressions are *small*, *medium* and *big*.

constructed on the basis of the above characteristics. The language J^{Ev} of T^{Ev} enables us to express formally notions of context, horizon, etc.

We are going to omit details of the formal axiomatic treatment of evaluating linguistic expressions. Interpretation of basic items of the language J^{Ev} and special formulas including extensions of evaluating expressions is schematically depicted on Fig. 1.

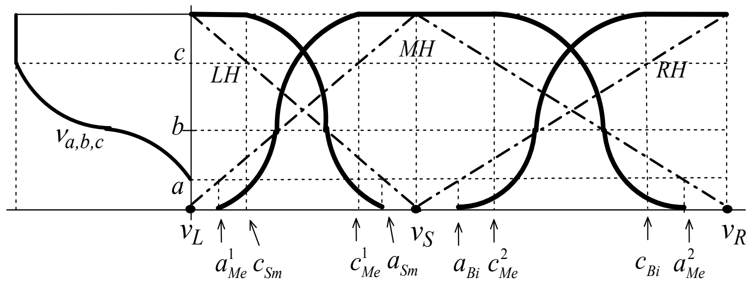


Figure 1: Scheme of the construction of extensions of evaluating expressions

In this picture, *LH*, *MH* and *RH* are interpretations of left, medium and right horizon, respectively, from item (ii) of the above list. The v_L, v_S and v_R are interpretations of left bound, central point and right bound, respectively, of the scale considered in item (i).

In our theory, we do not need to introduce a special elementary type for the context. Instead, we will assign it a formula $w_{\alpha o}$ of type αo, i.e., its interpretation is a function from the set of truth values to arbitrary objects of some type α. This definition is motivated by the idea that people keep in mind a certain image of a bounded scale which they modify according to the concrete situation. We will use a symbol ω as a (meta-)type for context.

To deal with elements of the context, we also need to introduce the inverse formula

$$w^{-1} \equiv \lambda y \cdot \iota_{o(oo)}(\lambda t \cdot y \equiv wt).$$

By the definition, $w^{-1}y$ is the truth value $t \in Form_o$ for which $y \equiv wt$ is true (provable) in the degree 1, and which is chosen using the description operator $\iota_{o(oo)}$. Clearly, $w^{-1} \in Form_{o\alpha}$.

To see how fuzzy type theory is utilized in the formal treatment of evaluating expressions, we present as an example formulas for the representation of them.

(i) *S-formula:* $Sm \coloneqq \lambda \mathbf{v}\, \lambda w\, \lambda x \cdot \mathbf{v}(LH\, w^{-1}x)$,
(ii) *M-formula:* $Me \coloneqq \lambda \mathbf{v}\, \lambda w\, \lambda x \cdot \mathbf{v}(MH\, w^{-1}x)$,
(iii) *B-formula:* $Bi \coloneqq \lambda \mathbf{v}\, \lambda w\, \lambda x \cdot \mathbf{v}(RH\, w^{-1}x)$.

where $LH, MH, RH \in Form_{oo}$ are special formulas representing the three above considered horizons and $\mathbf{v} \in Form_{oo}$ is a linguistic hedge. One can see that $Sm, Me, Bi \in Form_{o\alpha}$ where the type α is given by the chosen context $w \in Form_{\alpha o}$. It is also important to note that among possible linguistic hedges we include also the *empty hedge*, i.e., the evaluating expressions "small, medium, big" are taken as having the form "empty hedge ⟨atomic expression⟩". This approach enables us to develop a unified formal theory. Further, evaluating predications are linguistic expressions of the form

$$X \text{ is } \langle\text{linguistic hedge}\rangle\langle\text{atomic expression}\rangle$$

where atomic expression is one of "small, medium, big". Intensions of evaluating predications are formulas of type $(o\alpha)(\alpha o)$ defined by:

$$\mathrm{Int}(X \text{ is } \langle\text{linguistic hedge}\rangle\text{small}) := \lambda w\, \lambda x \cdot Sm_{\mathbf{v}} wx,$$

$$\mathrm{Int}(X \text{ is } \langle\text{linguistic hedge}\rangle\text{medium}) := \lambda w\, \lambda x \cdot Me_{\mathbf{v}} wx,$$

$$\mathrm{Int}(X \text{ is } \langle\text{linguistic hedge}\rangle\text{big}) := \lambda w\, \lambda x \cdot Bi_{\mathbf{v}} wx.$$

Extensions of evaluating predications are given as follows: let $w \in Form_{\alpha o}$ be a context and X be a variable representing objects of type α. Then

$$\mathrm{Ext}_w\,(X \text{ is } \mathcal{A}) \equiv \mathrm{Int}(X \text{ is } \mathcal{A})w \equiv \lambda x \cdot Ev_{\mathbf{v}} wx$$

where $Ev \in Form_{\varphi}$ is a general metavariable for intension of an evaluating predication, and $\varphi := (o\alpha)(\alpha o)$. It means that extension of the evaluating predication "X is $\mathcal{A}$" is a fuzzy set of elements of type α.

4. IF-THEN rules and perception-based logical deduction

In this section, we outline logical treatment of so-called IF-THEN rules. Its theory is heavily dependent on the theory of evaluating linguistic expressions presented in the previous section. Further we show a method called perception-based logical deduction, which serves as a tool for deduction over IF-THEN rules. Its present and perspective applications are numerous.

A fuzzy IF-THEN rule is a linguistic expression of the form

$$\mathcal{R} := \text{IF } X \text{ is } \mathcal{A} \text{ THEN } Y \text{ is } \mathcal{B} \tag{1}$$

where $\mathcal{A}, \mathcal{B}$ are evaluating expressions. The linguistic predication 'X is $\mathcal{A}$' is called *antecedent* and 'Y is $\mathcal{B}$' is called *consequent*.

Intension of a fuzzy IF-THEN rule $\mathcal{R}$ from (1) is the formula

$$\mathrm{Int}(\mathcal{R}) := \lambda w\, \lambda w' \cdot \lambda x\, \lambda y \cdot Ev^{A}\, wx \Rightarrow Ev^{C}\, w'y \tag{2}$$

where $x \in Form_{\alpha}$, $y \in Form_{\beta}$ represent objects of, possibly, different types and $w \in Form_{\alpha o}$, $w' \in Form_{\beta o}$ are the corresponding contexts. The symbols Ev^{A}, Ev^{C} denote intensions of the predications in the antecedent and consequent, respectively. We will also use a special (meta-)type $\rho := ((o\alpha)\alpha)\omega\omega$ for formulas being intensions of fuzzy IF-THEN rules of the form (2).

A *linguistic description* is a finite set $LD = \{\mathrm{Int}(\mathcal{R}_j)\, j = 1,\dots, m\}$ of (intensions of) fuzzy IF-THEN rules (1). In linguistic theory, there are important notions of *topic* and *focus*, see (Hajičová, Partee, & Sgall, 1998). Topic of linguistic description is a set of evaluating expressions $Topic^{\mathrm{LD}} = \{Ev_j^{A} \mid j = 1,\dots, m\}$ and focus is $Focus^{\mathrm{LD}} = \{Ev^{\mathrm{C}} \mid j = 1,\dots, m\}$. In general, we may take topic and focus as arbitrary (finite) sets of linguistic expressions.

Note that we can formally represent linguistic description, its topic and focus using special crisp formulas of FTT as follows:

$$LD \equiv \lambda z_{\rho} \cdot \bigvee_{j=1}^{m} \Delta(z_{\rho} \equiv \mathrm{Int}(\mathcal{R}_j)),$$

$$Topic^{LD} \equiv \lambda z_{\varphi} \cdot \bigvee_{j=1}^{m} \Delta(z_{\varphi} \equiv Ev_j^{A}),$$

$$Focus^{LD} \equiv \lambda z_{\varphi} \cdot \bigvee_{j=1}^{m} \Delta(z_{\varphi} \equiv Ev_j^{C}),$$

Let $x \in Form_{\alpha}$, $y \in Form_{\beta}$, $w \in Form_{\alpha o}$, $w' \in Form_{\beta o}$. Then the following scheme is a special inference rule of *perception-based logical deduction*:

$$r_{PbLD}: \frac{LPerc^{LD}\, wx\, Ev_i^{A}, \quad LD}{Eval\, w'\, \hat{y}_i\, Ev_i^{C}}$$

where $\hat{y}_i \equiv \iota_{\beta(o\beta)}(\lambda y \cdot Ev_i^{A}\, wx \Rightarrow Ev_i^{C}\, w'\, y)$, $i \in \{1,\dots,m\}$, $T \vdash Topic^{LD}\, Ev_i^{A}$ and $\mathrm{T} \vdash Focus^{LD}\, Ev_i^{C}$. The formula $LPerc^{LD}\, wx\, Ev_i^{A}$ says that Ev_i^{A} is a *perception* of x in the context w. The formula $Eval\, w'\, \hat{y}_i\, Ev_i^{C}$ means that element $\hat{y}_i$ is *evaluated* by Ev_i^{C} (i.e., it has a property expressed by Ev_i^{C} in the context w' in a non-zero truth degree).

Remark 1

Informal explanation of r_{PbLD} is the following: The linguistic description *LD* characterizes linguistically (i.e., imprecisely) some relation between *y*'s and *x*'s. Moreover, we can apply it in all couples of contexts w and w'. If a specific x_0 in a context w is given then *LD* should contain rules which characterize all *y*'s that might depend on x_0. If Ev_i^A is a perception of x_0 (in the context w) then by r_{PbLD} we conclude that $\hat{y}_i$ is evaluated by the corresponding expression Ev_i^C. This means that the formula $Ev_i^A\, wx_0 \Rightarrow Ev_i^C\, w'y$ represents an evaluation fuzzy set of those *y*'s (in the context w') whose dependence on x_0 can be characterized by *i*-th rule from *LD*. The best evaluated *y*'s form its kernel and the description operator ι takes one of them. In a model, ι is interpreted by a special operation called *Defuzzification of Evaluating Expressions* (DEE) which selects the worst[3] of those best evaluated *y*'s (for more details see Novák, 2005b).

The r_{PbLD} has abundantly many applications in control, decision-making, classification and others. Wider application of FTT to modeling of complex human reasoning, however, requires methods of non-monotonic logic (see, e.g. Bochman, 2001). More details can be found in (Novák & Dvořák, to appear).

Acknowledgement

The support of the project 201/04/1033 of Grant Agency of Czech Republic and the project MSM 6198898701 of the MŠMT ČR is gratefully acknowledged.

Antonín Dvořák and Vilém Novák
University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
701 03 Ostrava 1, Czech Republic
antonin.dvorak@osu.cz, vilem.novak@osu.cz
phone: +420 597091404, fax: +420 596120478

[3] This means, that we take the greatest y, if Ev_i^C is S-formula and the smallest y, if it is B-formula.

References

Bochman, A. (2001). *A logical theory of nonmonotonic inference and belief change*. Berlin: Springer.

Běhounek, L., & Cintula, P. (2005). Fuzzy class theory. *Fuzzy Sets and Systems*, *154*, 34–55.

Church, A. (1940). A formulation of the simple theory of types. *Journal of Symbolic Logic*, *5*, 56–69.

Dvořák, A., & Novák, V. (2005). Fuzzy logic as a methodology for the treatment of vagueness. In L. Běhounek & M. Bílková (Eds.). *The Logica yearbook 2004* (pp. 141–151). Prague: Filosofia.

Esteva, F., & Godo, L. (2001). Monoidal t-norm based logic: towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, *124*, 271–288.

Fitting, M. (2006). Intensional logic. In E. N. Zalta (Ed.). *The Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu/archives/fall2006/entries/logic-intensional/.

Hájek, P. (1998). *Metamathematics of fuzzy logic*. Dordrecht: Kluwer.

Hajičová, E., & Partee, B., & Sgall, P. (1998). *Topic-focus articulation, tripartite structures, and semantic content.* Dordrecht: Kluwer.

Henkin, L. (1950). Completeness in the theory of types. *Journal of Symbolic Logic*, *15*, 81–91.

Novák, V. (2005a). On fuzzy type theory. *Fuzzy Sets and Systems*, *149*, 235–273.

Novák, V. (2005b). Perception-based logical deduction. In B. Reusch (Ed.). *Computational intelligence, theory and applications* (pp. 237–250). Berlin: Springer.

Novák, V. (2006). Fuzzy logic theory of evaluating expressions and comparative quantifiers. In *Proc. 11th Int. Conf. IPMU, Paris, July 2006* (Vol. 2, pp. 1572–1579). Les Cordeliers, Paris: Editions EDK.

Novák, V. (to appear). A comprehensive theory of evaluating linguistic expressions. In S. Gottwald, & P. Hájek, & U. Höhle, & E. P. Klement (Eds.). *Proc. of 26th Linz Seminar on Fuzzy Set Theory.* (Research Report No. 71, http://irafm.osu.cz/)

Novák, V., & Dvořák, A. (to appear). A model of complex human reasoning in fuzzy logic in broader sense. *Fuzzy Sets and Systems*.

Novák, V., & Perfilieva, I., & Močkoř, J. (1999). *Mathematical principles of fuzzy logic*. Boston: Kluwer.

What is a Logical Constant? The Inference-marker View

María J. Frápolli

1. The realm of logic

In this paper I aim to offer a characterization of logical constants taking what we, speakers, do with this kind of expressions as the point of departure. There are several definitions of logical constants, but none of them include a comprehensive account of their meaning in the broad sense of the word; none of them propose a picture capable of dealing with the syntactic features, semantic value and pragmatic role of logical terms. This criticism applies to Tarski's proposal (Tarski, 1966) and the long list of sequels that are now known as "invariantist" theories. "We call a notion 'logical'", Tarski says, "if it is invariant under all possible one-one transformations of the world onto itself" (1966, p. 149).

I will not discuss the different existing definitions of logical constants in any detail. They, and also the standard criticisms that can be made against them, are well known to specialists. For specialists and non-specialists alike, it is important to be aware that, as Warmbrod says in a recent paper, "there is as yet no settled consensus as to what makes a term a logical constant or even as to which terms should be recognized as having this status" (Warmbrod, 1999, p. 503).

Warmbrod describes the present situation; I, on the other hand, would like to analyze some of its sources and offer a proposal. The unsatisfactory situation concerning logical constants can be attributed to two main causes: (i) the common understanding of the relations between mathematics and logic, and (ii) the common understanding of the relations between language and logic. Contrary to the standard view during the past century, logic and mathematics are completely disparate enterprises. The most visible point of contact between the two disciplines is that they are both formal; it remains to be seen whether "formal" has the same meaning in both cases. It is a fact that modern logic has become more akin to mathematical theories than to the study of inferential patterns in natural languages. Nonetheless, the legitimate methodology of applying mathematical tools to the study of logic and languages does not support the illegitimate identification of the aim of logic with that of mathematics. Undoubtedly, modern logic developed during the second half of the XIX century thanks to the work of mathematicians as Jevons, Boole, Peano and Frege, among others. It evolved from the previous enterprise of applying algebra to the study of natural language, which was already a revolutionary theoretical enterprise (*see* Goldfarb, 1979). But modern logic was born as an independent science when logicians understood the previous algebraic relations not as relations on sets but as relations on concepts and conceptual contents. Logic deals with judgeable contents, with propositions; propositions are the basic elements of arguments, and even when we use artificial calculi in which propositional structures are represented at the syntactic level, logical relations are not held between syntactic items as grammar understands them, but between the contents of some of our speech acts. Uninterpreted sentences are purely syntactic entities, and purely syntactic entities are not truth-bearers. Thus, *a fortiori*, uninterpreted sentences cannot be what logic, the science of valid arguments, is about. Logical constants are propositional operators; propositions, statements, thoughts – all these expressions are equivalent – are the arguments of logical constants, and not the sentences themselves. Since propositions can be seen as 0-adic predicables, logical constants can be characterized as predicables on 0-adic predicables, i.e. as higher-order predicables.

The formality of logic has been often defined by means of adjectives such as "syncategorematic" or "topic-neutral", and correctly understood, both characterizations are appropriate. Syncategoremata, as medieval logicians characterized them, were expressions that could neither be in subject positions nor in predicate positions. Indeed, as propositional operators, logical constants cannot combine as subjects with first-order predicates to form a complete sentence, and for the same reason they cannot be combined with singular terms as if they were ordinary verbs. This way of understanding syncategoremata is the syntactic medieval characterization (see Klima, 2006). There also is a traditional semantic characterization that justifies the description of logic as topic-neutral. The topic neutrality of logic is not related to the alleged meaninglessness of logical constants, but to the universal application of the principles of valid reasoning. Logical constants do have meaning; the point stressed by their topic neutrality is rather that their arguments can be propositions of any kind, propositions that deal with any subject matter. I propose to substitute these traditional characterizations of logical constants – being syncategorematic and topic-neutral – and by the more precise qualification of being higher order predicables with 0-adic predicables as arguments, that pick up a shared feature of all expressions with any logical relevance (see Williams, 1992b).

Nevertheless, syntax does not provide the right demarcation, which is something that medieval logicians already knew. Interjections, exclamations, adverbs, and punctuation marks are syncategoremata without being logical constants, and many of them are also topic-neutral. Logical constants are logically relevant expressions not because of their syntactic features but because of the role they play in the general task of drawing inferences. Words such as "if", "not", "or" and the rest have attracted the interest of logicians because the speakers use them essentially in their explicit inferential acts. It is their function in ordinary inferential practices that makes them *logically* interesting terms. Due to all of this, the project of defining logical constants exclusively by attending to their syntactic properties is completely misconceived.

2. Inferential meaning and inference-markers

My proposal is to bring logic back to language, its natural home, and to place the philosophy of logic within the philosophy of language. What logic is cannot be determined by backing out of the inferential linguistic practices of human beings, and the same can be said of the task of identifying the features that make a term a logical constant. The semantics and pragmatics of logical words provide us with more promising insights than the misguiding clues offered by the alternative syntactic approach.

What is the semantic value of a logical constant? Generally speaking, the semantic value of an expression is the component it contributes to the proposition expressed by the sentences in which the term occurs. Wittgenstein gave the appropriate answer to the previous question: *none*. Logical constants are not names of anything and their semantic function cannot be to add a further element to the proposition. Naming nothing does not mean having no meaning: *semantic value* is a theoretical notion that covers only an aspect of the broader and more informal term "meaning". The Wittgensteinian claim is often known as "logical expressivism". There is no general agreement as to the credibility of logical expressivism but, in spite of the theoretical protests, everybody follows it in practice. Consider the customary way in which one interprets sentences and formulae in formal semantics. One does it by defining an interpretation that attaches objects in the Universe to the individual constants in the formulae, sets of objects in the Universe to the predicates in the formulae, sets of ordered sets of objects to the relational expressions. But logical constants are not interpreted this way. One might retort that logical constants do not need interpretation precisely because the constancy of their meaning. This is part of the reason, indeed, but neither the complete answer nor the most relevant part of it. Numerals, for instance, are also constants; one already knows their meaning and thus there is no need to decide, in each new model, which entities would correspond to them. The entities that can be their values are of the same type as the entities that are the values of the rest of expressions, i.e. either objects in the Universe or sets of these objects. The case of logical constants is different. They correspond neither to objects of any kind or to any kind of properties; given a set of formulae and an interpretation, they help to find the truth-value of the formulae according to the interpretation without adding new entities to the model. This is their specific function and, correctly understood, this is the core of logical expressivism.

The pragmatic role of logical constants can be easily understood if we consider the following illustration. The proposition expressed by a sentence like (α) in a standard context,

(α) My daughter is called "Victoria",

stands in varied inferential relations with other propositions, as for instance, those expressed by sentences (β), (γ), and (δ) in the same context,

(β) I have a daughter
(γ) Victoria is a girl
(δ) Victoria is a human being.

The propositions expressed by (γ) and (δ) together form the following material inference:

(I) Victoria is a girl; Victoria is a human being.

The truth of (δ) follows from the truth of (γ). By asserting (γ), one is committed to assent to (δ).

Now, if for some reason one were interested in stressing the commitment one undertakes to (δ) by asserting (γ), one would have to display the implicit, meaning-based, transition from (γ) to (δ) as a rule of inference, either singular, as (R1):

(R1) *If* Victoria is a girl, *then* Victoria is a human being,

or general, as (R2):

(R2) *If* somebody is a girl, *then* she is a human being.

When the rule of inference is added to the previous material inference, it becomes a formal inference, as in

(II) *If* Victoria is a girl, *then* Victoria is a human being; Victoria is a girl; then Victoria is a human being.

What is the difference between inferences (I) and (II)? They both have the same conclusion, that Victoria is a human being, and the same premise, that Victoria is a girl. The conditional in (II) is *not* a further premise, but a principle of reasoning. Bolzano, Frege and Peirce already paid attention to this distinction between premises and principles of reasoning. Premises are claims; they are asserted propositions, judgements. Principles of reasoning are rules. Overlooking the distinction would lead us to Carroll's paradox. If (I) and (II) share their premise and their conclusion, in which sense are they different? The answer is obvious: they are different because in (II) the principle of inference used in both is explicitly displayed. To display it, indicating at the same time that it is a principle and not a claim, one has to use the appropriate kind of words: logical constants. In (II) the words "if ..., then ..." serve to make explicit an inferential connection between the antecedent and the consequent. When they occur in a sentence, the sentence in question does not express a proposition but a rule. The same effect might have been achieved by inserting "therefore" between the premise and the conclusion. "Therefore" is another logical term. One might think that with this explanation we are committing the sin that Quine seemed to find at the origin of modern modal logic, i.e. the sin of confusing use with mention. We are not. The difference between object language and metalanguage is not as straightforward in the actual use of natural languages as it is in formal artificial languages. But in any case, we are talking about propositions, not about sentences, and thus the distinction does not apply. The arguments of "if ..., then ...", understood as means of stressing an inferential relation, are propositions and not sentences, and exactly the same happens in the case of "therefore". Now we are in the position of stating the pragmatic role of logical constants: speakers use these words to display the structure of an inference. Logical constants are added to material inferences to exhibit their status as inferences; they are not essential to carry out inferential movements but their involvement becomes indispensable in order to present inferential connections between propositions as explicit inferences. It is only when we want to make the presence of an inference patent that they become useful.

In order to make inferential connections explicit an expression does not only need to have inferential meaning. All concepts have inferential meaning to some extent; the inferential connections between concepts justify the material inferences in which they are involved. In this sense the proposal I am putting forward here, which I will call "the inference-marker view", goes further than Gentzen's and Prawitz's views. The core of the inference-marker view is not that the meanings of logical constants can be given as sets of rules, introduction and elimination rules, but rather that the pragmatic significance of logical constants is to bring an implicit inference into the open.

It is important to realize that this pragmatic role does not imply that logical words always indicate *valid* inferences. The speaker uses logical words to indicate that, from her point of view, the relevant propositions she is expressing are somehow inferentially connected. But she might present as an inference one that is *invalid*, just because she might be wrong about some aspects of the case in hand. This doesn't undermine my general claim about the pragmatic role of logical constants. "If", and the other constants, have exactly the same meaning when they appear in deductive, inductive or simply invalid inferences. Inductive and deductive inferences have distinct properties, but they do not affect the meaning of constants. And the same applies to invalid inferences. The pragmatic role of "if" is constant across its various uses. "If" means the same when it appears in an instance of the fallacy of affirming the consequent as when it appears in an instance of our reliable Modus Ponens. It is because its meaning doesn't change that we classify the former, unlike the latter, as fallacious.

All logical words share the same general pragmatic role, but different logical words have different specific inferential meanings. In each case the rules that govern the relevant inferential movements depend on the particular meaning of the logical constants actually used. Thus, "if", "not" and "or" codify different inferential entitlements, that in formal calculi are represented by different sets of rules of inference that disclose the circumstances and consequences of their use.

I will propose a general definition of logical constants following the lines already mentioned, a definition that include their syntactic status, their semantic characterization and their pragmatic role. However, unlike many proposals that take syntax as the point of departure, my discussion of the subject will start from pragmatics. I take what we do with words, with *logical* words in this case, as the foundation level.

3. Syntax, semantic, and pragmatics of logical words

The definition:

[DEF] Logical constants are higher-order predicables that have 0-adic predicables as arguments. They don't name any kind of entity but rather are natural language devices for making inferential relations among concepts and propositional contents explicit.

DEF involves a syntactic claim, that logical constants are higher-order; a semantic claim, that they do not name; and a pragmatic claim, that by using them a speaker shows the presence of an inference. The semantic claim – logical expressivism – has been defended by John Buridan and Albert of Saxony in the Middle Ages (*see* Klima, 2006), and by Wittgenstein (1922), Austin (1962) and Brandom (1994) in the XXth century. Ramsey (1928), Ryle (1956) and Brandom (1994) supported the pragmatic view, that I have called "the inference-marker view".

Being higher-order – the syntactic claim – is only one of the necessary conditions for being a logical constant. And the same happens with the semantic aspect. In language, there are many different expressions that, strictly speaking, don't *name* anything, and shouldn't be catalogued as logical constants for this reason alone. But as logic is the science of inferences, logical constants are essentially inference-markers. The pragmatic role explains the semantic and syntactic features: logical constants are not components of the contents of inferences but have these propositions as arguments.

A 0-adic predicable is a predicable with 0 argument places, i.e. a proposition. That logical words are higher order predicables that have propositions as arguments should be obvious if one recalls that the basic notion of logic is validity, that validity is a property of inferences, and that inferences, considered in an objective sense as the result of acts of inferring, are sets of propositions. There is another way in which inferences can be understood, i. e. as movements, as transitions from sets of propositions to a proposition. The notion of inference has a dynamic sense, a sense that supports the static view of inferences as sets of propositions. This dynamic sense has been recently stressed by Dubucs and Marion (2003), by Martin-Löf (1996), and by Sundholm, among others. It is because an inference is a movement that genuine logical constants have an aspect of their significance that is dynamic. They show inferential bridges between concepts and propositions.

An immediate objection to my definition is that, although it fits sentential connectives well, it ignores identity and first order quantifiers. That first order identity is not a logical constant is nowadays an accepted point. First order quantifiers, on the other way, are generally considered as the logical constants that characterize first order calculi. The challenge that first order quantifiers pose to my view is not that they are first order (or that we call them "first order") for quantifiers are higher order functions. The difficulty here is that these quantifiers have n-adic predicables ($n > 0$), and not propositions (0-adic predicables), as arguments. A possible way out is provided by the fact that DEF can have two readings, one weaker than the other. They are the following,

$[\text{DEF}]_{\text{weak}}$ Logical constants are higher-order predicables that may admit 0-adic predicables among their arguments. They don't name any kind of entity but rather are natural language devices for making explicit inferential relations among concepts and propositional contents.

$[\text{DEF}]_{\text{strong}}$ Logical constants are higher-order predicables whose arguments are 0-adic predicables. They don't name any kind of entity but rather are natural language devices for making explicit inferential relations among concepts and propositional contents.

$[\text{DEF}]_{\text{weak}}$ predicates logical constanthood of types, while $[\text{DEF}]_{\text{strong}}$ predicates it of tokens. Under the former, a type, say a quantifier, is a logical constant if, among other characteristics, it has tokens that are functions of 0-adic predicables. Propositional quantification would be an obvious case that would provide quantifiers with the required feature. In any case, quantifiers are not an homogeneous kind, and it is reasonable to assume that different types with different functions may be distinguished. Under the stronger definition, what is classified as a logical constant is a token, i. e., a particular instance of a type together with its particular aspects. If one selects exclusively $[\text{DEF}]_{\text{strong}}$, it makes no sense asking whether quantifiers or any other kind of expression are or are not logical constants or not.

Fortunately, it is not necessary to choose one of the two options and reject the other. We can assume the charitable position of classifying types as logical constants in a weak sense if, and only if, they have tokens that are so in a strong sense.

DEF assembles three aspects that are individually necessary and jointly sufficient for being a logical constant.
My definition rules out:

(i) First-order predicables, and hence it discards first-order identity and membership as logical constants
(ii) Predicate-formers such as some uses of negation, conjunction and disjunction, higher-order identity and the reflexivity operator
(iii) Monadic sentential operators that act as circumstance-shifting operators, such as modal, epistemic and temporal operators.
(iv) Monadic sentence-formers, such as monadic quantifiers

Nevertheless,

(v) DEF doesn't imply that first-order identity, conjunction and disjunction should be removed from standard calculi. They shouldn't. They all have jobs to perform there.
(vi) DEF doesn't imply that modal, epistemic and tensed logics shouldn't be considered as logics. They should, although they all include at least two sets of constants, genuine logical constants, which earn for them the title "logic", and specific constants that make them *modal*, *epistemic* or *tensed* logics, in each case.
(vii) DEF doesn't imply that quantifiers are not logical constants. Rather, it distinguishes different kinds of quantifiers. Monadic quantifiers don't act as inference-markers, but binary quantifiers usually do. This does not mean any rejection of Frege's (1884) account. Frege rightly understood the nature of numerical expressions as higher-order concepts, and correctly defined existence as an expression of quantity. Numeric expressions and existence are monadic higher-order functions whose

arguments are concepts. They indicate sizes of concept's extensions. Nevertheless, they don't act as markers of inferences and so they are not logical constants.

(viii) My view does not suggest any criticism of Mostowski's insights that monadic quantifiers help us to construe propositions out of propositional functions, or that logical quantifiers cannot be used to single out individuals. Both theses are correct, but none of them define logical constanthood.

Points (i)–(viii) require explanation, although this is a task that I will not attempt here. A single paper of the length of the present one would not be enough for such a task that would eventually involve a revision of all terms that have ever been proposed as logical. Nevertheless, some comments would help. Point (i) is hardly controversial: there are many authors who do not count first order identity or membership among logical constants (see, for instance, Peacocke, 1976 and Warmbrod, 1999). Point (ii) refers to a relevant issue, that some uses of the words that are commonly accepted as logical terms actually have a combinatorial function. Some uses of negation, conjunction and disjunction have predicative expressions as arguments. When this happens, their function is helping to build complex concepts out of simple ones. Complex concepts such as "unhappy", "honest politician", "married woman", "homeless" are composed of more basic concepts by means of negation and conjunction. Although nowadays this function of concept construction seems to have been forgotten, medieval logicians were perfectly aware of it (*see* Klima, 2006). Point (iii) says that some monadic sentential functions are circumstance-shifting operators, i.e. operators that, although don't contribute a component to the proposition expressed by the sentence in which they occur, are relevant to the task of evaluating the propositions that act as their arguments. Point (iv) points to a significant feature of some quantifiers. All of us consider the Fregean treatment of existence in (Frege, 1884) as the first step towards the correct understanding of quantifiers. In § 53, Frege says: "In this respect existence is analogous to number. Affirmation of existence is in fact nothing but denial of the number nought". Although this is accurate in relation to existence, it is not in relation to generality. It is relevant here to acknowledge that quantifiers may be monadic higher-order operators or binary higher order operators. The monadic existential quantifier indicates, as Frege saw, that the extension of the concept that is its argument is not empty. But the monadic universal quantifier has a slightly different meaning: it indicates the scope of the concept that is its argument. In natural languages, both quantifiers standardly are binary operators. Frege also saw this: "It must be remarked that the words 'all', 'any', 'no', 'some' are prefixed to concept-words. In universal and particular affirmative and negative sentences, we are expressing relations between concepts, we use these words to indicate the special kind of relation. They are, thus, logically speaking, not to be more closely associated with the concept-words that follow them, but are to be related to the sentence as a whole." (Frege, 1892, p. 48) When they are binary operators, their meaning is dynamic and they indicate an inferential connection between the two concepts that are their arguments. From the quoted text it follows that Frege thought that both universal and existential quantifiers have uses in which they are binary. I totally agree. In artificial languages we are free to define the status of the operators that we introduce, but in natural languages we are not. These operators, as the rest of our expressions, are supported by the tasks the speakers use them for, and standardly the existential quantifier is used as a monadic operator to express the non-emptiness of an extension, and the universal quantifier is used as a binary operator to express a principle of reasoning.

All this is still very vague, but the main idea under the inference-marker view, that I am proposing in this paper should be clear by now. To sum up, logical constanthood is a functional concept. It applies to tokens of expressions depending on the role they perform. The central notion of logic is not truth but truth-preservation, and truth-preservation, i.e., validity, is a property of arguments, i.e. of sets of propositions. This is the standard explanation, and the correct one. My proposal takes it seriously as a guide into the inquiry about logical constants. And the result is that, in a strong sense, only some uses of negation, disjunction, conditional and quantifiers are genuine logical constants, their types being logical constants in a weak sense. In addition there are several kinds of expressions that play a role in the practice of drawing inferences, although their function is not presenting inferences as such. These kinds are typically (i) operators on predicative expressions that help forming complex predicables out of simpler ones (some uses of negation, conjunction, and disjunction), (ii) propositional operators, both monadic and binary, and among them, (ii. a) circumstance-shifting operators (modal and tense operators, for instance), and (iii) binary first order identity and membership. All these kinds deserve the logician's attention, although for different reasons.

If this conclusion sounds too unpalatable there is still the possibility of relaxing the requirements and considering any operator with a relevant role in the general task of drawing inferences as a logical constants. This would allow adding the expressions described in (i), (ii) and (iii). Still, the perspective should be pragmatic and the characterization should attend the task performed rather than the syntactic category. This broader characterization would be highly imprecise, and probably too liberal, but it would permit welcome back on board the familiar set of words. I don't object as far as we remain aware that, among the hospitable set, several well-defined types of operators can be distinguished.

I prefer the stronger characterization. In my view, negation and conditionality, both singular and general, constitute the core of our logical apparatus. And correctly understood, this view deeply respects tradition. It is faithful to Frege, Ramsey, and Peirce, to Wittgenstein, Prior, and Williams, to Sellars and Brandom, and in general to all those who consider what we do with words as the basic level of analysis.

Department of Philosophy
University of Granada (Spain)
Campus de Cartuja
Edificio Facultad de Psicología
18071 Granada, España
frapolli@ugr.es

References

Austin, J. (1962). *How to do things with words.* Oxford: Oxford University Press.

Brandom, R. (1994). *Making it Explicit. Reasoning, Representing, and Discursive Commitment*. Cambridge, Mass.: Harvard University Press.

Carroll, L. (1895). What Achilles said to the Tortoise. *Mind, 15*, 279–280.

Dosen, K. (1989). Logical Constants as Punctuation Marks. *Notre Dame Journal of Formal Logic, 30*(3), 362–381.

Dubucs, J. & Marion, M. (2003). Anti-realism and substructural logics. In A. Rojszczak & J. Cacho & G. Kurczewski (Eds.). *Philosophical Dimensions of Logic and Science, Selected Contributed Papers from the 11th International Congress of Logic, Methodology and the Philosophy of Science, Krakow, 1999* (pp. 235–249). Dordrecht: Kluwer.

Dummett, M. (1973). *Frege: Philosophy of Language*. Cambridge, Mass.: Harvard University Press.

Dummett, M. (1981). *The Interpretation of Frege's Philosophy*. Cambridge, Mass.: Harvard University Press.

Etchemendy, J. (1983). The doctrine of logic as form. *Linguistic and Philosophy,* 6, 319–334.

Feferman, S. (1999). Logic, logics, and logicism. *Notre Dame Journal of Formal Logic*, *40*, 31–54.

Frege, G. (1879). Conceptual Notation. A formula language of pure thought modelled upon the formula language of arithmetic. In G. Frege (1972). *Conceptual Notation and Related Articles*, translated and edited with a Biography and Introduction by Terrell Ward Bynum (pp. 101–203). Oxford: Clarendon Press.

Frege, G. (1884). *The Foundations of Arithmetic* (reprinted 1980). Oxford: Basil Blackwell.

Frege, G. (1892). On Concept and Object. In P. Geach, & M. Black (Eds.) (1952). *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell.

Geach, P. (1962). *Reference and Generality. An Examination of Some Medieval and Modern Theories.* Ithaca and London: Cornell University Press.

Gentzen, G. (1969). *Collected Papers*. Amsterdam: North-Holland.

Goble, L. (Ed.) (2001). *The Blackwell Guide to Philosophical Logic*. Oxford: Blackwell Publishers.

Goldfarb, W. (1979). Logic in the twenties. The nature of quantifiers. *Journal of Symbolic Logic, 44*, 351–368.

Gómez-Torrente, M. (2002). The problem of logical constants. *Bulletin of Symbolic Logic, 8*, 1–37.

Haack, S. (1978). *Philosophy of Logics.* Cambridge/N.Y.: Cambridge University Press.

Higginbotham, J. (2000). On second-order logic and natural language. In G. Sher, & R. Tieszen (Eds.). *Between Logic and Intuitionism. Essays in Honor of Charles Parsons*, (pp. 79–99). Cambridge/N.Y.: Cambridge University Press.

Klima, G. (2006). Syncategoremata. In *Elsevier's Encyclopedia of Language and Linguistic, 2nd, vol. 12* (pp. 353–356). Edited by Keith Brown. Elsevier: Oxford.

Koslow, A. (1992). *A Structuralist Theory of Logic*. Cambridge: Cambridge University Press.

Martin-Löf, P. (1996). On the meaning of the logical constants and the justifications of the logical laws. *Nordic Journal of Philosophical Logic, 1*(1), 11–60.

Peirce, C. S. (1932). *Collected Papers of Charles Sanders Peirce*. Edited by Ch. Hartshorne, & P. Weiss. Volume II. Elements of Logic. Cambridge: Harvard University Press.

Prior, A. (1976). What is logic? *Papers in Logic and Ethics*. Amherst: University of Massachusetts Press.

Ramsey, F. P. (1928). Universals of law and of fact. In Philosophical Papers. Edited by D. H . Mellor (pp. 135–51). Cambridge: CUP.

Read, S. (2003). Logical consequences as truth-preservation. *Logique et Analyse, 183-184*, 479–493.

Sher, G. (1991). *The Bounds of Logic. A Generalized Viewpoint*. Cambridge, Mass: The MIT Press.

Sher, G., & Tieszen, R. (Eds.) (2000). *Between Logic and Intuition. Essays in Honor of Charles Parsons*. Cambridge: Cambridge University Press.

Sundholm, G. (unpublished manuscript). *A century of judgement and inference 1837-1936*.

Tarski, A. (1966). What Are Logical Notions? In J. Corcoran (Ed.) (1986). *History and Philosophy of Logic* (Vol 7, pp. 143–154).

Warmbrod, K. (1999). Logical Constants. *Mind*, 8, July, 503–538.

Williams, C. J. F. (1989). *What is Identity?* Oxford: Oxford University Press.

Williams, C. J. F. (1992a). *Being, Identity and Truth*. Oxford: Oxford University Press.

Williams, C. J. F. (1992b). Towards a Unified Theory of Higher-Level Predication. *Philosophical Quarterly, 42(*169), 449–464.

Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus*.

Wittgenstein, L. (1978). *Remarks on the Foundations of Mathematics*. Edited by G. H. von Wright, & R. Rhees, & G. E. M. Anscombe. Revised edition. Cambridge, Mass.: The MIT Press.

“Realistic” Belief Dynamics*

Brian Hill

For several years now, the “realism” of the classical representations of beliefs proposed by logicians, philosophers, and economists has been the source of anxiety and debate. The realism of the models of doxastic actions which rely on such representations, such as those models proposed by decision theory, choice theory, and, more recently, belief revision, has given rise to similar worries. The purpose of this paper is to propose and motivate a framework which supports a more realistic model of doxastic states, of the changes they undergo, and of the role they play in action and decision. This framework shall be developed and applied to the case of belief revision, a paradigm example of an operation involving beliefs, and a field which has recently seen some concern about the realism of traditional approaches.

In the first part of the paper, the general framework shall be developed in two stages – firstly a representation of the instantaneous state shall be proposed and motivated, then an operation capturing the dynamics of this state shall be defined. In the second part of the paper, it shall be shown that the framework yields a model of belief revision which, firstly, recovers the Gärdenfors postulates as applying in particular circumstances; and secondly, can accommodate *iterated* revisions, recovering several proposed revision operators for iterated revisions as special cases.

1. General framework

1.1 Interpreted Algebras

All systems purporting to represent beliefs or operations involving them assume an underlying *language*, with its own *logic* (for the most part, the classical consequence relation). The fundamental observation motivating the proposed model is that, between any two moments, the languages which are *effective* or “*in play*” at these moments – the languages in which the beliefs active at these mo-

* This paper summarises the content of a presentation given at Logica in June 2006. Most of the content had been presented in April 2006 at the Philform seminar at the IHPST in Paris. The author would like to thank both audiences for their comments.

ments are couched – may differ. A similar point seems to hold for the logics of these languages, in so far as they are comparable. Let us call the combination of language and logic effective at a particular moment, the *local logical structure* at that moment. The model developed shall be more "realistic" or "sophisticated" in that it pays explicit attention to and indeed represents formally the local logical structures effective at particular moments, as well as the changes in these structures as new information comes into the fray.

Indeed, just the fact of explicitly representing the language and logical structure which are effective at a given moment allows one to deal with several of the most important weaknesses of traditional models of belief. On the one hand, traditional notions of belief generally imply that, if an agent (actively or explicitly) believes that he has a meeting at 10.00, then he also believes that he has a meeting at 10.00 and there are infinitely many primes. This unintuitive consequence of their models is avoided once one introduces the notion of a sentence or an issue being *in play*: the reason that the agent does not appear to have the latter belief is that, the *whole question of the number of primes* – the sentence "there are infinitely many primes", if you prefer – is *out of play* for him at that moment. This notion of 'in play', close to Fagin and Halpern's 'awareness' (Fagin and Halpern, 1988), cannot be captured by traditional models and needs some sort of *syntactic* apparatus distinguishing those sentences which are in play from those which are not. By modelling explicitly the set of sentences in play at a given moment – the *local language* (at that moment) – one avoids these troublesome cases of logical omniscience.

On the other hand, traditional notions of belief, which generally take a fixed logical structure, and thus a fixed consequence relation, generally have problems with subjects who apparently fail to recognise logical or intensional equivalence. They cannot represent an agent who accepts that he needs to go to the eyedoctor without accepting that he need go to the ophthalmologist, since the two sentences are (intensionally) equivalent. However, as soon as one considers not only the language, *but also* the logical structure on it as local, so that the logical relationships between sentences hold only *in so far as they figure in that local logical structure at that moment*, one can account for examples of this sort: the *local* logical structure relevant at such moments does not necessarily respect the *global* logical structure pertaining to some *global* language.

Thus, by taking the language and its associated logical structure as local, one avoids in one step a range of 'logical omniscience' problems which have previously needed different strategies, as in Fagin and Halpern (1988). It goes without saying that the idea that an agent can only operate in a fragment of his total linguistic range at any given moment of time is a simple, important, but particularly intuitive aspect of the finiteness of human thought, and is certainly tame compared to bolder models of human limitations which postulate a multiplicity of 'minds' or a particular type of mental architecture.

Interpreted algebras will be used to model formally the local logical structure effective at a given moment (for the classical propositional case considered here); they are defined as follows.

Definition 1 (Interpreted Algebra). An *interpreted algebra* $\mathbf{B}$ is a triple (B_I, B, q), where B_I is the free Boolean algebra generated by a set I (the *interpreting algebra*),[1] B is a Boolean algebra (the *base algebra*), and $q: B_I \to B$ is a surjective Boolean homomorphism.

An *element* of $\mathbf{B}$ is a pair $(\phi, q(\phi))$, $\phi \in B_I$; they shall be referred to by the appropriate elements of the interpreting algebra, and often be called "sentences". The *consequence relation* $\Rightarrow$ is defined on elements in the natural way: $\phi \Rightarrow \psi$ iff $q(\phi) \leqslant q(\psi)$.

The interpreting algebra models the local language effective at the moment in question, with I being the set of *locally* atomic or primitive sentences in play at that moment. It is, so to speak, the "syntax" of the local logical structure.

The base algebra is the *local logic* on this language. It is, so to speak, the "semantics" of the local logical structure. Just as the elements of the interpreting algebra may be thought of as the *sentences* of the local logical structure, the elements of the base algebra may be thought of as the (local) *propositions*. Accordingly, q is the map taking sentences to propositions, and may be thought of as the *valuation* of the sentences of the language.[2] Elements of the interpreted algebra consist of a sentence and the proposition which it expresses; the consequence relation on elements arises from relations between the propositions they express.

Intuitively, the local logical structure effective at a given moment is finite. The use of a Boolean algebra to model the local language allows one to circumvent the apparent contradiction between the finiteness of the language and the fact that recursion with Boolean operators yields an infinite set of sentences. The purported finiteness of the local language has two aspects: firstly, there a finite number of (locally) primitive sentences in play, but furthermore there are effectively only a finite number of linguistic entities which can be formed from them, since one naturally discounts such differences as those between '*A and A and A and A*' and '*A*'. The former aspect is captured by using interpreting algebras with finite I, the latter by the use of Boolean algebras, which automatically

[1] A Boolean algebra is a distributed complemented lattice; the order will be written as $\leqslant$, meet, join, complementation and residuation as $\wedge$, $\vee$, $\neg$, $\rightarrow$, the top and bottom elements as $\top$ and $\bot$. The free Boolean algebra generated by a set X shall be noted as B_X for the rest of the paper; details on this and the other notions used in this paper may be found in Koppelberg (1989).

[2] The fact that it is a Boolean homomorphism guarantees that the ordinary conditions on valuations are satisfied.

disregard the sort of differences mentioned.[3] The interpreting algebra will thus generally be assumed to be finite.

It follows that the base algebra will be finite, and thus atomic.[4] The atoms of the base algebra can be thought of as "states" or "*small* worlds" – worlds in the sense that every sentence of the local language receives a valuation in each world (thanks to q); *small* in the sense that *only* the sentences of the local language receive valuations in these worlds. It is of crucial importance that the expression in terms of Boolean algebras favoured here does not imply any rejection of the dominant "possible worlds", or as it might be called, *extensional* view: indeed, assuming the algebras are atomic, the two are technically equivalent. If the *algebraic* perspective is favoured, it is only because it proves more fruitful for considering the relationship with the local language, and for modelling the dynamics of local logical structures in general. Recourse shall be made, at times, to the extensional view, since it is, for many, simpler and more intuitive.

Finally, the assumption that the homomorphism q is surjective implies that there are no two "small worlds" which cannot be distinguished by sentences of the local language. This assumption follows from the idea that the local logical structure incorporates all and only the sentences in play at a given moment with all and only the logical structure on them at that moment. If one employed interpreted algebra with a non surjective homomorphism, then the logical structure would contain distinctions between elements of the base algebra (local propositions) which are beyond the linguistic resources in play, thus contravening the intuition behind local logical structures.

Finally, here are two examples of basic, but important, sorts of interpreted algebra.

Example 1. The *point interpreted* algebra for the sentence ϕ, $\mathbf{B}_{\phi_{\mathrm{p}}} = (B_{\{\phi\}}, 1, q)$, where 1 is the two element Boolean algebra $(\{\top,\bot\})$, and $q : \phi \mapsto \top$.

The simple interpreted algebra for the sentence ϕ, $\mathbf{B}_{\phi} = (B_{\{\phi\}}, \mathbf{2}, q)$, where **2** is the four element Boolean algebra $(\{\top,\bot,x,x'\})$, and $q : \phi \mapsto x$.[5]

Point algebras and simple algebras are the two basic possibilities for representing a (consistent) local logical structure which has essentially one sentence (ϕ) in play (that is, there is the one sentence and those which can be formed

[3] Apparent objections to this choice of model will generally be defused once one realises that no restrictions are put on the set I of primitive sentences. See Hill (2006b) for further discussion.

[4] Standard terminology is employed here: an atom of a Boolean algebra is an element $a \leqslant B$, such that, for all $x \in B$, if $\bot \leqslant x \leqslant a$, then either $x = \bot$ or $x = a$. Note furthermore that the assumption of finiteness is not required for any of the definitions or results in this paper; the weaker assumption that B is atomic is sufficient.

[5] Recall (footnote 2) that $B_{\{\phi\}}$ is the free Boolean algebra generated by $\{\phi\}$.

from it with Boolean connectives). In the point algebra, this sentence is accepted as a (local) logical truth in the language (in terms of small worlds, there is one world, where ϕ holds, $\neg\phi$ holding at no world in this interpreted algebra). The simple algebra admits the "possibility" that the sentence may be true as well as false (there are two worlds, one where ϕ holds, the other where $\neg\phi$ holds).

1.2 Fusion

Investment in a model which captures the logical imperfectness of an agent's instantaneous belief state seems worthless if it is not accompanied by an account of how this state can change. In terms of the framework proposed here, a proposed model of the local logical structure at particular moments is of little use unless it can also model the *changes* in the local logical structure which occur from one moment to the next. In this section, a *fusion* operation shall be defined which will model the change in the local logical structure as new information comes into play.

The changes to local logical structures which shall be dealt with here are those brought about by the incoming information. Typically, in models of belief (or knowledge) and their changes, new information comes in the form of a sentence (or set of sentences) of the language.[6] However, no global or overarching language is assumed in the current framework; indeed, given that the only language present is the local language of the current local logical structure, the whole problem is how to deal with sentences which do not necessarily belong to this language. It is therefore necessary to endow the incoming information with its own fragment of language, with the sort of basic logical structure which always accompanies such fragments of language. To put it another way, the new information comes in the form of (at least) a local language with a local logic. It shall thus be modelled using interpreted algebras.[7]

The flexibility of the notion of interpreted algebra permits it to capture the diverse, more or less complicated, forms which incoming information might take. At one end of the spectrum, rich local languages (large B_I) with interesting logical structures (B and q) can accurately model an input which does not consist of a simple sentence, but comprises a complex of diverse information, about how such a sentence comes into play, how it was learnt, what justifies it, and so on. At the other end of the spectrum, the simple traditional cases of a single sentence entering into play can be captured using simple or point interpreted algebras (Example 1).

[6] This is the case not only in belief revision, but equally in epistemic dynamic logic, or in typical Bayesian update theory.

[7] The current discussion concerns only the general framework; in Section 2.1, a richer model of new information, obtained by adding extra structure to the interpreted algebra, shall be proposed.

However, given that no overarching language is assumed, but only the local languages contained in the individual interpreted algebras, there is *a priori* no way of *identifying* sentences belonging to *different interpreted algebra*, and in particular sentences belonging to the algebra representing the current local logical structure and the one representing the new information. To represent the fact that the new information may involve sentences which already belong to the current local logical structure, supplementary technical apparatus is thus required. The identification of sentences between different interpreted algebras shall be represented using an appropriate relation, $\simeq$, *called identification*. For the purposes of this paper, identification relations can be considered to be Boolean congruence relations on the elements of the interpreted algebra, that is, equivalence relations which conserve Boolean structure.[8] In subsequent discussion, an identification relation $\simeq$ shall be assumed.

The question of changes in the local logical structure in the face of new information now becomes that of proposing an operation taking two interpreted algebras, with an identification of sentences between them, and yielding an interpreted algebra which *respects the identification of the sentences*. The operation of *fusion* of interpreted algebras does just this. It can be defined from two simple operations on interpreted algebras.

The first is the operation of free product on interpreted algebras, $\otimes$, which is obtained by taking the free product of the interpreting algebras, the free product of the base algebras, and the canonical homomorphism between them.[9] At the level of languages, the new local language obtained is the closure under Boolean operations of the disjoint union of the two initial local languages. On the semantic side, the set of small worlds or states in the resulting interpreted algebra is the cartesian product of the sets of small worlds or states of the initial algebras, and the valuation on these worlds (the homomorphism q) is the naturally derived valuation; one might think of the free product as "combining" small worlds, to give "enriched" small worlds. The second operation is the operation of quotient by the identification relation, obtained by taking the quotient of the interpreting algebra, and the induced quotient of the base algebra, with the canonical homomorphism between them. In terms of local languages, the quotient operation *identifies* or *renders identical* the sentences which were $\simeq$-equivalent in the initial local language. In terms of the local logic, the propositions corresponding to sentences which are $\simeq$-equivalent in the initial local logical structure are *identified* in the resultant structure. Equivalently, quotienting on the semantic level *removes* the small worlds which are witness to differences between any pair of $\simeq$-equivalent sentences ϕ and ψ; that is, worlds where the valuations of ϕ and ψ differ. The operation of *fusing* two interpreted algebra is defined as follows.

---

[8] For full technical details on this and other aspects, see Hill (2006a, 2006b).

[9] For more details on the product of Boolean algebras, see Koppelberg (1989).

Definition 2 (**Fusion** $*$). Given two interpreted algebras $\mathbf{B_1}$ and $\mathbf{B_2}$, with an identification relation $\approx$ between them, the *fusion* of the two algebras, $\mathbf{B_1} * \mathbf{B_2}$, is defined as the quotient of the free product of the two algebras by the relation $\approx$.

This operation models the change in the local logical structure under new incoming information: both the original local logical structure and the new information are modelled by interpreted algebra; the resulting local logical structure is the resulting interpreted algebra. This model is intuitive: in fusing the new information (with its fragment of language) with the existing logical structure, the "sum" of the two languages is taken (free product), and then appropriate sentences figuring in the different languages are identified (the quotient). Given that the operation to be modelled is that of "merging" or "combining" two fragments of language, one would expect it to be commutative: no priority should be given to one over the other.[10] The operation $*$ has this property.

Two examples shall serve to illustrate this sort of operation.

Example 2. For ϕ in $\mathbf{B}$, the fusions with the relevant simple and point algebras (Example 1) are as follows:

Simple algebra $\mathbf{B} * \mathbf{B}_\phi$ is isomorphic to $\mathbf{B}$;

Point algebra $\mathbf{B} * \mathbf{B}_{\phi_\mathrm{p}}$ is isomorphic to $(B_I, B/(\phi), q')$, where $B/(\phi)$ is the quotient of B by the smallest congruence relation such that $\phi \approx \top$, and q' the composition of q with the quotient homomorphism.

The first example illustrates that the fact of bringing into play a sentence which is already in play, in such a way that no extra logical structure is allocated to it, does not alter the algebra. The second example concerns fusion with a sentence already in play, but such that the sentence, in so far as it figures as new information, is endowed with extra logical structure: namely, it is taken to be equivalent to the true (of the local language). This leads to a change in the local logical structure to accommodate this information: the fusion results in a logical structure with the same local language, but such that the sentence is now equivalent to the true (or alternatively true in all small worlds).

This second example is interesting because, put in terms of small worlds, it essentially says that fusion with $\mathbf{B}_{\phi_\mathrm{p}}$ does not change the language but *removes* all the small possible worlds, or states, where ϕ is false. This sort of operation, which plays an important role in the literature on public announcement and dynamic epistemic logic (Gerbrandy and Groeneveld, 1997), is thus reproduced as a *special case* in the framework proposed here. More generally still, it is not

[10] Commutativity is expected only for the logicolinguistic structures in which beliefs and new information are couched. It will not be desired for full models of beliefs and new information, and shall not be present in the model proposed in Section 2.1.

difficult to see that the model of epistemic programs proposed by Baltag and Moss (2004) is based on the sort of fusion operation proposed here.[11]

A model has been proposed *both* of the local logical structure effective at a particular moment – interpreted algebras – *and* of the dynamics of this structure – the fusion operation. This technical apparatus is abstract, and thus can be applied to several different questions in several different fields; in each field, the basic notions assume different philosophical interpretations. In the next section, the power of the framework will be illustrated by using it to develop a realistic model of belief revision.

2. Belief revision

Models of belief revision typically consist of a model of the belief state, a representation of new information with which the state is to be revised, and a revision operation representing the revision of the former by the latter, which satisfies a certain number of belief revision postulates, such as the so-called Gärdenfors postulates (Gärdenfors, 1988). In the original AGM paradigm, the state of belief is taken to be a set of sentences (of a given language L) closed under a (given) logical consequence relation, and the new information consists of a sentence of this language.[12] A typical model of belief revision, proposed by Grove (1988), uses a *Grove order* $\preceq$ on the set S of maximal consistent sets ("possible worlds") of a language L, that is, a reflexive order which is connected, transitive and finitarily stoppered.[13] In this model, the set of beliefs is the set of sentences true in the $\preceq$-minimal worlds, and a sentence ψ is believed after revision by ϕ if it is true in all the $\preceq$-minimal worlds satisfying ϕ.[14]

Two further questions which have entered into the fray since the original AGM work are the question of *iterated belief* and that of *realism*. On the one hand, it is desirable to have a model such that, whatever results from the revi-

[11] Leaving aside the locality of languages, which is not present in Baltag's paper, and modalities, which are not (yet) present in the basic framework proposed here, Baltag's update product and the operation of fusion defined above turn out to be *technically* similar. Indeed, since this paper was presented, Baltag has applied his system to belief revision, in a way close to that presented in Section 2 (Baltag and Smets, 2006). Although there is little space to comment here, comparison with his paper brings out more clearly the difference in *philosophical* viewpoint: for example, Baltag et al show no interest in the question of realism, in belief revision postulates, and do not have equivalents of Theorem 1 or 2 below.

[12] In AGM theory, the operation of contraction – removal of a belief – is taken as primitive. Here only the operation of revision is considered. For the relationship between them, see Gärdenfors (1988).

[13] $\preceq$ is finitarily stoppered if and only if, for all $\phi \in L$, $|\phi| \neq \emptyset$ implies that $\{x \in |\phi| \mid x \preceq y, \forall y \in |\phi|\} \neq \emptyset$, where $|\phi| = \{x \in S | x \vDash \phi\}$. See Grove (1988) for details.

[14] For the uninitiated, it may be useful to compare this model with Lewis' semantics for counterfactuals; for a detailed comparison, see Grove (1988).

sion of belief, it can itself be revised in the face of subsequent information; famously, the traditional AGM models, and indeed the Grove model described above, do not satisfy this condition. There have however been proposals for modelling iterated revision, and indeed a variety of postulates which one might want iterated revisions to satisfy; see Rott (2003) for some examples. On the other hand, concern has surfaced about the realism of the proposed theories of belief revision, Hansson (2003) and Rott (2004) being just two examples where such worries have been expressed.

In the following section, a model of belief revision shall be proposed. This model is realistic in its conception, to the extent that it is based on the local logical structures introduced above, and inherits from these structures the ability to avoid some of the problems described above. Furthermore, the model satisfies the Gärdenfors postulates in appropriate cases, and models iterated revision in such a way as to recover several iterated revision operations proposed in the literature as special cases. This will be interpreted as a sign that the model does not make the sort of *idealistic assumptions* that force other models to deal only with such special cases.[15]

2.1 A model of belief revision

In Section 1.1, a model of the local logical structure effective at a particular moment was proposed, in the form of what was called interpreted algebra. The locality of this language and of its logic respond well to certain limits in real agents' belief states. The basic proposal for modelling the belief state of an individual is to employ traditional models of beliefs, but, instead of using some fixed language and logical structure, considering the beliefs of the agent at a particular moment as couched in a local logical structure which is effective at that moment. This is, so to speak, a model of the beliefs of which the agent is "aware", in the agent's own language, at a particular moment.

The simplest model of beliefs would be as a set of sentences closed under logical consequence – that is, the logical consequence of the local logical structure which is operative at the appropriate moment.[16] However, it has been suggested that correct representations of the belief *states* of an individual should include information not only about his current beliefs, but also about how he would revise them, or, alternatively, about how "entrenched" they are

[15] The model is also realistic in the stronger sense that it permits an analysis of counterexamples to Gärdenfors postulates, such as that proposed by Rott (2004), which makes clear in exactly what sense the examples do not deal with the special cases to which the postulates are to apply. See Hill (2006b) for more details.

[16] This model implies that the agent is locally logically omniscient. However, this seems a generally correct assumption: if an agent believes '*A*', 'if *A*, then *B*', recognises these beliefs as such, and '*B*' is in play, then it would seem that he believes '*B*'.

(Darwiche and Pearl, 1997). Such a model of belief states shall be employed. It consists in adding a Grove order – representing not only the agent's beliefs but potential revisions of these beliefs – to an interpreted algebra – representing the local logical structure in play at the moment in question. The resulting structure is called an ordered algebra.

Definition 3 (Ordered algebra). An ordered algebra is a pair $(\mathbf{B}, \preceq)$ where $\mathbf{B} = (B_I, B, q)$ is an interpreted algebra and $\preceq$ is a reflexive order on the atoms of B which is connected, transitive and finitarily stoppered.

Example 3. The *point ordered algebra* for sentence ϕ, $(\mathbf{B}_{\phi_p}, \preceq_{\phi_p})$ has, for interpreted algebra, the point algebra for ϕ, and, for order, the only possible one.

The *simple ordered algebra* for sentence ϕ, $(\mathbf{B}_\phi, \preceq_\phi)$, has, for interpreted algebra, the simple algebra for ϕ, and, for order, the order favouring ϕ: $q(\phi) \preceq q(\neg\phi)$.

As a point of terminology, we shall say that the *centre* of an ordered algebra $(\mathbf{B}, \preceq)$ is the set of elements of $\mathbf{B}$ true in all the small worlds minimal with respect to $\preceq$. An element of the centre is a generator if it is true only in the $\preceq$-minimal small worlds.[17] Finally, an element is a local tautology if it is true in all the small worlds.

Ordered algebras provide a particularly rich representation of the agent's doxastic state at a given moment. For a sentence ϕ, ordered algebras can capture two senses in which it may be "believed". It might in the centre of the algebra; furthermore, it may be a local tautology. The first case is what are called "beliefs" in Grove's model (Gärdenfors, 1988); this set of "beliefs" may be revised if new information forces one to move to worlds where not all of them hold. The second case corresponds what have been called "doxastic commitments" or "irrevocable beliefs" (Segerberg, 1998); no revision of such beliefs is admissible, since there is no world (of the ordered algebra) where they do not hold. However, whereas in the literature, where a fixed language and notion of logical consequence are presupposed, "irrevocable" beliefs end up being just the tautologies of this language, in the framework proposed here, where the language and the logic are local, it is *not* necessary that the *local* tautologies are tautologies of some fixed language (see Section 1.1); in this sense, the believer is not modelled as omniscient. Moreover, as opposed to most traditional models, not only the centre of the ordered algebra modelling the agent's belief state, but also the local tautologies, may change in time. They are thus to be understood as those opinions which the agent cannot envisage giving up *at that particular moment* – his *local* commitments, if you like. The sentences in the centre are the

[17] Since q is surjective, there is always a generator. This is natural in the finite case appropriate here.

most preferred sentences *amongst those which are in play* – his (explicit, instantaneous) beliefs. For an ordered algebra representing the agent's belief state, the centre is thus the set of beliefs.

What allows this departure from the tradition is the fact that ϕ may *not* be believed in *two* general senses. Firstly, it can not figure in the local algebra at all: it can be out of play. It is this possibility that is not permitted by previous theories, and it is by changing the sentences which are in and out of play that the beliefs or commitments of the agent can change in ways in which previous models cannot capture. Secondly, ϕ may be in play for the agent, true in some small worlds, but not all of the $\preceq$-minimal ones. The ordinary method for revising beliefs in Grove models applies to such sentences: the set of beliefs after revision by ϕ are those sentences true in all $\preceq$-minimal small worlds where ϕ. The ordered algebra thus represents the agent's opinion on how he *would* revise his beliefs by sentences which are in play for him at that moment; that is, it provides *envisaged revision*. However, this is not a full measure of *actual revisions*, because the Grove order in the ordered algebra cannot take account of revision by sentences which do not belong to the local language of this algebra. In order to propose a general operator for revision, which applies to such cases, it is necessary to have a representation of such new information; this is the task to which we now turn.

Often new information with respect to which beliefs are to be revised is treated as a simple sentence of some fixed language. However, in the framework proposed here, where no use is made of such a fixed language, incoming information will generally require a local logical structure of its own, which can be modelled as an interpreted algebra (Section 1.2). However, the interpreted algebra only models the logical structure in which the incoming information is couched; it does not specify which sentences in this structure are learnt, or the extent to which the sentences of the local language are to be accepted. For example, if the local logical structure pertinent for a case where ϕ is learnt is modelled as a simple algebra (Example 1), some supplementary structure on this algebra would be needed to represent the fact that it is ϕ and not $\neg\phi$ which is to be accepted. It would seem natural to represent this fact with an order on the states or small worlds of this algebra which favours (the small world where) ϕ to (that where) $\neg\phi$. So doing, one obtains an ordered algebra – in fact, one obtains a *simple ordered algebra* (Example 3). In this basic case, new information can be represented as ordered algebra; the suggestion is that this sort of representation is appropriate in general.

Indeed, representing new information with ordered algebras inherits the advantages of ordered algebras which have been emphasised above. As with the case of belief states, different statuses of the different elements of incoming information may be captured by ordered algebras. A sentence learnt *irrevocably* – accepted without any envisaged possibility of challenging the new information – can be represented as a local tautology of the ordered algebra representing the incoming information. On the other hand, information learnt in a context such that it is reliable only under certain conditions – say, the result of a scientific experiment, which is valid only under certain assumptions relating to the details of the experiment – would belong to the centre of an ordered algebra which also contains sentences expressing the appropriate conditions. To be more pedantic, what is learnt is characterised *precisely* by any sentence which is true only in the minimal (or most preferred) worlds of the ordered algebra – that is, by any *generator* of the ordered algebra (see above). Incoming information shall be modelled by an ordered algebra, where the sentence learnt can be thought of as a generator of the algebra.

Under the current proposal, both the belief state and the new information are represented by interpreted algebras with appropriate orders on them (ordered algebras); the revision operation will somehow combine these algebras. The operation which combines the interpreted algebras has already been defined and motivated: it is the fusion operation $*$ of Section 1.2. It remains to specify how to combine the orders on the algebras. There is a selection of operations which may be employed here, several of which have been discussed in some form or another in the literature. For the purposes of this paper, where the general framework is at issue, it would not be appropriate to enter into detailed considerations and debates; it will suffice to pick a natural candidate and develop a revision operation built on this operation on orders. Although this candidate, and the revision operation constructed from it, has several interesting, attractive and useful properties, let it be emphasised that other operations on orders may prove equally useful, and may result, using a similar procedure to that carried out below, in equally interesting revision operations. The operation on orders used here is the lexicographic product, $\times_L$, which, loosely speaking, follows the *latter* order, unless the two elements are equivalent under this order, in which case it follows the former order.[18] It has the advantage of being non commutative, which fits well with the idea that new information should have priority over previous beliefs.

Definition 4 (Fusion $*$ of ordered algebras). Let $(\mathbf{B_1},\preceq_1)$ and $(\mathbf{B_2},\preceq_2)$ be ordered algebras.[19] The fusion $(\mathbf{B_1},\preceq_1) * (\mathbf{B_2},\preceq_2) = (\mathbf{B_1} * \mathbf{B_2},\preceq_1 \times_L \preceq_2)$.

$(\mathbf{B_1},\preceq_1)$ represents the initial belief state: its centre is the set of beliefs. $(\mathbf{B_2},\preceq_2)$ represents the new information: the sentence learnt is a generator. $(\mathbf{B_1},\preceq_\mathbf{1}) * (\mathbf{B}_2,\preceq_2)$ represents the resulting belief state: its centre is the new set of beliefs.

[18] Formally, $(a, b) \preceq_1 \times_L \preceq_2 (c, d)$ iff either $b \prec_2 d$ or $b \approx_2 d$ and $a \preceq_1 c$. The order which has priority is a question of convention. The latter order is chosen here to simplify the discussion below.

[19] Recall that the appropriate identification relation is presupposed.

2.2 Properties of the model

The operator $*$, with the interpretation of ordered algebras as representations of belief states and incoming information, provides a model of belief revision in so far as it satisfies an appropriate translation of the well-known Gärdenfors postulates for belief revision into the proposed framework. Since the representation of the belief state after revision (ordered algebra) is of the same format as the representation before revision (and thus appropriate for further revision), it is automatically an *iterated* revision operator; furthermore, two important iterated revision operators proposed in the literature (Segerberg, 1998; Nayak, 1994) can be recovered in the proposed framework as *special cases* corresponding to particular constraints placed on the incoming information. These properties are expressed by the following two theorems.[20]

Familiarity with the Gärdenfors postulates for belief revision is assumed (for the canonical presentation, see Gärdenfors (1988)). Rather than reproducing them in all their glory, and to avoid getting bogged down in technical details, the following informal version of the theorem is stated.

Theorem 1. *Let* $(\mathbf{B_1},\preceq_1)$ *be an non trivial ordered algebra with centre K, let* $(\mathbf{B_2},\preceq_2)$ *contain sentences* ϕ *and* ψ *and have generator* ϕ*, and let* $(\mathbf{B_2},\preceq_3)$ *have generator* $\phi \wedge \psi$*. Let* $K * \phi$ (resp. $K * (\phi \wedge \psi)$) be the centre of $(\mathbf{B_1},\preceq_1) * (\mathbf{B_2},\preceq_2)$*, (resp.* $(\mathbf{B_1},\preceq_1) * (\mathbf{B_2},\preceq_3)$)*. Then the Gärdenfors postulates, applied to K,* $K * \phi$ *and* $K * (\phi\wedge \psi)$*, and using the notions of consequence in the appropriate interpreted algebras* $(\mathbf{B_1}, \mathbf{B_2}, \mathbf{B_1} * \mathbf{B_2})$*, are satisfied.*

There are two subtleties in this theorem, with respect to the simple Gärdenfors formulation of the postulates. On the one hand, where one normally assumes a fixed language and logical consequence relation, there are several in play here, so it is necessary to specify which one is relevant for each postulate; in all cases, the theorem holds for the most natural candidate. On the other hand, whereas the Gärdenfors postulates are normally expressed in terms of sentences and sets of sentences, the basic notion here is that of ordered algebra. This is a more flexible and general representation of beliefs and new information, which offers several notions of "belief" or "sentence learnt", of which only one is pertinent to the theorem: namely, the interpretation of beliefs and sentences learnt as the "most preferred sentence" (centres and generators) of the respective algebras. The postulates *do not* necessarily apply to the *local tautologies* of the respective algebra, that is to the "commitments" or "irrevocable sentences". This is one concrete sense in which the model of belief revision proposed here recov-

[20] Proofs, and indeed rigorous formulations, of these theorems shall not be presented here; for details, see Hill (2006b).

ers the traditional theory as an *idealisation*: the Gärdenfors postulates hold, but only in the *special cases* where centres and generators (and the corresponding notions of "belief" or "information") are being used. One of the desiderata of a more realistic model of belief dynamics – namely, to exhibit in which sense previous theories are idealisations – is thus fulfilled.

As noted above, $*$ is an iterated revision operator, in that it yields a structure (ordered algebra) fit for subsequent revision (using $*$). Furthermore, two iterated revision operators, called "radical" and "moderate" revision by Rott (2003), which have been suggested and defended by Segerberg (1998) and Nayak (1994) respectively, can be recovered by placing conditions on the ordered algebra representing the incoming information. This is the sense of the following theorem.[21]

Theorem 2. *Let K be the centre of the ordered algebra* $(\mathbf{B}, \preceq)$. *Then*

(Rad) *If* ϕ *and* ψ *are modelled by* $\mathbf{B}_{\phi_{\mathrm{p}}}$ *et* $\mathbf{B}_{\psi_{\mathrm{p}}}$ *respectively, the postulate for radical revision is satisfied.*

(Mod) *If* ϕ *and* ψ *are modelled by* $\mathbf{B}_{\phi}$ *et* $\mathbf{B}_{\psi}$ *respectively, the postulate for moderate revision is satisfied.*

This theorem counts as a further illustration of the fruitfulness of this model of belief revision: iterated revision operations proposed in the literature are apparently recovered as special cases of the form of the input information. In the sense in which they suppose that the input information takes a particular form, they are *idealisations*; in the sense in which the model proposed here does not make this supposition, and indeed can accommodate a multiplicity of possible formats for the incoming information, it is more *realistic*.

One can conclude that the model of belief revision proposed in the second part of this paper is more realistic, and indeed seems to open up fruitful possibilities of development into a full theory of belief revision. The general framework on which this model rests, and which was presented in the first part of this paper, has proved promising in the case of belief revision; it would not be exaggerated to expect similar success when applied to other questions where belief is involved and realism is an issue.

[21] See any of the cited papers for a formulation of the iterated revision postulates.

Brian Hill
GREGHEC / IHPST
1, rue de la Libération 7
78350 Jouy-en-Josas, France
hill.brian@wanadoo.fr
www.brian-hill.org

References

Baltag, A., & Moss, L. S. (2004). Logic for epistemic programs. *Synthèse*, *60*, 1–59.

Baltag, A., & Smets, S. (2006). Dynamic belief revision over multiagent plausibility models. In G. Bonanno, W. van der Hoek, & M. Wooldridge, M. (Eds.). *Proceedings of the 7th Conference on Logic and the Foundations of Game and Decision Theory* (LOFT06), (pp. 11–24).

Darwiche, A., & Pearl, J. (1997). On the logic of iterated belief revision. *Artificial Intelligence*, *89*, 1–29.

Fagin, R., & Halpern, J. Y. (1988). Belief, awareness, and limited reasoning. *Artificial Intelligence*, *34*, 39–76.

Gerbrandy, J., & Groeneveld, W. (1997). Reasoning about information change. *Journal of Logic, Language, and Information*, *6*, 147–169.

Grove, A. (1988). Two modelings for theory change. *Journal of Philosophical Logic, 17*, 157–170.

Gärdenfors, P. (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* Cambridge: MIT Press.

Hansson, S. O. (2003). Ten philosophical problems in belief revision. *Journal of Logic and Compututㅤion, 13*, 37–49.

Hill, B. (2006a). *Jouer avec le Faux. Recherches sur les processus mentaux à l'œuvre dans la lecture des textes de fiction.* Doctorate Thesis. Université Paris 1 Panthéon-Sorbonne (retrievable on HAL database – Hyper Articles en Ligne, France).

Hill, B. (2006b). Towards a "sophisticated" model of belief dynamics. *Cahiers de Recherche GREGHEC*, 868/2007.

Koppelberg, S. (1989). General theory of boolean algebras. In J. D. Monk, & R. Bonnet (Eds.). *Handbook of Boolean Algebras*, volume 1. North Holland.

Nayak, A. C. (1994). Iterated belief change based on epistemic entrenchment. *Erkenntnis, 41*, 353–390.

Rott, H. (2003). Coherence and conservatism in the dynamics of belief. Part II: Iterated belief change without dispositional coherence. *Journal of Logic and Computation, 13*, 111–145.

Rott, H. (2004). A counterexample to six fundamental principles of belief formation. *Synthèse, 139*, 225–240.

Segerberg, K. (1998). Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic, 39*, 287–306.

Six Ways of Knowing Whether[1]

Bjørn Jespersen

Introduction

It is shown how to analyse and formalise (possible-world) propositional and hyperpropositional empirical attitudes of the form, "*a* knows whether *A*" in their two *de dicto* and at least two *de re* variants. The logic of *knowing whether* is developed within Transparent Intensional Logic, whose notion of *construction* will explicate the notion of hyperintensionality. (For background and further details, see Tichý, 1988, 2004.)

Let *A* be an arbitrary object of knowledge. Then *knowing whether A* is construed as a special case of a general case. The general case is

knowing which disjunct (if any) of $A \vee B$ *is true.*

The general case should not be confused with *knowing whether A or B (if any)*. The difference is the difference between knowing which disjunct is true and knowing whether their disjunction is true. (Syntactically, the difference is predicated on whether $\vee$ includes K_a in its scope.) The disjunction $A \vee B$ in the general case may well be inclusive, for all that is required to know which disjunct (if any) of $A \vee B$ is true is knowing of at least one of *A*, *B* that it is true. The only exception is when $B = \neg A$, in which case $\vee$ needs to be exclusive.

The most important difference between *knowing that A* and *knowing whether A* is that the latter is *not factive*; *knowing whether A* is logically compatible with $\neg A$.[2]

---

[1] A version of this paper (entitled "Russell's first puzzle") was read at the conference *100 Years of 'On Denoting'*, Department of Philosophy, University of Genova, 18 December 2005. It coincides in part with material also appearing in Duží *et al.* (Ms.). I am indebted to Marie Duží and Gert-Jan Lokhorst for comments on an earlier draft.

[2] Rescher calls it an "epistemic resolution regarding a proposition [*A*] when the knower [*a*] knows whether *A* is true or not: $K_aA \vee K_a\neg A$" (2005, p. 24). See also Hintikka (1975) and Lewis (1998). $K_aA \vee K_a\neg A$ is not a tautology, for *a* may know neither *A* nor $\neg A$, and should not be confused with the classical tautology $K_aA \vee \neg K_aA$. (See Genesereth & Nilsson, 1987, p. 227.) Hart *et al.* argue that *knowing whether* and *knowing that* are interdefinable, such that *a* knows that *A* iff *A* and *a* knows whether *A* (1996, p. 254). They also point out that *knowing whether* is 'invariant under complementation'; *a* knows whether *A* iff *a* knows whether $\neg A$. This is due to the non-factivity of *knowing whether*, and is symptomatic of its poverty of information.

Therefore, for instance, the following standard principle of transmission of knowledge does not hold for *knowing whether*:[3]

> *a* knows that *b* knows that *A* is true;
> therefore, *a* knows that *A* is true.

The reason is because the principle would translate into

> *a* knows whether *b* knows whether *A* is true;
> therefore, *a* knows whether *A* is true.

If *a* knows which disjunct of $A \vee B$ is true, it is because any one of the following four options obtains:

- *a* knows that *A*
- *a* knows that *B*
- *a* knows that *A and B*
- *a* knows that *neither A nor B*.

The third option presupposes that $B \neq \neg A$ on pain of rendering knowledge inconsistent.[4] The fourth option presupposes that if $B = \neg A$ then if *A* is a proposition then it must be a properly partial function; and if *A* is a hyperproposition then it must yield a properly partial function.

An ascription of *knowledge whether* does not reveal which of the four options obtains. Nor need the ascriber know which obtains in order to make a true ascription. But the ascriber must know that the ascribee knows which it is. For illustration, imagine that you know that Fermat had a proof of whether his Last Theorem is indeed a theorem, but do not know which way the proof went. Then you know that Fermat knew whether the Theorem is a theorem, while you may not know what Fermat knew. What you do know is that Fermat would have been the one to turn to for a conclusive answer.

Let George IV know whether Scott is the author of *Waverley*. Understood *de dicto*, George IV knows whether the proposition that Scott is the author is true, or George IV knows whether the hyperproposition that Scott is the author yields a true proposition. Understood *de re*, either George IV knows of the particular individual who is singled out as the author whether he or she is

---

[3] Consider these two mixed cases: *a* knows whether *b* knows that *A* is true; therefore, *a* knows that *A*. And: *a* knows that *b* knows whether *A* is true; therefore, *a* knows that *A*. The former is valid; the latter, invalid.

[4] A paraconsistent epistemic logic holds that some instances of $A \wedge \neg A$ may figure as pieces of knowledge; namely, when it is known that a self-contradiction is true. This is a non-vacuous claim, since paraconsistent logicians argue that some (though not all) self-contradictions are indeed true.

Scott, or George IV knows of the particular individual who is singled out as the author whether he or she is the author. Both variants have their passive forms as well: The individual who is singled out as the author is such as to be known by George IV whether to be Scott or whether to be the author. The active variant is the variant with an anaphoric reference; the passive variant ascribes to the particular individual the property of being such as to be known by George IV whether to be Scott or the author.[5]

The *intension/hyperintension* (*proposition/hyperproposition*) distinction concerns, within epistemic logic, whether the piece of knowledge is intensionally or hyperintensionally individuated. Therefore, *A* in *knowing whether A* is ambiguous between intensions and hyperintensions, and rigorous disambiguation is called for. By 'intensional entity' I mean intension as defined by possible-world semantics, which defines intensions as functions from possible worlds and identifies any two such logically equivalent functions. 'Proposition' denotes only possible-world propositions in this paper. By 'hyperintensional entity' I mean entities whose principle of individuation is finer than logical equivalence (see Cresswell, 1975.) In popular terms, hyperintensional logic is able to distinguish between a half-full glass and a half-empty glass. This distinction presupposes the possibility of operating with two or more different (yet equivalent) modes of presentation or conceptualisations of the same inverse relation. In the case of knowledge, we need to be able to operate with two or more different (yet equivalent) hyperintensional 'modes of presentation' of the same proposition. The relevance to epistemic logic is that even though *a* knows, hyperintensionally, that the glass before him is half-empty, it does not follow that *a* knows that the glass is half-full (or *vice versa*). The same proposition is conceptualised in two different manners; first, in terms of the glass being half-*empty*, then in terms of the glass being half-*full*. Likewise, even though *b* knows, hyperintensionally, that the figure before her is tri*angular*, it does not follow that *b* knows that the figure is tri*lateral* (or *vice versa*).

Logical foundations

In order to define *knowing whether* in its intensional and hyperintensional (constructional) variants within TIL, we define this theory's concepts of *construction* and *simple and ramified types*. The former types are also known as *types of order 1 over an ontological base*.

DEFINITION 1 (*type of order 1 over ontological base B*)
Let *B* be a collection of pairwise disjoint, non-empty sets. Then

- Each member of *B* is a *type of order 1 over B*.

[5] This active/passive distinction is explained in detail in Duží *et al*. (Ms.)

- If α, $\beta_1,...,\beta_n$ are arbitrary *types of order 1 over B*, then the set $(\alpha\beta_1...\beta_n)$ of all partial functions whose arguments are tuples with elements of the *types* $\beta_1,...,\beta_n$, respectively, and whose values are elements of *type* α is also a *type of order 1 over B*.
- Nothing else is a *type of order 1 over B*. □

Remark. An ontological base of ground types must be decided upon before launching a type-theoretic analysis. For the purposes of the analysis of attitudes, not only individuals and truth-values but also times and possible worlds are needed: ο (truth-values), ι (individuals), τ (times), ω (possible worlds). τ is also the type of real numbers; hence time is modelled as a continuum.

Remark. The type of an intensional entity is polymorphous, namely $(\alpha\omega)$, α an arbitrary type. An intensional entity is a (perhaps properly) partial function from possible worlds to α-objects. The intensional entities occurring in this paper are all of the type $((\alpha\tau)\omega)$, abbreviated $\alpha_{\tau\omega}$: Functions from possible worlds to functions from times to α-objects. For instance, a *proposition* is of type $o_{\tau\omega}$: A function from possible worlds to a function from times to truth-values. This is so in order to model both modal and temporal variability.

DEFINITION 2 (*construction*)

- (*Variable*) Let a total valuation function *v* be given that associates variables $x_0^\alpha, x_1^\alpha,..., x_n^\alpha...$ with a sequence *Seq* of objects $a_0, a_1, ..., a_n, ...$ of type α. Then the *variable* x_n^α *v*(aluation)-*constructs* the n^{th} object *a* of *Seq* relative to *v*.
- (*Trivialization*) The *construction* ^0X consists in *constructing X* without the mediation of other *constructions* and leaves *X* unchanged.
- (*Double Execution*) The *construction* ^2X *v-constructs* what is *v-constructed* by what is *v-constructed* by *X iff X* is a *construction* that *v-constructs* a *v-proper construction*, a *construction* being *v-proper* if it *v-constructs* an entity. Otherwise ^2X is *v-improper* in the sense of failing to *v-construct* anything.
- (*Composition*) Let *X* be a *construction* that *v-constructs* a function *f*, of type $(\alpha\beta_0...\beta_n)$, and let $X_0,..., X_n$ be *constructions* that *v-construct* the entities $b_0,..., b_n$, respectively, of types $\beta_0,... \beta_n$, respectively. Then $[XX_0...X_n]$ is a *construction* called *Composition*. If *f* is undefined at $<b_0,..., b_n>$ or if any of $b_0,..., b_n$ is not *v-constructed*, then $[XX_0...X_n]$ is *v-improper* by failing to *construct* anything. Otherwise $[XX_0...X_n]$ *v-constructs* the value of *f* at the arguments $b_0,..., b_n$.
- (*Closure*) Let $x_0^\alpha,..., x_n^\alpha$ be pairwise distinct variables and *Y* a construction. Then $[\lambda\, x_0^\alpha,..., x_n^\alpha\, Y]$ is a *construction* called λ-*Closure* (or simply

Closure). It *v-constructs* the following function g. Let v' be a valuation identical with v at least up to assigning objects b_i to variables x_i, $1 \leq i \leq n$. If Y is v'-improper, g is undefined on $<b_0,..., b_n>$. Otherwise the value of g on $<b_0..., b_n>$ is the object v'*-constructed* by Y.

- Nothing else is a *construction*. □

Remark. The following examples may help keep Variable, Trivialisation, and Double Execution apart. If the 5th slot in *Seq* is 5 then x_5 *v*-constructs 5. $^0 5$ constructs 5. $^0 x_5$ *v*-constructs, for all valuations v, the variable x_5. If X *v*-constructs x_5 then $^2 X$ *v*-constructs 5. The Trivialization of an entity X is a primitive, non-perspectival, 'direct' construction of X and makes use of no other constructions in constructing X. A rough linguistic counterpart would be the device of quotation. Just as "quotation" quotes, or mentions, the word 'quotation', so $^0 X$ quotes, or mentions by constructing, the entity X. If c is a variable ranging over propositional constructions, then the Double Execution $^2 c$ consists in, first, descending from variable to propositional construction and, second, descending from propositional construction to proposition. That is, $^2 c$ is a construction *v*-constructing a proposition.

DEFINITION 3 (*ramified type hierarchy*)

- T_1 (*simple types*) *Simple types are of order 1.*
- C_n (*construction of order n*):
    - o If x is a variable ranging over a type of order n then x is a *construction of order n*.
    - o If X belongs to a type of order n then $^0 X$ is a *construction of order n*.
    - o If $X, X_1,..., X_m$ are *constructions of order n* then $[XX_1...X_m]$ is a *construction of order n*.
    - o If $x_1,...,x_m$, Y are *constructions of order n* then $[\lambda x_1...x_m Y]$ is a *construction of order n*.
- T_{n+1} (*type of order n+1*) Let $*_n$ be the set of all *constructions* of order n. Then:
    - o $*_n$ and every type of order n are *types of order n+1* ('type raising').
    - o If $\alpha,\beta_1,...,\beta_m$ are *types of order n+1* then $(\alpha\beta_1...\beta_m)$ is also a *type of order n+1*.
    - o Nothing else is a *type of order n+1*. □

Remark. The functions Sub^n (for 'substitution') and Tr^α (for 'Trivialization function') are indispensable for the logical analysis of attitudes *de re*. (See Tichý 1988, p. 68 for Tr^α, p. 75 for Sub^n, and Materna 1997, p. 337 for both.) Sub^n and Tr^α make bound variables amenable to manipulation by, first, 'untying' them from the context they are bound in and, second, substituting Trivializations for them. (Intuitively speaking, a constructional (i.e., hyperintensional) context is one 'mentioning' a construction rather than 'using' the construction to obtain the entity it constructs. Thus there is a methodological, though not substantial, parallel between the hyperintensional epistemic logics of TIL and sententialism.) In the cases below the relevant bound variables are bound by Trivialization, like ^0x or ^0X, where x occurs at least once in X. Let X, Y, Z be constructions of order n, at least Y a variable. Then the function Sub^n, of type $(*_n *_n *_n *_n)$, is a mapping which, when applied to $\langle X, Y, Z\rangle$, returns the construction that is the result of correctly substituting X for Y in Z. Next, let α be a type of order n, o an object of type α. Then Tr^{α}, of type $(*_n\ \alpha)$, is a function which, when applied to o, returns the Trivialization of o. For instance, if a/ι, $A/\iota_{\tau\omega}$, then $[^0Tr^{\iota}\ {}^0a]$ constructs ^0a (i.e., the Trivialization of the individual a). The Composition $[^0Tr^{\iota}\ {}^0A_{wt}]$ v-constructs the Trivialisation of the individual (if any) v-constructed by $^0A_{wt}$. The Composition $[^0Sub\ [^0Tr\ {}^0A_{wt}]\ {}^0y\ {}^0[...y...]]$ is v-improper if $^0A_{wt}$ is v-improper. Otherwise, if a is the individual v-constructed by $^0A_{wt}$ then it v-constructs the construction v-equivalent with $^0[...{}^0a...]$.

To express *knowing that neither A nor* $\neg A$ in logical notation, we need to introduce the propositional property of being undefined (*Und*).

DEFINITION 4 (*Undefined*)

Let $\forall/(o(o\omega))$; $\forall'/(o(o\tau))$; $\neg/(oo)$; $=/(ooo)$; $P/o_{\tau\omega}$; $True, False, Und/(o\ o_{\tau\omega})_{\tau\omega}$. Then

$$^0\forall\lambda w\ [^0\forall'\lambda t\ [^0Und_{wt}\ {}^0P] = [\neg[^0True_{wt}\ {}^0P] \wedge \neg[^0False_{wt}\ {}^0P]]].\ \square$$

Knowing whether requires two definitions. In the case of empirical attitudes, *knowing* is a relation (-in-intension) either to a proposition or a propositional construction, while mathematical attitudes are invariably relations to constructions of the truth-value **T**. Thus,

$K/(o\ \iota\ o_{\tau\omega})_{\tau\omega}$ ('to know a proposition')
$K^*/(o\ \iota\ *_1)_{\tau\omega}$ ('to know a (first-order) construction').

Let $Q/o_{\tau\omega}$; $C, D/*_1$; $p/*_1 \rightarrow o_{\tau\omega}$; $c, d/*_2 \rightarrow *_1$, $^2c \rightarrow o_{\tau\omega}$, $^2d \rightarrow o_{\tau\omega}$; $=_1/(o\ o_{\tau\omega}\ o_{\tau\omega})$; $=_2/(o*_1\ *_1)$; $\iota/(o_{\tau\omega}\ (oo_{\tau\omega}))$; $\iota^*/(*_1(o*_1))$. '$x/*_n \rightarrow \alpha$' means that x is a construction of type $*_n$ ranging over the type α. Terms for truth-functions occur in infix notation without Trivialization for better readability. Here C, D are propositional constructions, and c, d variables ranging over propositional constructions. We only define the cases in which $^0Q =_1 \lambda w\lambda t\ [\neg{}^0P_{wt}]$ and $^0D =_2 {}^0[\lambda w\lambda t\ [\neg C_{wt}]]$ to keep the definitions as economic as possible. The respective general cases may be readily reconstructed from the definitions.

DEFINITION 5 (*knowing whether P*)

a knows whether P iff
$\lambda w\lambda t\ [^0K_{wt}\ ^0a\ [^0\iota\lambda p\ [p_{wt} \wedge [[p =_1\ ^0P] \vee [p =_1 \lambda w\lambda t\ [\neg{}^0P_{wt}]]]]]]$. □

Remark. Provided C constructs P, the *definiens* of DEF. 5 may be equivalently constructed by

$$\lambda w\lambda t\ [^0K_{wt}\ ^0a\ [^0\iota\lambda p\ [p_{wt} \wedge [[p =_1 C] \vee [p =_1 \lambda w\lambda t\ [\neg C_{wt}]]]]]].$$

When it is known whether P is true or false or neither, what is known is this:

DEFINITION 6 (*P being true or false or neither*)

$[^0\iota\lambda p\ [p_{wt} \wedge [[p =_1\ ^0P] \vee [p =_1 \lambda w\lambda t\ \neg[^0P_{wt}]] \vee$
$[p =_1 \lambda w\lambda t\ [^0Und_{wt}\ ^0P]]]]]$. □

Remark. The third disjunct, $[p =_1 \lambda w\lambda t\ [^0Und_{wt}\ ^0P]]$, can be dispensed with when P is a total function.

DEFINITION 7 (*knowing* whether C*)

a knows whether C iff*
$\lambda w\lambda t\ [^0K_{wt}\ ^0a\ [^0\iota{}^*\lambda c\ [[^2c]_{wt} \wedge [[c =_2\ ^0C] \vee [c =_2\ ^0[\lambda w\lambda t\ [\neg C_{wt}]]] \vee$
$[c =_2\ ^0[\lambda w\lambda t\ [^0Und_{wt}\ C]]]]]]]$. □

Philosophical application (I): Mathematical attitude

Mathematical attitudes must be relations to constructions of truth-*values* and not also of truth-*conditions* (propositions), since mathematical truths and false-hoods are not sensitive to worlds and times. Therefore, sentences ascribing mathematical attitudes are susceptible to only one reading *de dicto* and at most two readings *de re*, depending on the particular example. For instance, one thing is to know* of 2 that it is the only even prime; another thing is to know* of the only even prime that it is 2 (see the following section).

Knowing* whether Fermat's Last Theorem is true is to know* which of two constructions constructs **T**. The *analysandum* is the sentence (disregarding tense)

> "Fermat knows whether there are positive integers a, b, c, n (n>2) such that $a^n + b^n = c^n$."

Let ν be the type of *Pos* (positive integers) such that $a, b, c, n, x/*_1 \rightarrow \nu$; $Pos/(o\tau)$; $2/\nu$; $\forall, \exists/(o(o\nu))$, $c/*_2 \rightarrow *_1$, ${}^2c \rightarrow o$; $Fermat/\iota$. We write 'x^n' for '$[{}^0Exp\ n\ x]$', $Exp/(\nu\nu\nu)$ the power function taking x to its n^{th} power. Then the analysis *de dicto* is the Closure

$$\lambda w \lambda t\ [{}^0K^*_{wt}\ {}^0Fermat\ [{}^0\iota{*}\lambda c\ [[{}^2c] \wedge [c =_2 {}^0[{}^0\exists\lambda abcn\ [[{}^0Pos\ a] \wedge [{}^0Pos\ b] \wedge [{}^0Pos\ c] \wedge [{}^0{>}\ n\ {}^02] \wedge [{}^0{=}\ [{}^0{+}\ a^n\ b^n]\ c^n]]] \vee c =_2 {}^0[{}^0\forall\lambda abcn\ [[{}^0Pos\ a] \wedge [{}^0Pos\ b] \wedge [{}^0Pos\ c] \wedge [{}^0{>}\ n\ {}^02] \supset \neg[{}^0{=}\ [{}^0{+}\ a^n\ b^n]\ c^n]]]]]]].$$

The analyses *de re* are reconstructible from the analyses below.

Philosophical application (II): Empirical attitude

Here follows a six-way disambiguation of

> "*a* knows whether Scott is the author of *Waverley*."

Taken together, the disambiguations express *knowing whether* and *knowing* whether* in their two *de dicto* and all their four *de re* variants. We need to employ the fourth option mentioned in the Introduction, since the individual concept *the author of Waverley* is a properly partial function.

The propositional and constructional attitudes *de re* will both have two variants, as soon as we allow that the ascribed sentence may also be read as, "*a* knows whether the author of *Waverley* is Scott".[6]

The disambiguations are the following paraphrases:

- (*propositional, de dicto*) "*a* knows whether the proposition that Scott is the author is true or not"
- (*propositional, de re*)
  - *(i)* "*a* knows of Scott whether the proposition that he is the author is true or not"
  - *(ii)* "*a* knows of the author whether the proposition that he/she is Scott is true or not"
- (*constructional, de dicto*) "*a* knows* whether the construction constructing the proposition that Scott is the author constructs a true proposition or not"

[6] I owe to Marie Duží the observation that propositional as well as constructional attitudes *de re* are susceptible to disambiguation in terms of knowing whether *a* is the *F*, as opposed to knowing whether the *F* is *a*.

- (*constructional, de re*)
  - *(i)* "*a* knows* of Scott whether the construction constructing the proposition that he is the author constructs a true proposition or not"
  - *(ii)* "*a* knows* of the author whether the construction constructing the proposition that he/she is the author constructs a true proposition or not".

Let s/ι (Scott); $AW/\iota_{\tau\omega}$ (the individual concept of the author of *Waverley*); $=/(o\iota\iota)$; $Sub/(*_1 *_1 *_1 *_1)$; $Tr/(*_1\iota)$; $y/*_1 \rightarrow \iota$. Then:

(*propositional, de dicto*)

$$\lambda w\lambda t\ [{}^0K_{wt}\ {}^0a\ [{}^0\iota\lambda p\ [p_{wt} \wedge [[p = \lambda w\lambda t\ [{}^0AW_{wt} = {}^0s]] \vee$$
$$[p = \lambda w\lambda t\ \neg[{}^0AW_{wt} = {}^0s]] \vee [p =_1 \lambda w\lambda t\ [{}^0Und_{wt}\lambda w\lambda t\ [{}^0AW_{wt} = {}^0s]]]]]]]$$

(*propositional, de re*)

(i)
$$\lambda w\lambda t\ [{}^0K_{wt}\ {}^0a\ [{}^0\iota\lambda p\ [p_{wt} \wedge {}^2[{}^0Sub\ {}^0s\ {}^0y\ {}^0[[p =_1 \lambda w\lambda t\ [y = {}^0AW_{wt}]] \vee$$
$$[p =_1 \lambda w\lambda t\ \neg[y = {}^0AW_{wt}]] \vee$$
$$[p =_1 \lambda w\lambda t\ [{}^0Und_{wt}\lambda w\lambda t\ [y = {}^0AW_{wt}]]]]]]]]$$

(ii)
$$\lambda w\lambda t\ [{}^0K_{wt}\ {}^0a\ [{}^0\iota\lambda p\ [p_{wt} \wedge$$
$${}^2[{}^0Sub\ [{}^0Tr\ {}^0AW_{wt}]\ {}^0y\ {}^0[[p =_1 \lambda w\lambda t\ [y = {}^0s]] \vee$$
$$[p=_1 \lambda w\lambda t\ \neg[y = {}^0s]]]]]]]$$

(*constructional, de dicto*)

$$\lambda w\lambda t\ [{}^0K^*_{wt}\ {}^0a\ [{}^0\iota^*\lambda c\ [[{}^2c]_{wt} \wedge [[c =_2 {}^0[\lambda w\lambda t\ [{}^0AW_{wt} = {}^0s]]] \vee$$
$$[c =_2 {}^0[\lambda w\lambda t\ \neg[{}^0AW_{wt} = {}^0s]]] \vee$$
$$[c =_2 {}^0[\lambda w\lambda t\ [{}^0Und_{wt}\lambda w\lambda t\ [{}^0AW_{wt} = {}^0s]]]]]]]$$

(*constructional, de re*)

(i)
$$\lambda w\lambda t\ [{}^0K^*_{wt}\ {}^0a\ [{}^0\iota^*\lambda c\ [[{}^2c]_{wt} \wedge$$
$${}^2[{}^0Sub\ {}^0s\ {}^0y\ {}^0[[c =_2 {}^0[\lambda w\lambda t\ [y = {}^0AW_{wt}]]] \vee$$
$$[c =_2 {}^0[\lambda w\lambda t\ \neg[y = {}^0AW_{wt}]]] \vee$$
$$[c =_2 {}^0[\lambda w\lambda t\ [{}^0Und_{wt}\lambda w\lambda t\ [y = {}^0AW_{wt}]]]]]]]]$$

(ii)
$$\lambda w\lambda t\ [{}^0K^*_{wt}\ {}^0a\ [{}^0\iota^*\lambda c\ [[{}^2c]_{wt} \wedge {}^2[{}^0Sub\ [{}^0Tr\ {}^0AW_{wt}]\ {}^0y$$
$${}^0[[c =_2 {}^0[\lambda w\lambda t\ [y = {}^0s]]] \vee [c =_2 {}^0[\lambda w\lambda t\ \neg[y = {}^0s]]]]]]]].$$

Epistemic shift

Factivity can be restored to propositional *knowledge whether* by specifying which of *P*, *Q* (possibly both) is true, and to constructional *knowledge whether* by specifying which of *C*, *D* (possibly both) constructs a true proposition. For example, let *a* know* *whether* Scott is the author of *Waverley* and let it be true that Scott is indeed the author of *Waverley*. Then it can be validly inferred that *a* knows* *that* Scott is the author of *Waverley*. But then the question arises what the rule of factivity of knowledge is to look like for constructional knowledge. This is the problem of what I call *epistemic shift*. The problem is how to account logically for the shift from a *known* construction* to a *true proposition* or to **T**. The problem of epistemic shift arises for any hyperintensional logic within which hyperintensional objects of knowledge are not also truth-bearers. (In an epistemic logic based on possible-world semantics, propositions serve in both capacities.) The rule for propositional knowledge is obvious:

$$\frac{[{}^0K_{wt}\ {}^0a\ p]}{p_{wt}}$$

The rules for the empirical and mathematical attitudes, respectively, are as follows. Let $c/*_2 \rightarrow *_1$; ${}^2c \rightarrow o_{\tau\omega}$. Then if c is known* the proposition that is v-constructed by what is v-constructed by c is true:

$$\frac{[{}^0K^*_{wt}\ {}^0a\ c]}{[{}^2c]_{wt}}$$

Let $d/*_2 \rightarrow *_1$; ${}^2d \rightarrow o$. Then if d is known* the truth-value v-constructed by what is v-constructed by d is **T**:

$$\frac{[{}^0K^*_{wt}\ {}^0a\ d]}{{}^2d}$$

Conclusions

- To know whether *A* is to know which disjunct, if any, of (*A* or *B*) is true (inclusive disjunction, provided $B \neq \neg A$)
- A particular case is knowing which disjunct, if any, of (*A* or $\neg A$) is true (exclusive disjunction)
- '*A*' in '*knowing whether A*' is ambiguous between denoting either a proposition *P* or a construction *C*

- '*Knowing*' in '*knowing whether A*' is ambiguous between denoting a relation-in-intension between an epistemic agent and either *P* or *C*
- "*a* knows (propositionally) whether *P*" and "*a* knows* (constructionally) whether *C*" are ambiguous between interpretations *de dicto* and *de re*
- Transparent Intensional Logic can provide a principled, non-*ad hoc* logic to capture the distinctions between propositional and constructional knowledge *de dicto* and *de re*. The theory can also solve the problem of epistemic shift.

Bjørn Jespersen
Section of Philosophy
Delft University of Technology
The Netherlands
b.t.f.jespersen@tbm.tudelft.nl

References

Cresswell, M. J. (1975). Hyperintensional logic. *Studia Logica, 34*, 25–38.

Duží, M., & Jespersen, B., & Materna, P. (Ms.). *Transparent Intensional Logic*. In submission.

Genesereth, M. R., & Nilsson, N. J. (1987). *Logical Foundations of Artificial Intelligence*. Los Altos: Morgan Kaufmann.

Hart, S., & Heifetz, A., & Samet, D. (1996). 'Knowing whether', 'knowing that', and the cardinality of state spaces. *Journal of Economic Theory, 70*, 249–56.

Hintikka, J. (1975). Different constructions in terms of the basic epistemological verbs: a survey of some problems and proposals. In: *The Intentions of Intentionality and Other New Models for Modalities*, Dordrecht/Boston: D. Reidel.

Lewis, D. (1998). 'Whether' report. In: *Papers in Philosophical Logic*. Cambridge: Cambridge University Press.

Materna, P. (1997). Rules of existential quantification into 'intensional contexts'. *Studia Logica, 57*, 331–43.

Rescher, N. (2005). *Epistemic Logic*. Pittsburgh: University of Pittsburgh Press.

Tichý, P. (1988). *The Foundations of Frege's Logic*. Berlin, New York: de Gruyter.

Tichý, P. (2004). *Collected Papers in Logic and Philosophy*. Edited by V. Svoboda, & B. Jespersen, & C. Cheyne. Prague: FILOSOFIA, Czech Academy of Sciences, and Dunedin: University of Otago Press.

Recapturing the Epistemic Dimension of Logic

John T. Kearns

1. The two dimensions

Historically, the subject matter logic has had both an epistemological, or *epistemic*, and an ontological, or *ontic*, dimension. From the time of Aristotle until the mid-nineteenth century, the focus was primarily epistemic. Logic was concerned with arguments, deductions, and proofs. Following the work of Boole and Frege, logic took an ontic turn. This is perhaps most obvious in the case of Boole, who showed little interest in deductive derivations. Frege, in contrast, did have epistemic concerns. He developed the modern style of deductive system, and regarded his deductions as models of rigor, in which fallacious appeals to intuition would have no place. But Frege was concerned to reason carefully and correctly, not to study reasoning. For Frege, logic is no more a study of knowledge and how we get it than physics is a study of these things.

Perhaps Frege's conception of logic was influenced by his aversion to the psychologism that he saw in Kant's account of mathematics, especially arithmetic. In order to defend the universality of mathematics, or, anyway, arithmetic, and show that its truths would hold in any world whatever, Frege took up the project of showing that arithmetic belongs to a logic whose truths have this character. The project of logic as he understood it was to develop a perspicuous language for describing reality, a language in which grammatical categories reflect ontological ones, and to establish logical laws that have the form of statements about reality.

The ontological dimension of logic is a legitimate object of logical investigation. It was an important advance when logic was reconceived to incorporate ontology. But this advance need not, and should not, lead us to abandon the epistemic dimension of logic. *Illocutionary logic* provides the resources to accommodate both the ontic and the epistemic dimensions of logic, and I want to extol some of the virtues of illocutionary logic.

2. The logic of speech acts

Illocutionary logic as a distinct subject matter was invented, and pioneered, by John Searle and Daniel Vanderveken. However, their understanding of the field is somewhat different from my own, and there is not much overlap between the topics they investigate and the results that I have obtained. I will explain illocutionary logic from my own perspective.

Illocutionary logic is the logic of speech acts, or *language acts*. These are meaningful acts performed with expressions. There are a great variety of language acts. I shall focus on *sentential acts*, which are performed with an entire sentence. Some sentential acts are true or false, and I call these *statements*. This is a special, stipulated use for the word ‘statement,’ because the word is often used as a near synonym for ‘assertion.’ On my conception, language acts are the primary bearers of such semantic features as meaning and truth; expressions have syntactic features and can be regarded as syntactic objects.

Some sentential acts are performed with a certain illocutionary force, and constitute *illocutionary acts*. Examples are promises, warnings, assertions, declarations, and requests. Statements themselves can be used to perform a variety of illocutionary acts.

We now understand a logical theory to have three components: (1) a specialized or formal language, (2) a semantic account for this language, and (3) a deductive system for codifying some logically distinguished items in the language. A system of illocutionary logic is obtained from a standard system by making three changes:

(i) Illocutionary-force indicating expressions, *illocutionary operators*, are added to the formal language.
(ii) The semantic account of truth-conditions is supplemented with an account of the *rational commitments* generated by performing illocutionary acts. Asserting this or denying that will commit a person to make further assertions and denials.
(iii) The deductive system is modified to accommodate illocutionary operators.

3. A simple system

I will illustrate a simple system of propositional illocutionary logic. The language *L* contains atomic sentences and compound sentences obtained from them by using these connectives: *~, v, &*. (The horseshoe is a defined symbol.) The atomic and compound sentences are *plain sentences of L*. The plain sentences represent natural-language statements.

The illocutionary operators are the following:

$\vdash$ – the sign of assertion	$\dashv$ – the sign of denial
$\llcorner$ – the sign of supposing true	$\urcorner$ – the sign of supposing false

A plain sentence prefixed with an illocutionary operator is a *completed sentence of L*; there are no other completed sentences. Completed sentences represent illocutionary acts.

The language *L* contains two kinds of logical operators. The logical operators in plain sentences are the connectives, which represent things we actually say in making statements. These things we say belong to the statements that we make. But the illocutionary operators don't represent things we say. They represent things we *do* with statements. We may sometimes use expressions to make explicit just what we are doing with a statement, as when we say "suppose." But we generally don't say "I assert" in making an assertion, and we often don't say "suppose" when we are supposing something.

An assertion is understood to be an act of producing and coming to accept a statement as representing what is the case, or an act of producing and reaffirming one's (continued) acceptance of statement. (An assertion in this sense doesn't need an audience, and all such assertions are sincere.) A denial is an act of coming to reject a statement (for being false), or an act reflecting one's rejection of the statement. Supposing a statement *A* to be true or false is not a subjunctive or counterfactual consideration of how things *would be if A were true*. Instead we consider how things *are*, if in addition to what we know or believe, *A* is also true. Once made, a supposition remains in force until it is discharged (canceled) or simply abandoned. An argument which begins with assertions and denials can reach a conclusion which is an assertion or denial. But we cannot correctly begin with at least one supposition, and conclude with an assertion or denial. The conclusion must have the force of a supposition, and will be called a supposition.

The semantic account for the language *L* has two tiers, or levels. The first tier presents the ontology encoded by the language, giving truth conditions of plain sentences and the statements that these represent. An interpreting function assigns truth and falsity to the atomic plain sentences, and determines a *truth-value valuation* of all the plain sentences.

The second tier of the semantic account is epistemic, and deals with *rational commitment*. This is a commitment to perform or not perform some act, or to continue in some state or condition like that of accepting a statement. Commitments generated by performing acts of assertion, denial, or supposition are conditional rather than absolute. A person who accepts a statement will be committed to accept (or to reaffirm her continued acceptance of) some further statement, but only if the matter comes up and she chooses to think about it, and only so long as she continues to accept the first statement.

The second semantic level depends on the first, for the language user must understand the truth conditions of the statements she asserts, denies, or supposes. Since a commitment to perform or not perform an act is always *someone's* commitment, we develop the commitment semantics for an idealized person, the *designated subject*. We consider the designated subject at some particular moment. There are certain statements which she has thought about and accepted, which she remembers and continues to accept. There are similar statements that she has considered and rejected. These *explicit* beliefs and disbeliefs commit her to accept further statements and to reject further statements. We use '+' for the value of assertions and denials that she is committed, at that moment, to perform.

A *commitment valuation* assigns this value to some of the assertions and denials in *L*. A commitment valuation *V* is *based on* an interpreting function *f* if, and only if (from now on: iff) (i) If $V(\vdash A) = +$, then $f(A) = \mathrm{T}$, and (ii) If $V(\dashv A) = +$, then $f(A) = \mathrm{F}$. A commitment valuation is *coherent* iff it is based on an interpreting function.

A coherent commitment valuation V_0 can be understood to register the designated subject's explicit beliefs and disbeliefs at a given time. The *commitment valuation determined by* V_0 is the function *V* such that (i) $V(\vdash A) = +$ iff *A* is true for every interpreting function on which V_0 is based, and (ii) $V(\dashv A) = +$ iff *A* is false for every interpreting function on which V_0 is based. *V* indicates the assertions and denials which the designated subject is committed to perform by her explicit beliefs and disbeliefs. An *acceptable* commitment valuation is one determined by a coherent commitment valuation. Acceptable valuations have these matrices (the letter '*b*' stands for *blank* – for those positions in which no value is assigned):

$\vdash A$	$\vdash B$	$\dashv A$	$\dashv B$	$\vdash {\sim} A$	$\dashv {\sim} A$	$\vdash [A \,\&\, B]$	$\dashv [A \,\&\, B]$	$\vdash [A \vee B]$	$\dashv [A \vee B]$
+	+	*b*	*b*	*b*	+	+	*b*	+	*b*
+	*b*	*b*	*b*	*b*	+	*b*	*b*	+	*b*
+	*b*	*b*	+	*b*	+	*b*	+	+	*b*
b	+	*b*	*b*	*b*	*b*	*b*	*b*	+	*b*
b	*b*	*b*	*b*	*b*	*b*	*b*	+,*b*	+,*b*	*b*
b	*b*	*b*	+	*b*	*b*	*b*	+	*b*	*b*
b	+	+	*b*	+	*b*	*b*	+	+	*b*
b	*b*	+	*b*	+	*b*	*b*	+	*b*	*b*
b	*b*	+	+	+	*b*	*b*	+	*b*	+

The first row shows the commitments of accepting/asserting both *A, B*, the second shows the commitments of accepting *A* and neither accepting nor rejecting *B*. Etc. In some cases, the values (or non-values) of assertions and denials of simple sentences are not sufficient to determine the values of assertions and denials of compound sentences. For example, if ⊣*A* and ⊣*B* have no value, and *A, B* are irrelevant to one another, then ‘⊣*[A & B]*’ should have no value. But if ⊣A, ⊣~A have no value, the completed sentence ‘⊣*[A & ~A]*’ will have value +.

4. Semantic concepts

The language *L* and the two tiers of the semantic account for *L* provide the conceptual resources to understand, explain, and explore many logic-related phenomena. For example, it is common to attempt to distinguish inductive from deductive arguments by considerations relating to truth and probability. But these are first-tier concepts. To properly distinguish deductive from inductive arguments, we must employ second-tier concepts. What characterizes deductive arguments, or correct deductive arguments, is that they are based on rational commitment. In contrast, performing the premiss acts of an inductively satisfactory argument won't commit the arguer to performing the conclusion act, the premiss acts only *authorize* him, to a greater or lesser degree, to perform the conclusion act.

The truth conditions of a statement determine what the world must be like for the statement to be true. In standard logic, many concepts are defined in terms of truth conditions. For example, a set *X* of plain sentences of *L (truth-conditionally) implies* a plain sentence *A* iff there is no interpreting function of *L* for which every sentence in *X* has value T, while *A* has value F.

An illocutionary counterpart of implication links completed sentences of *L* and the illocutionary acts that these represent. Instead of speaking of *illocutionary implying*, I prefer to speak of *logical requiring*. In order to define this concept, some preliminary definitions are required.

Let V_0 be a coherent commitment valuation of *L*, let *V* be the commitment valuation determined by V_0, and let *A* be a completed sentence of *L* that is either an assertion or denial. Then V_0 *satisfies A* iff *V(A)* = +.

Suppositions are not assigned values by commitment valuations. But supposing certain statements will commit a person to supposing others. In supposing a statement either true or false, we consider truth values to determine what further statements we are committed to suppose.

Let *f* be an interpreting function of *L*, and let *A, B* be plain sentences of *L*. Then (i) *f satisfies* ⌞*A* iff *f(A)* = T, and (ii) *f satisfies* ¬*B* iff *f(B)* = F.

Let *f* be an interpreting function of *L* and *V* be a commitment valuation of *L* based on *f*. Then <*f, V*> is a *coherent pair for L.*

Let <*f, V*> be a coherent pair (for *L*), and let *A* be a completed sentence of *L*. Then <*f, V*> *satisfies A* iff either (i) *A* is an assertion or denial and *V* satisfies *A*, or (ii) *A* is a supposition and *f* satisfies *A*.

Let *X* be a set of completed sentences of *L* and let *A* be a completed sentence of L. Then *X logically requires A* iff (i) A is an assertion or denial and there is no coherent commitment valuation which satisfies the assertions and denials in X but does not satisfy A, or (ii) A is a supposition and there is no coherent pair for L which satisfies every sentence in X, but fails to satisfy A. If a set X of completed sentences logically requires a further completed sentence, then anyone performing the acts represented by the sentences in the set is committed to perform the act represented by the further sentence.

It is necessary to have two clauses in the definitions of illocutionary implication, because if A is an assertion or denial, its value is independent of the values assigned to suppositions. For example, consider these completed sentences:

$$\llcorner A,\ \neg A,\ \vdash B;\ \vdash [B \mathbin{\&} A]$$

There is no coherent pair which satisfies $\llcorner A,\ \neg A,\ \vdash B$ and fails to satisfy '$\vdash [B \mathbin{\&} A]$,' because there is no coherent pair which satisfies $\llcorner A,\ \neg A,\ \vdash B$. However, the first three sentences do not logically require '$\vdash [B \mathbin{\&} A]$,' for suppositions make no "demands" on assertions and denials. Incoherent suppositions logically require that we suppose true and suppose false every plain sentence, but they do not require that we assert or deny anything.

5. Reasoning

The natural deduction system S uses tree proofs. Steps in a proof are completed sentences, and the rules take account of both truth conditions and illocutionary force. An initial step in a tree proof is an assertion, a denial, a positive supposition, or a negative supposition. An initial assertion or denial is not a hypothesis of the proof. Instead, an initial assertion or denial should express knowledge or justified (dis)belief of the arguer. Not every asserted sentence is eligible to be an initial assertion in a proof constructed by a given person. In contrast, any supposition can be an initial supposition. Initial suppositions are hypotheses of the proof.

The rules of inference for conjunction are *elementary:*

& Introduction & Elimination

$$\frac{\vdash/\llcorner A \quad \vdash/\llcorner B}{\vdash/\llcorner [A \mathbin{\&} B]} \qquad \frac{\vdash/\llcorner [A \mathbin{\&} B]}{\vdash/\llcorner A} \qquad \frac{\vdash/\llcorner [A \mathbin{\&} B]}{\vdash/\llcorner B}$$

The expression '$\vdash/\llcorner$' indicates that the illustration holds both for assertions and positive suppositions. For each rule, the conclusion is an assertion only if all premisses are assertions. If at least one premiss is a supposition, then the conclusion must be a supposition.

The following arguments are incorrect:

$$\frac{\llcorner A \quad \llcorner B}{\vdash [A \,\&\, B]} \qquad\qquad \frac{\vdash A \quad \llcorner B}{\vdash [A \,\&\, B]}$$

even though they are truth-preserving, for a supposition as premiss will not support a conclusion which is an assertion. These arguments are correct:

$$\frac{\vdash A \quad \vdash B}{\vdash [A \,\&\, B]} \qquad\qquad \frac{\vdash A \quad \llcorner B}{\llcorner [A \,\&\, B]} \qquad\qquad \frac{\llcorner A \quad \llcorner B}{\llcorner [A \,\&\, B]}$$

Elementary rules move directly from assertions, denials, or suppositions as premisses to an assertion, denial, or supposition as conclusion. Non-elementary rules include at least one premiss which is a subproof, and cancel, or discharge, a hypothesis of the subproof. The rule $\supset$ *Introduction* is a non-elementary rule:

$$\frac{\begin{array}{c}\{\llcorner A\}\\ \llcorner B\end{array}}{\vdash/\llcorner\, [A \supset B]}$$

The premiss of this rule is an entire subproof with '$\llcorner A$' as a hypothesis, and '$\llcorner B$' as conclusion. Following a use of this rule, the hypothesis '$\llcorner A$' is canceled. The conclusion is an assertion if the subproof contains only one uncanceled hypothesis, '$\llcorner A$.' If the subproof contains additional hypotheses, the conclusion is a (positive) supposition.

Given sentences *A, B*, the following is an example of a simple argument in the deductive system *S*:

$$\cfrac{\cfrac{\cfrac{\llcorner A \quad \overset{x}{\llcorner B}}{\llcorner [A \,\&\, B]}\,\&I}{\llcorner A}\,\&E}{\llcorner [B \supset A]}\supset I,\ \textit{cancel } `\llcorner B\text{'}$$

An '*x*' is placed above canceled hypotheses. This argument shows that the premiss '$\llcorner A$' logically requires the conclusion '$\llcorner [B \supset A]$.'

Since illocutionary logic is concerned with epistemology, and correct arguments, as well as being concerned with ontology and logical laws, it is important that arguments in the deductive system be perspicuous, and that the difference between direct and indirect arguments be clearly indicated. From the perspective of illocutionary logic, arguments and proofs are not simply instruments for establishing various results; they are also objects to be studied. The tree proofs and the illocutionary operators play an important role in achieving this goal.

6. Conditional assertions

A theory, or system, of illocutionary logic has an empirical character. It is intended to represent, to *capture*, our actual practice in using language. It is true that when it comes to reasoning, and arguments, we are concerned with how people *should* reason rather than with how people *in fact* reason. But the practice of using language is *normative* in the sense that there are norms for correct speaking, and for constructing correct arguments. These norms are implicit in the practice, in spite of the fact that people often speak and reason in ways that violate the norms. Systems of illocutionary logic are intended to illuminate and explain our practice in using language, and must be judged on the basis of whether they do fit this practice.

By recognizing that conditional assertions are a distinctive form of illocutionary act, a form intended to establish a commitment from accepting or supposing true the antecedent to accepting or supposing true the consequent, illocutionary logic is able to provide an intuitively satisfactory treatment of conditional assertions. This account is part of a larger account of conditional illocutionary acts of various kinds, like conditional promises, conditional warnings, conditional requests, and many more. Standard accounts of conditionals cannot accommodate these other kinds of conditional acts. I described the illocutionary account of conditional assertions at *Logica 2003,* and a longer account is soon to appear in *Linguistics and Philosophy*.

7. Semantic modalities

Distinctive concepts of necessity and possibility are associated with each semantic level of an illocutionary logical theory. A statement is *ontically necessary* if its truth conditions cannot fail to be satisfied. Ontic necessity is *analytic truth*. Whether a statement is ontically necessary, or analytic, depends on what might be called the "total meaning" of the statement. A statement is *ontically possible* if it is not contradictory. An illocutionary logical version of the modal system *S5* is the appropriate system for exploring analytic truth and logical truth.

Epistemic necessity is relative to a person, or a community, and that person's or that community's knowledge at a given time. It is most convenient to develop an illocutionary system of epistemic modal logic from the perspective of the designated subject. A statement is epistemically necessary at a given time if its assertion follows, in the sense of commitment, from the designated subject's knowledge at that time. And a statement is *epistemically possible* at a time if it is not ruled out by the designated subject's knowledge at that time.

Illocutionary logic provides the most convenient, and intuitive, framework for developing epistemic modal logic. If we consider a context in which assertions have the force of knowledge claims, it is clear that these inference principles are correct.

$\Box$ Introduction

$$\frac{\vdash A}{\vdash \Box A}$$

$\Box$ Elimination

$$\frac{\vdash \Box A}{\vdash A}$$

For positive supposition, we also have a principle *Elimination*:

$$\frac{\llcorner \Box A}{\llcorner A}$$

Someone who asserts a statement with the force of a knowledge claim is committed to claiming that the statement follows from what she knows. And if she claims that a statement follows from what she knows, then she is clearly committed to assert that statement itself with the force of a knowledge claim. Similarly, to suppose that statement *A* follows from current knowledge is to suppose that *A* is true. But to suppose that *A* is true is not to suppose that *A* follows from current knowledge. Instead of $\Box$ *Introduction*, we need these principles for supposition:

(T)

$$\frac{\vdash/\llcorner \Box A \quad \vdash/\llcorner \Box [A \supset B]}{\llcorner \Box B}$$

(S4)

$$\frac{\llcorner \Box A}{\llcorner \Box \Box A}$$

I spoke about this at *Logica 2005*, when I talked about the logical difference between knowledge and justified belief. The illocutionary version of epistemic modal logic provides an explanation which dissolves the puzzle in Moore's paradox, and explains what is going on in the surprise execution puzzle or paradox.

8. Referring

The topic of referring has been important in logic, at least since the work of Frege and Russell, although neither Frege nor Russell used the word ‘refer’ as a technical term for a type of speech act. However, both men were concerned with our use of language to “get at” things in the world. In “On Sense and Reference,” Frege claims that the senses of proper names and definite descriptions provide *modes of access* to their referents, while Russell believed that it is the expressions he called *logically proper names* that directly connect our statements with objects of our acquaintance.

There are various puzzles associated with singular terms and statements made with them. Perhaps the original puzzle is that noted by Frege, who wanted to understand why some identity statements seem trivial, while others are informative; even though all true identity statements simply say that a thing is itself. Another puzzle is to explain how a descriptive singular term can sometimes be used to identify an object which doesn’t satisfy the description (as one might use ‘the man with a martini’ for a person who isn’t drinking a martini). Or what is the difference between expressions which provide the direct access to an object which has been characterized as rigid designation, and expressions which designate non-rigidly?

Various theories have been proposed to explain the workings of names, definite descriptions, demonstratives, indexicals, and other singular terms. Some of these theories have been extended to cover common nouns and adjectives. Logical theories of great complexity have been devised to explain a practice that doesn’t seem to ordinary language users to be either mysterious or especially complicated.

No entirely successful or satisfactory account has been provided which explains the uses of singular terms. Certainly no logical theory provides much insight. This is partly due to the standard understanding of logic, which fails to adequately accommodate both the ontic and epistemic dimensions of logic. Most singular terms have at least two distinct uses, the *referring* use and the *predicative* use. For some singular terms, the referring use is primary, and there are singular terms which are used only to refer. In a logical theory, the predicative use of singular terms is best understood, and explicated, at the ontic level of logical theory. The referring use of a singular term is epistemic. Each person in referring exploits features which are peculiar to herself. Although different people can assert the same statement, they can’t make the same assertion. Jones’ assertion commits Jones but not Smith, while Smith’s assertion plays a similar role for Smith. And when Jones refers to someone, say Napoleon, he exploits a connection linking him to Napoleon in directing his attention to Napoleon. Smith can also refer to Napoleon, but his connection to Napoleon is different from Jones’.

In a first-order illocutionary theory, the semantic difference between predicative and referring uses of a singular term should be marked syntactically, even though this is not done in English. I mark the distinction by underlining individual constants used to refer, and use plain individual constants to represent their predicative use. A constant can be used predicatively to say that an individual satisfies criteria associated with the constant, or, if the constant is a proper name, that the individual is called by that name. A constant can be used predicatively to talk about whatever individual (uniquely) satisfies the criteria or is called by that name.

A person who performs a referring act uses a singular term to direct her attention to a particular object. In doing this, she exploits a connection (a *mode of access*) that she knows about linking her to that object. This connection might be based on her own experience of the object, or be derivative from the connections of other people who have informed her of the object. There are still other sources of these connections. Since the connections may not be supplied as a matter of language, we can explain how a person might use a singular term like 'the man drinking a martini' to refer to something other than a man drinking a martini.

A system of first-order illocutionary logic includes a domain of modes of access as well as a domain of individuals. An interpreting function for the language assigns individuals to some or all individual constants, while a commitment valuation assigns modes of access to some or all individual constants. The modes of access are construed as functions yielding individuals as values. I haven't the time or the space here to develop the formal details of a suitable logical treatment of our use of singular terms. All that I want to note in this place is that an adequate account of referring expressions and referring acts belongs to the epistemic level of logic, not the ontic level. Once logic is enlarged to accommodate the epistemic, it is a relatively straightforward task to devise a simple and intuitive account that accommodates both referring and non-referring uses of singular terms.

9. Summing up

Logic is a very old academic subject, and field of research. But there are many new topics and new areas for logical research. Illocutionary logic, the logic of speech acts or language acts, is one of these. Illocutionary logic accommodates, or incorporates, standard logic, and provides the resources to integrate logic's traditional concern with epistemology into modern logical theory. This more adequate conception of logic provides a perspective which allows us to solve or dissolve certain long-standing problems, and to carry out research which illuminates our linguistic and cognitive practices. I hope I can encourage other students and scholars to investigate this relatively unexplored area of logic.

John T. Kearns
Department of Philosophy and Center for Cognitive Science
University at Buffalo, State University of New York
Buffalo, NY 14260
kearns@buffalo.edu

On Circular Acceptance*

Katarzyna Kijania-Placek

1. Introduction

This paper is a part of a larger project,[1] in which I attempt to give a logical analysis of the consensus criterion of truth, i.e. the criterion according to which a given sentence is true when it is universally accepted by the members of some group. In this analysis, universal agreement is treated as a special case of majority agreement, and the criterion is stated as follows

> A given sentence, *p*, is true when it is accepted by the majority of some group, *B*.

In my previous work I assumed that the expression "person *x* accepts sentence *p*", where *p* is atomic, is basic and have not analyzed it. This means that I excluded from the scope of investigation all issues having to do with what inclines people to accept particular atomic sentences, and in particular what criteria they use and whether they use any criteria at all. I also left out of consideration the issue of the rationality of decisions regarding the acceptance or rejection of atomic sentences. My intention now is to remain neutral in these respects. There is, however, one aspect of these considerations, which I would like to focus upon, since its character is purely logical and, if left dubious, might undermine the whole project.

2. Majority agreement applied to the case of atomic sentences

The issue concerns the use of the majority criterion itself in deciding about the acceptance of atomic sentences. We might wonder whether in the case when all members of a group base their decisions entirely on decisions made by other members, the whole process does not become logically corrupt due to the circularity of this case. What I will try to show below is that even if we take the

* The research for this paper was supported by the Foundation for Polish Science. Thanks to Anil Gupta and Nuel Belnap as well as the participants of Logica 2006 for comments and suggestions.

[1] See Kijania-Placek (2000).

most extreme case, in which the acceptance of an atomic sentence by a person is, for each person, *defined* in terms of the acceptance of the sentence by the other members of a group, this does not by itself lead to logical inconsistencies. In my work I will rely on the theory of circular definitions developed by Anil Gupta and Nuel Belnap.

First, let me define such a concept of acceptance, to expose its circularity. Let a, b, c, d, e and $a_1, \ldots, a_n$ refer to the members of a group B. We are dealing with finite groups only. Assume additionally that a_i's decision is defined in terms of the decisions of all the other members of the group. Otherwise, the authority of some members might save us from the circularity and the problem does not arise. We will consider the more reasonable case, in which the person whose acceptance we are checking is not explicitly taking his or her own decision into consideration when counting the majority of the group. This decision is not only more reasonable but logically relevant. The other case, in which a simple majority of the whole group matters, does not lead to a circular definition. The majority is then an objective characteristic of the group and does not vary with different member's opinions being taken into account. Let $A_p(a_i)$ be an abbreviation for the expression „the sentence p is accepted by person a_i", with a_i ranging over the members of a particular group. We can define circular agreement by:

(1) $A_p(a_i) =_{\mathrm{df}}$ Most members of the group B_i accept sentence p,

where B_i stands for the members of the group B except for a_i. Stated as it is, the definition is not explicitly circular. However, since a_i is a member of the group as well, and since all other members base their acceptance on that of other members, the majority depends, in some cases, on whether a_i him/herself accepts the sentence or not. This makes the definition circular. I will not show the circularity of the definition in the general case but proceed to examples, since in order to show that a concept is circular it is sufficient to show its circularity for a particular situation.

2.1. Example

Consider a group of close friends, a, b, c and d standing for Arthur, Betty, Chris and Daniel, who value only the opinions of the other members of this closed group as far as fashion, partying, music, etc. are concerned. And they are pretty conformist – they follow the majority. Let p stand for the sentence „The graduation party is worth attending". For our group, whether the party is worth attending depends on what other members of the group think about it; in fact, if most members of the group think it worth attending, this *makes* it worth attending. From (1) we can infer partial definitions of their respective acceptance of sentence p:

$$A_p(a) =_{\text{df}} (A_p(b) \wedge A_p(c)) \vee (A_p(b) \wedge A_p(d)) \vee (A_p(c) \wedge A_p(d))$$
$$A_p(b) =_{\text{df}} (A_p(a) \wedge A_p(c)) \vee (A_p(a) \wedge A_p(d)) \vee (A_p(c) \wedge A_p(d))$$
$$A_p(c) =_{\text{df}} (A_p(a) \wedge A_p(b)) \vee (A_p(a) \wedge A_p(d)) \vee (A_p(b) \wedge A_p(d))$$
$$\mathrm{A}_p(d) =_{\text{df}} (A_p(a) \wedge A_p(b)) \vee (A_p(a) \wedge A_p(c)) \vee (A_p(b) \wedge A_p(c))$$

Even though Arthur's acceptance of p is not explicitly defined in terms of his own acceptance, but in terms of the acceptance of his friends, their acceptance is defined in terms of his acceptance, which makes the definition circular.

Such circular definitions were believed to be formally inadequate, as they do not provide definite extensions to the defined concepts and lead to contradictory conclusions in certain circumstances. But the work of Anil Gupta and Nuel Belnap shows that we can accept circular definitions and work with them. Although circular definitions do not provide extensions for the concepts defined, we can make semantic sense of them without risking a contradiction. I will briefly introduce Gupta and Belnap's theory of circular definitions, or rather those parts of it that are relevant to our simple case.

3. Circular definitions

Consider an example given by Gupta in „On circular concepts"

(*) $F(x) =_{\text{df}} (x = \text{Socrates} \vee (x = \text{Plato} \wedge {\sim}F(x)))$

This definition is circular, as it contains the definiendum, F, in its definiens. Thus, no classical extension can be consistently assigned to it. Nothing categorical can be said about Plato. That does not mean, however, that the definition is useless. We can consistently assign hypothetical extensions to F: assuming a hypothesis about the extension of F, for example the hypothesis that the extension is empty, we can conclude on the basis of the definition that its extension should instead be {Socrates, Plato}. This being our new and better hypothesis, we can revise it again on the basis of the definition and obtain another, still better hypothesis. This basic intuition is the source of the concept of a rule of revision, which is a rule for moving from one hypothesis to another. Definition (*) generates the following rule of revision δ:

input (antecedent hypothesis)	output (consequent hypothesis)
$\varnothing$	{Socrates, Plato}
{Socrates}	{Socrates, Plato}

{Plato}		{Socrates}
{Socrates, Plato}	{Socrates}	
{Arystotle}	{Socrates, Plato}	
{Arystotle, Plato}	{Socrates}	
...		

In general, a rule of revision δ for a circular definition *D*:

(*D*) $G(x) =_{\mathrm{df}} \phi(x, G)$

is an operation on the powerset of the domain such that for all objects *d* in the domain and every hypothesis *h* about the extension of *G*, $d \in \delta(h)$ iff *d* satisfies $\phi(x, G)$ assuming that *h* is the extension of *G*. The rule of revision takes an antecedent hypothesis about the extension of a concept and yields an improved consequent hypothesis about the extension. Taking all possible extensions as initial hypotheses, we obtain by the rule of revision infinite revision sequences. All such sequences form a revision process for the rule. A central thesis of the theory of circular definitions developed by Gupta and Belnap is that the meaning of a circular concept yields this rule of revision, instead of a way of demarcating objects into those that fall under the concept and those that do not.

Even though the revision process has a hypothetical character, we can sometimes make categorical judgments on its basis. Some of the hypotheses may turn out to be reflexive, i.e. the repeated application of the revision rule to the hypothesis results in the original hypothesis.[2] There are two reflexive hypotheses: {Socrates} and {Socrates, Plato} for definition (*), because δ^2({Socrates}) = {Socrates} and δ^2({Socrates, Plato}) = {Socrates, Plato}. These reflexive hypotheses are those which occur again and again in the revision process; others do not survive the process of improving the hypothesis about the extension of the concept. If a claim holds under all reflexive hypotheses, it can be said to be true categorically; if it holds under none of them, it is false categorically. Thus, we can conclude from the revision process based on (*) that Socrates is categorically *F* and that everything other than Socrates or Plato is categorically not *F*. We cannot conclude anything categorical about Plato.

---

[2] There is a natural number *n*, such that $\delta^n(h) = h$, where $\delta^n(h)$ is the result of *n* applications of δ to *h*. See Gupta (2000), p. 125.

4. Finite circular definitions

These concepts of categoricity, although not satisfactory for the general theory of definitions,[3] work well for finite definitions. Gupta defined finite definitions only for first order languages,[4] but we can directly generalize his definition to all languages

Definition (1). A definition D in a language L is finite in L iff, for all interpretations of the non-logical symbols of the language L other than those defined by D, there is a natural number n such that for all hypotheses h about the extension of the defined concept, $\delta^n(h)$ is reflexive.

By mimicking Gupta's argument of (2000, p. 126) we can show that definition (1) is finite.[5] The concept A_p (accepting sentence p) defined by (1) is always relativised to a group of people, B. The groups are finite and as a result there are finitely many hypotheses about the extension of the concept A_p. For every interpretation of B there is a number 2^n, where n is the number of members of the group, such that for each hypothesis h about the interpretation of A_p, $\delta^{2^n}(h)$ is reflexive – the hypothesis at stage 2^n of each revision sequence for (1) is reflexive. Suppose it is not. Then none of the hypotheses occurring at earlier stages are reflexive, because a hypothesis occurring after a reflexive hypothesis is always reflexive. This means that all the preceding hypotheses are distinct and that they exhaust the range of possible hypotheses. It follows that the hypothesis at stage 2^{n+1} must be one of those occurring earlier, which contradicts the claim that none of them was reflexive. Hence, for all hypotheses $\delta^{2^n}(h)$, is reflexive.

5. $A_p(x)$ defined as a circular concept

In our example, the four elements – group members – yield 16 initial hypotheses: $\varnothing$, $\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$, $\{a, b\}$, $\{a, c\}$, $\{a, d\}$, $\{b, c\}$, $\{b, d\}$, $\{c, d\}$, $\{a, b, c\}$, $\{a, b, d\}$, $\{a, c, d\}$, $\{b, c, d\}$, $\{a, b, c, d\}$. It may be the case that initially nobody

---

[3] See Gupta and Belnap (1993).

[4] Gupta (2000) and (2004).

[5] Gupta used this argument to show the finiteness of two classes of first order definitions. The first class contains definitions of the form:

$$G(x) =_{\text{df}} (x = a_1 \vee x = a_2 \vee \ldots \vee x = a_n) \wedge \varphi(x, G),$$

with φ first-order. In fact, his argument generalizes to arbitrary φ and generalized in this way applies to definition (1) reformulated as:

(1*) $A_p(x) =_{\text{df}} (x = a_1 \vee x = a_2 \vee \ldots \vee x = a_n) \wedge$ most members of the group consisting of $a_1, a_2, \ldots, a_n$ except for x accept sentence p.

finds the party worth attending, that only Arthur does, or only Betty, or only Chris, or only Daniel; it maybe that just Arthur and Betty find the party worth attending, or just Arthur and Chris, or just Arthur and Daniel, or just Betty and Chris, or just Betty and Daniel, or just Chris and Daniel; or it maybe the case that initially exactly Arthur, Betty and Chris find the party worth attending, or exactly Arthur, Betty and Daniel, or exactly Arthur, Chris and Daniel, or exactly Betty, Chris and Daniel; the final hypothesis is that all four of them find the party worth attending. These initial hypotheses, together with definition (1), yield a rule of revision δ_I. If initially nobody accepted p, then none of the partial definientia holds, so

$$\delta_I(\varnothing) = \varnothing$$

If initially only Arthur found the party worth attending, his vote would not suffice to convince others and, because in the second step of the revision process his own opinion depends on the opinions of others, he himself would be willing to change his mind and abstain from finding the party worth attending.[6] As a result of this, at the second stage of the revision process nobody would find the party worth attending.

$$\delta_I(\{a\}) = \varnothing$$

The hypotheses that initially only Betty, or that only Chris, or that only Daniel finds the party worth attending result in analogous outcomes:

$$\delta_I(\{b\}) = \varnothing$$
$$\delta_I(\{c\}) = \varnothing$$
$$\delta_I(\{d\}) = \varnothing$$

The interesting case comes when initially only Arthur and Betty find the party worth attending. Both Chris and Daniel, the remaining members, who initially were not interested in attending the party, or were not aware of whether they are, would find the party worth attending, because the majority of the group does. Recall that we have assumed that they do not count themselves, so Arthur and Betty form the majority. Both Chris and Daniel would now be inclined to accept p, but Arthur and Betty would do the opposite. Arthur, excluding his own

[6] To tell a story for the sake of an example, I use grammatical tenses like “was willing to attend the party” and “will change his mind”. But we have to bear in mind that the case considered is the logical possibility of simultaneous interdependence when the revision process is happening “at the same time”, or better, at no time at all. In the case that the process really takes place in time – is extended in time – no circularity arises. Thanks to Nuel Belnap for encouraging me to make this point explicit.

opinion, would find only Betty willing to go to the party. Finding himself in the minority, he would be willing to change his mind. The same reasoning applies to Betty, so she would change her mind as well. We would have only Chris and Daniel finding the party attractive at the next stage.

$$\delta_1(\{a, b\}) = \{c, d\}$$

The arguments works in the opposite direction for the hypothesis that initially only Chris and Daniel find the party worth attending, so

$$\delta_1(\{c, d\}) = \{a, b\}$$

And analogously,

$$\delta_1(\{a, c\}) = \{b, d\}$$
$$\delta_1(\{a, d\}) = \{b, c\}$$
$$\delta_1(\{b, c\}) = \{a, d\}$$
$$\delta_1(\{b, d\}) = \{a, c\}$$

If exactly Arthur, Betty and Chris find the party attractive (or any other three of them), the situation immediately leads to a consensus. All three are secured in their opinion by the opinion of the other two. And the fourth, Daniel, changes his mind, realizing that he alone is of a different opinion.

$$\delta_1(\{a, b, c\}) = \{a, b, c, d\}$$
$$\delta_1(\{a, b, d\}) = \{a, b, c, d\}$$
$$\delta_1(\{a, c, d\}) = \{a, b, c, d\}$$
$$\delta_1(\{b, c, d\}) = \{a, b, c, d\}$$

Similarly for the last hypothesis: the opinion of each of them is secured by the opinion of the other members, as all of them find the party worth attending under the last hypothesis.

$$\delta_1(\{a, b, c, d\}) = \{a, b, c, d\}$$

The rule of revision δ_1 for (1) yields the following 16 revision sequences:

s_1: $\emptyset, \emptyset, \emptyset, \ldots$
s_2: $\{a\}, \emptyset, \emptyset, \ldots$
s_3: $\{b\}, \emptyset, \emptyset, \ldots$
s_4: $\{c\}, \emptyset, \emptyset, \ldots$
s_5: $\{d\}, \emptyset, \emptyset, \ldots$

s_6: $\{a, b\}, \{c, d\}, \{a, b\}, \{c, d\},...$
s_7: $\{a, c\}, \{b, d\}, \{a, c\}, \{b, d\},...$
s_8: $\{a, d\}, \{b, c\}, \{a, d\}, \{b, c\},...$
s_9: $\{b, c\}, \{a, d\}, \{b, c\}, \{a, d\},...$
s_{10}: $\{b, d\}, \{a, c\}, \{b, d\}, \{a, c\},...$
s_{11}: $\{c, d\}, \{a, b\}, \{c, d\}, \{a, b\},...$
s_{12}: $\{a, b, c\}, \{a, b, c, d\}, \{a, b, c, d\},...$
s_{13}: $\{a, b, d\}, \{a, b, c, d\}, \{a, b, c, d\},...$
s_{14}: $\{a, c, d\}, \{a, b, c, d\}, \{a, b, c, d\},...$
s_{15}: $\{b, c, d\}, \{a, b, c, d\}, \{a, b, c, d\},...$
s_{16}: $\{a, b, c, d\}, \{a, b, c, d\}, \{a, b, c, d\},...$

Nobody categorically accepts sentence *p* according to definition (1). That should not be surprising, since the mutual interdependence of the acceptance of *p* by a member of a group on its acceptance by other members was bound to result in the members sometimes changing their views on and off without being able to make up their mind and sometimes agreeing in their opinion. The revision process, yielded by the meaning that we have assigned to A_p, gives us a clear description of these interdependencies. Sequences 1–5 and 12–16 stabilize after the second application of the rule of revision. The interesting case, where the circularity of the definition is at work, is exhibited in sequences 6–11, where members of the group who accept *p* at one stage do not accept it at the next stage because their own acceptance of the sentence turns out to be critical to the majority *via* the decisions taken by the others. This is clearly a by-product of the fact that the number of members in the group is odd. Consider a group *B'* of five people: a', b', c', d', e'. The concept A'_p relativized to the group is defined by a definition parallel to (1):

(1') $A'_p(a_i) =_{\mathrm{df}}$ Most members of the group B'_i accept sentence *p*.

The definition generates the rule of revision δ', on which the following 32 revision sequences are based:

s'_1: $\varnothing, \varnothing, \varnothing,...$
s'_2: $\{a'\}, \varnothing, \varnothing,...$
s'_3: $\{b'\}, \varnothing, \varnothing,...$
s'_4: $\{c'\}, \varnothing, \varnothing,...$
s'_5: $\{d'\}, \varnothing, \varnothing,...$
s'_6: $\{e'\}, \varnothing, \varnothing,...$
s'_7: $\{a', b'\}, \varnothing, \varnothing,...$
s'_8: $\{a', c'\}, \varnothing, \varnothing,...$
s'_9: $\{a', d'\}, \varnothing, \varnothing,...$

s'_{10}: $\{a', e'\}, \varnothing, \varnothing, \ldots$
s'_{11}: $\{b', c'\}, \varnothing, \varnothing, \ldots$
s'_{12}: $\{b', d'\}, \varnothing, \varnothing, \ldots$
s'_{13}: $\{b', e'\}, \varnothing, \varnothing, \ldots$
s'_{14}: $\{c', d'\}, \varnothing, \varnothing, \ldots$
s'_{15}: $\{c', e'\}, \varnothing, \varnothing, \ldots$
s'_{16}: $\{d', e'\}, \varnothing, \varnothing, \ldots$
s'_{17}: $\{a', b', c'\}, \{d', e'\}, \varnothing, \varnothing, \ldots$
s'_{18}: $\{a', b', d'\}, \{c', e'\}, \varnothing, \varnothing, \ldots$
s'_{19}: $\{a', b', e'\}, \{c', d'\}, \varnothing, \varnothing, \ldots$
s'_{20}: $\{a', c', d'\}, \{b', e'\}, \varnothing, \varnothing, \ldots$
s'_{21}: $\{a', c', e'\}, \{b', d'\}, \varnothing, \varnothing, \ldots$
s'_{22}: $\{a', d', e'\}, \{b', c'\}, \varnothing, \varnothing, \ldots$
s'_{23}: $\{b', c', d'\}, \{a', e'\}, \varnothing, \varnothing, \ldots$
s'_{24}: $\{b', c', e'\}, \{a', d'\}, \varnothing, \varnothing, \ldots$
s'_{25}: $\{b', d', e'\}, \{a', c'\}, \varnothing, \varnothing, \ldots$
s'_{26}: $\{c', d', e'\}, \{a', b'\}, \varnothing, \varnothing, \ldots$
s'_{27}: $\{a', b', c', d'\}, \{a', b', c', d', e'\}, \{a', b', c', d', e'\}, \ldots$
s'_{28}: $\{a', b', c', e'\}, \{a', b', c', d', e'\}, \{a', b', c', d', e'\}, \ldots$
s'_{29}: $\{a', b', d', e'\}, \{a', b', c', d', e'\}, \{a', b', c', d', e'\}, \ldots$
s'_{30}: $\{a', c', d', e'\}, \{a', b', c', d', e'\}, \{a', b', c', d', e'\}, \ldots$
s'_{31}: $\{b', c', d', e'\}, \{a', b', c', d', e'\}, \{a', b', c', d', e'\}, \ldots$
s'_{32}: $\{a', b', c', d', e'\}, \{a', b', c', d', e'\}, \{a', b', c', d', e'\}, \ldots$

After four applications of the revision rule, all of the sequences stabilize. They do not stabilize to one hypothesis – after an initial changing of minds, it is either the case that nobody finds the party attractive or that everybody finds it attractive. There is no initial hypothesis that results in the members changing their minds infinitely. But again, no categorical claims about who finds the party attractive can be subtracted from the revision process. However, the fact that revision processes sometimes give categorical judgments, as in the case of definition (*), makes it possible to build upon them semantics for circular definitions and to give a calculus C^0 that is sound and for finite definitions complete.

6. Conclusion

So, even if we were to understand the acceptance of atomic sentences by a member of a group in terms of the decisions concerning the sentence made by the other members of the group, whose decisions in turn are determined by the decision of the member in question, arguably a very extreme case to be taken into account, there is a logically legitimate way out and this is the way shown to us by Gupta and Belnap. Since no categorical judgements can be based on revision processes for (1),[7] such procedures would not by themselves lead to positive decisions being made by members of the groups, but would not put them at risk of contradiction either.

Institute of Philosophy
Jagiellonian University
52 Grodzka St.
31-044 Krakow, Poland
uzkijani@cyf-kr.edu.pl

References

Gupta, A. (2000). On circular concepts. In A. Chapuis, & A. Gupta (Eds.). *Circularity Definition and Truth* (pp. 123–154). New Delhi: Indian Council of Philosophical Research.

Gupta, A. (2004). Finite circular definitions. In T. Bolander, & V. F. Hendericks, & S. A. Pedersen. *Self-Reference* (pp. 53–67). Stanford: CSLI Publications.

Gupta, A., & Belnap, N. (1993). *Revision Theory of Truth*. Cambridge, MA: MIT Press.

Kijania-Placek, K. (2000). *Prawda i konsensus. Logiczne podstawy konsensualnego kryterium prawdy* (Truth and consensus. A logical basis of the consensus criterion of truth). Krakow: Jagiellonian University Press.

[7] Because one of the reflexive hypotheses is always the empty set. Thanks to Wen-fang Wang for drawing my attention to the need to clarify this point.

Logicism and the Recursion Theorem*

Vojtěch Kolman

Old questions such as “Is logicism dead?” or “Are the truths of arithmetic synthetic a priori?“ are once again being revived by scholars like Crispin Wright or George Boolos. But we have not learnt to be more cautious and still tend to answer them rashly without taking into account the broader background against which they were originally posed. As a consequence we are facing the same insurmountable difficulties as Poincaré, Wittgenstein or Russell, no matter whether their (rash) answers were “yes” or “no”.

This paper has two aims. The first is to portray Frege’s logicism in the spirit of Lakatos’ logic of mathematical discovery as a bold conjecture eventually rejected. The second aim consists in showing that this rejection was based on different and more serious reasons than we are usually told. What I maintain is that one can agree with the neologicists that Frege’s system is not so badly affected by Russell’s paradox as was once thought, but still claim that this is not enough to render the whole project successful according to Frege’s own standards. So the second aim of my paper is to indicate why – contrary to the neologicists’ plan – the logicist idea cannot be saved *and* rendered compatible with Frege’s original intentions, which, in my opinion, are quite sound. In view of this, of course, it is necessary to outline to some extent the original intentions and standards of Frege.[1] The recursion theorem and its role in the early foundational development turn out to be central to both parts of my argument.

1.

The so-called recursion theorem (RT), first proved by Dedekind and later by Frege, pertains to the method of how functions on natural numbers can be uniquely defined, namely in the *usual* recursive way by (1) setting the value for 0 and (2) laying down the rules for computing the value at $n+1$ from the value at n. The theorem says that there is exactly one, i.e. one and only one, function

* Work on this paper has been supported in part by the grant No. 401/06/0387 of the Grant Agency of the Czech Republic and in part by the research project MSM 0021620839 of the Ministry of Education of the Czech Republic.

[1] For a detailed account see Kolman (2005).

of this kind, thus guaranteeing the correctness and meaningfulness of the definition by recursion. Just for the record, the exact wording of the theorem (in one of its versions) goes like this:

> If a is an element of a set S and g is a function from S to itself, there is exactly one function f from the natural numbers to S such that (1) $f(0) = a$, (2) $f(n+1) = g(f(n))$ for every natural number n.

From the usual or "intuitive" point of view, however, there is no need, and in fact no room for such a license, because recursion is "intuitively" held for the most simple and natural way of proving or defining in arithmetic.

This intuitive point of view, moreover, seems to be supported by the fact that, for a long time, the theorem itself was omitted from books on foundations even by mathematicians such as Peano or Landau who, furthermore, did not explicitly share the idea of induction being the essence of arithmetical reasoning (in other words: who did not adhere to the Kantian philosophy of mathematics, unlike, e.g. Poincaré). Of course, these mathematicians may simply have missed the need for the theorem, as Landau actually thought he had when he – in his own words – added the proof of it to his *Foundations of Analysis* only after an intervention from a colleague.[2] As a result, it is possible to argue that it was only because of Frege's and Dedekind's foundational work and their proofs of certain seemingly "intuitive" theorems that the standards of rigor and the techniques of proof were gradually improved and mathematicians finally became aware of their necessity, as happened ostentatiously to Landau.

One can, furthermore, link Dedekind's proof with the general distrust of arguments based on intuition. Among the 19th century mathematicians this became a relatively common attitude to Kant's philosophy of mathematics with its thesis that the roots of arithmetic and geometry ought to be sought in the spatio-temporal structures imposed on reality by reason. The official standpoint, then, was that instead of trying to (mentally) intuit something we should (verbally) prove it, as Bolzano allegedly did in the case of the intermediate value theorem (IVT) or actually in its special case, the so-called Bolzano theorem which is, however, equivalent to the first one:

> Let, for two reals a and b, $a<b$, a function f be continuous on a closed interval $[a,b]$ such that $f(a)$ and $f(b)$ are of opposite signs. Then there exists a number $x \in [a,b]$ with $f(x)=0$.

Let us consider the tempting standard parallel between Bolzano's and Dedekind's theorems:

[2] This observation is due to Michael Potter, see (Potter, 2000, p. 83).

We are usually told that Bolzano proved something self-evident on an analytic (purely conceptual or verbal) basis merely to avoid the spatial intuition behind it. Along these lines we can conclude that Frege and Dedekind tried to expel the temporal intuition from elementary arithmetic by replacing the recursive definition with sophisticated expressive and deductive tools. This actually does make sense because, according to Kant, the sequence of natural numbers is a result of counting in time, and this counting or construction in time is based on the routine recursive procedure: (1) produce the numeral 0, (2) provided n has already been constructed produce the numeral $n+1$. Despite the fact that Frege and Dedekind themselves would have happily approved of this parallel, I suggest discarding it as misleading and replacing it with another, less obvious but certainly deeper as far as its consequences go.

What I want to draw your attention to is that both Bolzano's and Dedekind's proofs are in fact not proofs in the usual sense whereby a proof amounts to the "choice" between two basic possibilities, namely a conjecture's being true or false, as the proof of Fermat's theorem is or the proof of Goldbach's conjecture one day will be. The situation with the IVT and RT is different. Similarly to many other famous "proofs", such as Cantor's argument for the nondenumerability of the reals or Brouwer's proof that every total real function is (locally) uniformly continuous, what such "proofs" consist in is rather something like a resolution to proceed in a very specific way which is only one among many ways not delimited in advance. Such resolutions (or, as Germans would say, prototheoretical justifications) can then underlie real proofs to be articulated later.

To illustrate the point, let us take a closer look at the IVT. There is no doubt that Bolzano's notorious example of an everywhere continuous but nowhere differentiable function can be read as the indication of certain troubles to which the explicit (not the intuitive) definitions of conceptualized analysis (such as those of continuity, convergence and derivative) can lead. But something very similar holds for the IVT as well. In accordance with the well-known "epsilon-delta-type" definition of continuity and somewhat surprisingly, the function

$$f(x) = \begin{cases} 1 & \text{if} \quad x^2<2 \vee x<0 \\ -1 & \text{if} \quad x^2\geq 2 \vee x\geq 0 \end{cases}$$

is continuous on the rational line and yet fails to meet the IVT there. A similar argument holds for other non-Cantorian continua like the Euclidian one (consisting of numbers constructed by means of ruler and compass), the Cartesian one (consisting of algebraic numbers) or that built on *lawlike* sequences of rational numbers which actually goes back to the Pythagorean definition of proportion by means of a reciprocal subtraction (*anthyphairesis*).[3]

Of course, we have not yet said what a continuum is, but then, neither did

[3] See Fowler (1999).

Bolzano (or Leibniz or Cauchy), so we might as well work merely on the basis of ancient standards where "being continuous" amounts to the simple opposite of "being finitely divisible" and is therefore met already by the totality of rational numbers.

Hence what I say is that since Bolzano did not possess a clear concept of real number, one cannot conclude that he tried to prove – or even proved – some self-evident truth by analytical means. He could not provide a definite "choice" between the truth or falsity of a conjecture since he had no clear concept of truth for arithmetical sentences at all. What we are authorized to claim is merely that Bolzano indicated that for the IVT to hold continuum must be complete in a very specific (order-complete) sense. So instead of a proof we are (at best) facing a decision to define real numbers in a certain holistic way.

In comparison with the IVT, the case of the RT seems to be even trickier since it requires the scrutiny of the concept of definition itself. According to Frege, i.e. from the logician's and logicist's point of view, the usual types of definition, like the definition by recursion, are too specific or unreliable (as Brouwer would say) and should therefore be replaced by, or reduced to, an explicit definition as the only admissible form. For the usual recursive form consisting of two basic steps this means that it has to be expressed by a single formula. The first success of conceptualization or "logification" of arithmetic along these lines was achieved by Frege in his *Begriffsschrift*. He came up with an explicit second-order definition of a closure licensing him to capture natural numbers as the smallest set containing number 0 and closed under a one-to-one successor function s which does not assign 0 to any of its arguments. Dedekind achieved the same result putting it in the more familiar form of the so-called Peano's axioms (PA2):

(1) $(\forall x,y)(s(x)=s(y) \rightarrow x=y)$

(2) $(\forall x)(0 \neq s(x))$

(3) $(\forall F)(F(0) \wedge (\forall x)(F(x) \rightarrow F(s(x))) \rightarrow (\forall y)F(y))$

Just notice that because of the third axiom being second-order this version of arithmetic, contrary to its first order variant (PA1), can actually be conceived as a single axiom.

2.

This initial success encouraged Frege to claim that logicism is a feasible *hypothesis* according to which we can expect that (a) numbers will be conceptually separated by means of a single predicate from the domain of all objects or at least from the domain of all "logical objects" (whatever they may be), (b) arithmetical functions will be delimited in a similar way, whereas their "intuitive"

– recursive – formation will first have to be shown to be logistically admissible. This is the task for the RT, as, by the way, Frege's polemic with Grassmann in his *Grundlagen* clearly indicates.

Indeed, thanks to the RT we can dispense with the usual four axioms for addition and multiplication as used in PA1 and, to the same effect, we can introduce the basic arithmetical operations *via* explicit definition

$$x + y = z \quad \text{iff} \quad (\forall f)((\forall x)(f(x,0)=x) \wedge (\forall x,y)(f(x,s(y))=s(f(x,y))) \rightarrow f(x,y)=z),$$

in the same way Frege introduced the concept of natural number

$$Z(y) \quad \text{iff} \quad (\forall F)(F(0) \wedge (\forall x)(F(x)\rightarrow F(s(x))) \rightarrow F(y)).$$

All of this comprises (1) the *expressive*, *semantic* part of Frege's logicist project. At the very beginning Frege supplemented it with (2) a *deductive, inferential* pendant, according to which: All arithmetical propositions are to be derived from logical axioms, the conceptual truths of Frege's new logic, by logical rules alone. One of these axioms, the so called G*rundgesetz V* – Frege's axiom of extensionality – provides the ontological basis from which the numbers are to be taken out. This ties both parts of the logicist project – the deductive and the expressive one – together. Let me inspect more closely how and why.

It is not difficult to see that conceptual separation, the process of picking out some objects as falling under a given concept while neglecting others, necessarily presupposes *two* ingredients: (a) the separating concept on the one hand, and (b) the matter or domain from which the objects are to be separated on the other. I call them the (a) *descriptive* and (b) *ontological* ingredients of the semantic part respectively.

It is obvious that the ontological basis we are looking for must be described by logistically acceptable means. The explicit definition, however, cannot be included since it presupposes its *definiens* as already given. What are the options now? Frege chooses to enlarge the logical vocabulary by introducing a second-order term-forming operator $\{x:Fx\}$ the meaning of which he wishes to fix contextually through *Grundgesetz V* (GV):

$$\{x:Fx\}=\{x:Gx\} \leftrightarrow (\forall x)(Fx\leftrightarrow Gx).$$

In Frege's opinion, this stipulation should supply us with objects utterly independent of non-logical, descriptive vocabulary, or *logical objects*, as he calls them. He does not say it explicitly but the implicit practice in his *Grundgesetze* shows that these are the well-known *pure sets* like the empty set $\{x:x\neq x\}$, the singleton of the empty set $\{x:x=\{x:x\neq x\}\}$, and the like.

In this sense, both Frege and Cantor proposed a set-theoretical solution of the foundational problems of arithmetic. But Frege, *unlike* Cantor, realized that now he has to face a new problem. Instead of

"What are numbers and how are they given to us?",

which was the key issue of his *Grundlagen*, the key question of his *Grundgesetze* is:

"What are sets and how are they given to us?".

Frege's general answer, however, is always the same: Numbers, sets or objects in general are something one may be acquainted with only within the context of a proposition. Unfortunately, in the case of sets this proposition – the GV – turned out to be contradictory, bringing Russell's paradox in its train.

The one thing we must grant to the neologicists is the proven fact that, contrary to common wisdom, Russell's paradox does not constitute a serious threat to any of the aforementioned parts of the logicist project. GV can be replaced by a similar principle, referred to in the literature as Hume's Principle (HP),

Card(F)=Card(G) iff F eq G,

whereby two concepts have the same or cardinal number if they are equinumerous (eq), which means that there is a one-to-one correspondence between their extensions. Moreover, Burgess and others (Hodes, Hazen, Boolos) showed with the help of very simple analytical model that unlike GV this new principle is deductively consistent with the underlying logic.

As a consequence, HP is inferentially weaker than GV, yet, as already Frege has shown, it is strong enough to entail all the axioms of PA2. That is why this result is known as *Frege's Theorem* and the L2 together with HP as *Frege Arithmetic*, both due to Boolos. Moreover, by means of RT one can prove the categoricity of the systems in question, as Dedekind explicitly and Frege implicitly did. The semantical completeness of these systems easily follows, which means that Frege and Peano Arithmetic entail every true arithmetical formula and nothing else.

Despite the fact that the concepts of derivation and entailment were by no means clearly established at Frege's or Dedekind's time, these results seem to make the idea of logicism being vindicated quite plausible. In the remainder of my paper I shall show why it is plausible only superficially.

3.

The first reason is that the whole story has a dramatic sequel, namely the appearance of incompleteness phenomena. According to Gödel's results, not only is arithmetic incomplete in the axiomatic-deductive sense, but the logic it should be prospectively reduced to is deductively incomplete too. The reasoning goes as follows: As we already remarked, PA2 consists of a single axiom, hence if A is an arithmetical formula, then so is the implication with A in the consequent and this axiom in the antecedent. But PA2 is categorical, so if A is the truth of arithmetic, this implication becomes the truth of logic and *vice versa*. From this the incompleteness of the underlying logic easily follows, or to be accurate: it follows that it cannot be weakly complete. (The strong incompleteness is cheaper.)

The fact that we usually do not phrase Gödel's result in this way is motivated by a tacit effort to keep the logic-in-question at least semi-decidable, which is also the real reason lying behind the first- and higher-order distinction and the current paradigm of first-order theories. But since there is no theoretical reason for granting semi-decidability any special status with respect to the logicist project one has to conclude that Gödel's discovery ruined, in effect, its inferential part. As far as the expressive part of the project is concerned our reasoning must be more subtle.

Let us deal with its ontological ingredient first. What we need is an infinite stock of objects serving as the basis for the subsequent conceptual separation. Moreover, this basis should be given in a way independent of any recursive formation. Neologicists try to achieve this by an indirect route adopting Hilbert's standpoint: we cannot say what "point", "line" or "number" are, but only describe their structural properties by means of an axiomatic system. This is the idea of the so-called implicit definition, which, in fact, is already present in Dedekind's logicist account.

Instead of separating numbers from the domain of all objects or logical objects by some *predicate* (like Frege) Dedekind attempted to separate them from the realm of all domains as the unique domain satisfying some *formula*, assuming that there is no need and in fact no way of describing all these possible domains in advance. Unlike Hilbert, Dedekind was fully aware of the fact that his axioms cannot "define" a unique system by themselves, since by definition the set of formulae has always plenty of systems fulfilling them, on condition that they have at least one! But showing that there is such a system that fulfils the axioms of PA2 reduces, as Dedekind quickly realized, to showing again that there is some infinite domain of objects. Hence, we are going round in circles.

To sum up this part: According to the neologicists' view HP seems to provide an infinite domain independent of any recursive formation, but we already know that the reverse is the case. Their HP does not entail but rather presupposes the existence of an infinite domain in order to be consistent.

It is hardly an accident that Boolos' proof of the consistency of HP builds on the model-theoretic-construction consisting of natural numbers, because obviously (1) *natural numbers* constitute the most prominent prototype of an infinite set (2) built up by a simple *recursive process*. The cumbersome examples of an infinite set due to Bolzano and Dedekind:

> proposition A,
> proposition that A is true,
> proposition that the proposition that A is true is true,
> etc.

as much as Frege's, Zermelo's and von Neumann's definition of cardinal and ordinal numbers

$$\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\{\{\varnothing\}\}\}, \text{etc.}$$
$$\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}, \text{etc.}$$

show very graphically that there is no way of circumventing this direct recursive construction by inventing a larger domain of which they would be species because this domain has to be specified recursively too. In the light of this, we can easily accept Kant's view that what the word "infinity" stands for is only a form of recursion.

Having accepted that the ontological part of the project is not feasible, we can still hope that the descriptive part remains sound; i.e. that provided somebody gives us the individual numbers, we are able to capture them by means of Frege's number-predicate

$$(\forall F)(F(0) \wedge (\forall x)(F(x) \rightarrow Fs(x)) \rightarrow F(y)).$$

as a single whole, i.e. as the set to which all and only successors of 0 belong. This is actually no trifling matter since it cannot be done *via* a first order formula or even formulae. Taking a closer look at the arguments for the soundness of the aforementioned number-predicate we find out that they stand or fall with the supposition that its second-order variable ranges over *arbitrary* subsets of the underlying universe including the set of all natural numbers. Hence, even having left aside the problem of impredicativity, we are facing a clear vicious circle, because the set of natural numbers is something we wanted to capture.

To put this last argument into perspective, let us take this infinite formula

$$x=1 \vee x=2 \vee \text{etc.}$$

or, in the general case of an arbitrary number set, a formula like this

$$x=1 \vee x=45 \vee x=89 \vee \text{etc.}$$

As far as the intended use is concerned they are sound enough, i.e., they capture the respective sets adequately. But of course they are not formulas in the same sense that second-order predicates (open formulas) are, because they are not finite sequences of characters, and this was actually the only reason Frege employed the second-order expressions in his project.

But what our aforementioned argument pointed at was that the finiteness of these expressions is only apparent because, if they are to work correctly, their variables are to be interpreted as appealing to an arbitrary set. But what else is the arbitrary set but an infinite sequence of numbers or an infinite formula without a generating rule (the meaning of "and so on") which one can actually follow and which therefore must be finite. Hence, from the logicist's and even logician's point of view both – infinite and second-order – predicates must be equally (un)acceptable as long as the aim of writing everything down is to avoid an appeal to somebody's intuition as to what an arbitrary set is, or to what the words "and so on" can mean without knowing "how on".

4.

So much for the argument, now comes the conclusion. It was the very decision to define arithmetical objects in the exclusively explicit way which, from the very outset, doomed Frege's foundational program to failure even more decisively than Russell's paradox could. This is because the paradox affects only the project's formal part (which later turned out to be reparable), whereas the logicists (and the so-called neologicists as well) are inevitably forced to employ recursive formations, and not only within the basic formula- or proof-building operations, but also within the justifications of their (second-order) explicit definitions (like that of a closure).

To sum up: Frege's attempt to state or prove something like RT amounts to a decision to perform arithmetic in a certain, very abstract way. In this project, the recursive formations are not conceived as *names* of arithmetical objects, but as their *definite descriptions*, which ought to be checked additionally as for their ability to represent uniquely. This plan turned out to be infeasible, at least in its entirety: a recursion is apparently the simplest way of to constitute or *name* things in arithmetic. It seems probable that Frege eventually realized this and therefore gave up the whole project.

The mainstream of the subsequent foundational movement (which today's neologicism attempts to develop) is a lame compromise between the formalism of the first-order syntax and the Platonism of model or set-theoretical semantics which does not meet the ambitions of Frege's original plan. Ironically enough, it was the arch-enemy of the verbalized mathematics, the intuitionism of Brouwer, which alone picked up the baton of Frege's basic approach to mathematics, resurrecting as meaningful and non-trivial the seemingly straightforward prescientific questions like "What does it mean for an arithmetical sentence to be true?" and "What are the natural numbers (the so-called standard model of formalized arithmetic) for?".

Vojtěch Kolman
Department of Logic,
Faculty of Philosophy and Arts
Charles University Prague
Czech Republic
vojtech.kolman@ff.cuni.cz

References

Boolos, G. (1998). *Logic, Logic, and Logic.* Cambridge, MA: Harvard University Press.
Fowler, D. (1999). *The Mathematics of Plato's Academy*. Oxford: Oxford University Press.
Kolman, V. (2005). Lässt sich der Logizismus retten? *Allgemeine Zeitschrift für Philosophie, 30* (2), 159–174.
Potter, M. (2000). *Reason's Nearest Kin*. Oxford: Oxford University Press.
Shapiro, S. (1991). *Foundations without Foundationalism. A Case for Second-order Logic*. Oxford: Oxford University Press.
Stekeler-Weithofer, P. (1986). *Grundprobleme der Logik.* Berlin: de Gruyter.

A Tale of two Schemata: Tarskian (Finitary) Truth and Ramseyan Mental States

Arnold Koslow

> "Our task, then, is to elucidate the terms true and false as applied to mental states, and as typical of the states with which we are concerned we may take for the moment beliefs."[1]

1.

I want to connect up into a coherent theory three primitive ideas which no one in their right mind would normally want to consider together. The first is this: In an (1922), unpublished (seriously unpublished) typescript on Truth and Simplicity, Frank Ramsey speculated on the idea that truth is an *incomplete symbol* and that the claim that "'p' is true" can be expressed in certain linguistic contexts by adding on the phrase "and p" in that context. As we shall see, that can't be right, but, as we shall also see, that insight is part of a more general account of truth for finite languages introduced by Tarski. The second is also an insight of Ramsey about belief, which he called a "truism". It can be expressed by saying that "Richard believes that p" is true if and only if "Richard believes that p", and p (we use "p" as a schematic letter). We shall see, that can't be right either, but it will become part of an account that we shall give that might be right. The third idea of Ramsey, was a central idea in an unpublished four chapter manuscript on logic; a late writing. The proposal was to organize an account of logic in terms of the truth and falsity of *belief states* rather than sentences, statements, or propositions. It was a very bold idea for that time and ours as well. That account never emerged, and it is these three ideas that I will try to combine into a coherent, simple theory.

Ramsey, in a remarkable passage of his unpublished *Facts and Propositions,* (1927), proposed a thesis about belief according to which (**RBT**) any belief that p, is true if and only if p. Here is that prescient passage:

> "It is, perhaps also immediately obvious that if we have analyzed judgment we have solved the problem of truth; for taking the mental fac-

[1] "The Nature of Truth" in Rescher and Majer (1991, Chapter 1, p. 7).

> tor in a judgment (which is often itself called a judgment), the truth or falsity of this depends only on what proposition it is that is judged, and what we have to explain is the meaning of saying that the judgment is a judgment that *a* has *R* to *b*, i.e. is true if *aRb*, false if not. ... In order to proceed further, we must now consider the mental factors in a belief."[2]

This doesn't seem to be right when one thinks of it as a thesis about beliefs. However, if we adjust Ramsey's proposal so that it is a requirement on belief states , then it yields what I shall call *Ramsey's Belief Schema,* (**RBS**).[3]

(**RBS**). A belief state that p, is a true belief state, if and only if p.

Given our account of the truth and falsity of belief states, we will be able to prove this schema and it will be evident that it is a parallel that it is a close cousin of the Tarski T-schema. We turn first to a consideration of Tarski's definition of truth for finite sets of sentences

2.

Astute scholars of Ramsey and Tarski have noted that there is some connection, perhaps even an anticipation of Tarski's T-Schema by Ramsey's endorsement of what is sometimes called the *Redundancy Theory of Truth*. Truth be told, something akin to the T-schema goes back to Aristotle, resonates to Frege, and was endorsed by the early twentieth century Cambridge philosopher W. E. Johnson as a redundancy claim that he expressed in a form that is tied to assertions:

> "the assertion of p is equivalent to the assertion that p is true ..."[4]

This is close to Ramsey's formulation but of course is not to be identified with the T-condition as formulated by Tarski.

There is however, another claim about truth and belief that Ramsey regarded as a "truism", which in its own way appeals to another kind of redundancy. He describes the idea this way:

(**BTp**) A belief is true if it is a belief that p, and p

---

[2] "Facts and Propositions" in Mellor (1990, p. 39–40).

[3] (**RBS**) will be expressed below by using schematic letters rather than propositional variables, and it is, as we shall see, analogous to the Tarski T-schema. This is a departure from Ramsey's way of expressing the condition, but not of any significance for the present discussion.

[4] This history is succinctly and aptly documented in Sahlin (1990, Chapter 2).

We know from the published and unpublished papers of Ramsey that although he termed it a "truism", it was nevertheless a central element of his thought about judgements, assertions or beliefs.[5] It is also the sort of conjunctive construction that Ramsey made much of in an earlier unpublished talk which he gave to the Apostles in 1922.[6]

It is obvious that there are several important differences that should be noted. First, that the T-schema as it figured in Tarski's work, and the variants of it that figure in the earlier accounts from Frege to W.E. Johnson are very different from each other, and second, that Ramsey's version of the T-schema and his other leading dictum about true belief (**Btp**) are important but patently different. The difference is that Ramsey's version of the T-schema leans heavily on epistemic notion of assertion or belief, while Tarski's version involves sentences, and is well known for its eschewal of any reference to assertion, judgement, or belief. In light of these obvious differences, it would be plausible to conclude that there is no connection whatever between Ramsey's version of the T-schema and his doxastic truism, and even less connection between Tarski's version of the T-schema and that truism. Nevertheless, I want to argue that there is a way of showing how two basic ideas: Ramsey's idea that the truth of p can be regarded in certain contexts as the conjunction of "and p" to that context, and his endorsement of the doxastic truism, can be brought into a very simple framework when they are both placed against the background of the Tarski T-schema in a *finitary* context. Consequently, when we look at that famous paragraph in "Facts and Propositions" where Ramsey says that

> "It is, perhaps also immediately obvious that if we have analyzed judgment we have solved the problem of truth; for taking the mental factor in a judgment (which is often itself called a judgment), the truth or falsity of this depends only on what proposition it is that is judged, and what we have to explain is the meaning of saying that the judgment is a judgment that *a* has *R* to *b*, i.e. is true if *aRb*, false if not. ... In order to proceed further, we must now consider the mental factors in a belief."[7],

we shall see how so much of it falls easily into place –including an argument for the thesis that if there is a belief state that p, then it is a true belief state if and only if p –which we shall call *Ramsey's Belief Thesis* (**RBT**). [8]

[5] This point is deftly made by U. Majer (1991).

[6] Ramsey proposed as early as 1922, that "p is true" is conveyed in certain linguistic contexts by using "and p" as a suffix in those contexts. Consequently he referred to "truth" as an incomplete symbol, similar to Russell's definite descriptions and classes in the *Principia.* Cf. A. Koslow (2005).

[7] "Facts and Propositions" in Mellor (1990, p. 39–40).

[8] (**RBT**) will be expressed below by using schematic letters rather than propositional variables, and it is, as we shall see, analogous to the Tarski T-schema. This schematic version is a departure from Ramsey's way of expressing the condition, but not of any significance for the present discussion.

3.

FiniteTarski. It is best to begin the discussion with the problem of defining "true" or better, the predicate "is true" in the case where we consider the case of defining such a predicate for a specified finite number of sentences. Tarski showed how this could be done using an appropriate metalanguage, without need for a notion of satisfaction. Before we consider his construction, it is instructive to look at some very simple special cases

Let's begin with the simplest case: to define a truth predicate for the sentence A, such that it is provable that $Tr(|A|) \leftrightarrow A$, where any sentence flanked by two vertical lines will count as a name of that sentence. The definition is straightforward:

$$Tr(x) : (x = |A|) \wedge A.$$

To provide a proof that $Tr(|A|) \leftrightarrow A$, we assume that in the metalanguage we have $|A|$, and the axiom $|A| = |A|$. It is trivial to prove the equivalence of A and Tr(A). First note that $Tr(|A|)$ is just $(|A| = |A|) \wedge A$., which implies A. For the converse, note that from A together with the axiom that $|A| = |A|$, we have their conjunction $(|A| = |A|) \wedge A$, which, by definition, is just $Tr(|A|)$. It was assumed in this argument, and in the ones to come that enough classical sentential logic is at hand to run the argument.

So for any single sentence we can define a predicate that is tailor-made to insure that it is a provable case of the Tarski T-schema. It is also obvious that for any specific sentence B, other than A, the sentence "$Tr(|B|)$" is provably incorrect, provided that for the sentence B, we have the sentence $|B| \neq |A|$ as an axiom of the metalanguage. For then $Tr(|B|)$ is the conjunction $(|B|) = |A|)$ $\wedge$ A, and the first conjunct contradicts one of the axioms of the metalanguage.

This construction, of a truth predicate, sentence-by-sentence, is not to everyone's liking. True we have a schema such that for every A, $Tr(|A|) \leftrightarrow A$ is provable in the metalanguage, but it's always a different predicate that is indexed to each sentence. It would be more precise to say that for every sentence A, the following instance of a truth-schema can be proved for a truth predicate that is indexed to A. i.e.

$$Tr_A(|A|) \leftrightarrow A.$$

The special feature of this schema is that it is doubly schematic. The "A" is a schematic letter, but the predicate is schematic too: there's a *different* predicate "Tr_A" for each replacement of the schematic A. Now that may be a cause for complaint. It is normally assumed that the T-schema $Tr_A(|A|) \leftrightarrow A$ always uses the same predicate, no matter what particular sentence replaces the schematic letter. One normally assumes that the sentences "'A' is true" and "'B' is true" use the same predicate. It is a familiar constraint on the various instances of the T-schema that is so common that it usually goes without saying. The doubly schematic version of the T-schema is highly relativized, but it is a coherent way of considering the relation between a statement A and the statement that A is true. It's just not the usual thing.

There's more dissatisfaction with this notion of truth that is indexed to single sentences: There seems to be no systematic relation between any of these truth predicates with each other, nor do any of them behave well with respect to logical operations. For example the truth of a conjunction does not imply the truth of its conjuncts, if truth is given by any of these statement-indexed predicates.

Many of these defects can be remedied by using a construction which Tarski employed in order to show, as he said, that "Under certain special assumptions the construction of a general definition of truth is easy". He considered a finite fragment of English, or some object language with finitely many sentences.[9] We shall use a slight modification which singles out some finite set of sentences $\Gamma_n = \{A_1, ..., A_n\}$ either of a fragment of English or of some object language. The idea is that instead of defining a truth predicate for sentence A, and another for sentence B, that we define one truth predicate for both of them. In all the finite cases, that can be done disjunctively, as Tarski noted.

It is worth considering a few simple examples before describing his construction that covers the case of any finite set of sentences:[10]

(1) Let $\Gamma_1 = \{A, B\}$. Then set $T_{\Gamma 1}(x) = [(x = |A|) \wedge A] \vee [(x = |B|) \wedge B]$. We assume that the metalanguage contains $|A| = |A|$, $|B| = |B|$, and $|A| \neq |B|$ as axioms. Then it follows that $T_{\Gamma 1}(A) \leftrightarrow A$, and $T_{\Gamma 1}(B) \leftrightarrow B$ are provable in the metalanguage.

It is obvious that the mini-example of (1) will not guarantee that the truth of $(A \wedge B)$ will be equivalent to the conjunction of the truth of A and the truth of B. However if the conjunction of A with B is included in the set containing A and B, and the metalanguage has the self identities of $|A|$, $|B|$, $|(A \wedge B)|$, and the non-identities $|A| \neq |B|$, $|A| \neq |(A \wedge B)|$, and $|B| \neq |(A \wedge B)|$ in the metalanguage, then the result is easily obtained:

(2) Let $\Gamma_2 = \{A, B, (A \wedge B)\}$. Then set $T_{\Gamma 2}(x) = [(x = |A|) \wedge A] \vee [(x = |B|) \wedge B] \vee [(x = |(A \wedge B)|) \wedge (A \wedge B)]$. It follows that (i) $T_{\Gamma 2}(|(A \wedge B)|) \leftrightarrow (A \wedge B)$, (ii) $T_{\Gamma 2}(|A|) \leftrightarrow A$, and (iii) $T_{\Gamma 2}(|B|) \leftrightarrow B$ are provable in the metalanguage. Consequently, $T_{\Gamma 2}(|(A \wedge B)|) \leftrightarrow T_{\Gamma 2}(|A|) \wedge T_{\Gamma 2}(|B|)$ is also provable.

---

[9] Tarski (1969, p. 63–77).

[10] Following Tarski, we shall also assume that none of the sentences has occurrences of the predicate "true". The aim after all is to define such a predicate.

As a final example we can consider how well this truth predicate behaves with respect to negation, if the negation of a sentence in included in the set of sentences for which the truth predicate is to be defined, and the metalanguage contains the self-identities of $|A|$, $|\neg A|$, and the non-identity $|A| \neq |\neg A|$.

(3) Suppose that $\Gamma_3 = \{A, \neg A\}$ and that $T_{\Gamma 3}(x) = [(x = |A|) \wedge A] \vee [(x = |\neg A|) \wedge \neg A]$. Thus (i) $T_{\Gamma 3}(|A|) \leftrightarrow A$, and (ii) $T_{\Gamma 3}(|\neg A|) \leftrightarrow \neg A$ are provable, and consequently, $T_{\Gamma 3}(|\neg A|) \leftrightarrow \neg T_{\Gamma 3}(|A|)$ is also provable. It follows from this that even in this simple case, $T_{\Gamma 3}(|A|) \vee T_{\Gamma 3}(|\neg A|)$ is provable (one of the conditions of adequacy that Tarski set for a definition of truth).

We can now proceed to Tarski's provision of a truth predicate for an arbitrary finite subset of the sentences of some object language, $\Gamma = \{A_1, A_2, \ldots, A_n\}$:

$$T_\Gamma(x) : [(x = |A_1|) \wedge A_1] \vee \ldots \vee [(x = |A_n|) \wedge A_n],$$

When the metalanguage contains as axioms the n identities $|A_1| = |A_1|, \ldots, |A_n| = |A_n|$, and all the non-identities $|A_i| \neq |A_j|$ (for all $i \neq j$ between 1 and n), it is easily proved that every instance, $T_\Gamma(|A_i|) \leftrightarrow A_i$ of the T-schema is provable in the metalanguage.

It is also interesting to note that if B is a consequence of the set $\Gamma = \{A_1, A_2, \ldots, A_n\}$ but not a member of it, then one can form the larger set $\Gamma^* = \{A_1, A_2, \ldots, A_n, B\}$, and for the truth predicate for the larger set, we will have preservation of the truth of a conclusion, given the truth of the premises: that is, if $A_1 \wedge A_2 \wedge \ldots \wedge A_n \rightarrow B$ is a theorem of the metalanguage, then $T_{\Gamma*}(A_1) \wedge T_{\Gamma*}(A_2) \wedge \ldots \wedge T_{\Gamma*}(A_n) \rightarrow T_{\Gamma*}(B)$ is also a theorem of the metalanguage (provided the usual assumptions of self identify and non-identity of the sentences of Γ^* are axioms of the metalanguage).

It is readily apparent that the truth predicates may differ with a difference of the sets for which they are defined. There is one fact which mitigates this feature, and it is that the truth predicate in the finite case is cumulative: If we suppose that Γ and Γ^* are two finite sets of sentences of the object language, such that one is a subset of the other, $\Gamma \subseteq \Gamma^*$, and "$T_\Gamma$" and "$T_\Gamma *$" are their respective truth predicates, then in shifting from one set of sentences to a second larger one, even though there is a corresponding shift in the truth predicate, nevertheless, everything that is true with respect to the first predicate will also be true with respect to the second predicate. In this sense, truth is *cummulative.* Roughly stated: no truth's are lost in expanding the set of sentences for which truth is being defined. The reason lies mainly with the disjunctive character of the definitions of the truth predicate in the finite case. For example if Γ^* has only one more sentence say "B", beyond the sentences of Γ, then $T_\Gamma(x)$ is $[(x = |A_1|) \wedge A_1] \vee \ldots \vee [(x = |A_n|) \wedge A_n]$, and $T_\Gamma*(x)$ is just $T_\Gamma(x) \vee [(x = |B|) \wedge B]$. So of course, it's provable that $[T_\Gamma(x) \rightarrow T_\Gamma*(x)]$.

When the sentences to be covered are infinitely many, and where the structure of some of them involves quantification, the Tarskian construction proceeds differently from the finite case, using a more detailed theory concerning subsentential parts and the satisfaction relation. Reference becomes important, and the theory required begins to look more substantial than what was needed for the finite case.

4.

True Belief States. With these various familiar results for the notion of truth in the finite cases in place, it is plausible that even in such a simple case, a viable notion of truth can be provided that has many of the features that are so familiar. The Tarski construction for the finite case provides an account of the predicate "x ix true" (T_Γ (x)) for finite sets of sentences Γ of an object language essentially by exploiting a device that conjoins "and A" to a context indexed to A ("x = | A | ") forming their conjunction ("[(x = | A |) $\wedge$ A], and then forming the disjunction of each such conjunction for each of the sentences of the set Γ. This part of the construction resonates with Ramsey's early "conjunctive" view (1922) that "true" is an incomplete symbol. Our task now is to take to heart Ramsey's proposal when, in his late study *The Nature of Truth,* he said that the

> " ... the task, then is to elucidate the terms true and false as applied to mental states, and as typical of the states with which we are concerned we may take for the moment beliefs."[11]

What we now wish to explain is how this task of elucidation of the truth or falsity of *mental states* might be accomplished by transferring or recreating in a parallel fashion, the Tarskian proposal for truth for finite sets of sentences to the nonlinguistic terrain of finite sets of mental states. One of the consequences of this elucidation is the proof for belief states of a counterpart of the T-schema for sentences, which we think is something that Ramsey anticipated.[12]

---

[11] N. Rescher and U. Majer (1991, p. 7).

[12] The focus on the problem of elucidating an account of the truth and falsity of mental states like belief states is explicit in the four unpublished chapters on logic probably written during 1927- 8 according to R.B.Braithwaite (1931, p. xiii –xiv). It indicates a radical departure of course from the usual assumption that the predicates of truth and falsity apply to the usual suspects: sentences, statements, or propositions. In fact this focus on states of belief rather than the usual targets is what makes possible the development of the present account. It is not an account which Ramsey proposed. I agree with the judgment on those chapters given by Braithwaite who noted that Ramsey was profoundly dissatisfied with them, and accordingly did not include them in the 1931 collection.

Let us suppose that there are some special belief states, the belief state that p, the belief state that q, the belief state that r, ...It is for these kinds of mental acts for which a definition of "true" is intended. They are belief states with a special kind of content indicated by the use of "p", "q", "r", ... Ramsey, in speaking of these mental states says that

> "... whether or not it is philosophically correct to say that they have propositions as objects, beliefs undoubtedly have a characteristic which I make bold to call *propositional reference*. A belief is necessarily a belief that something or other is so-and-so, for instance that the earth is flat; and it is this aspect of it, its being "that the earth is flat" that I propose to call its propositional reference."[13]

It is unimportant how this is decided, as long as they stand for the kinds of things that can enter into logical relations. Furthermore, just as in the Tarski construction, we needed some standard way to refer to the sentences under consideration, so too we shall use "tbtp", "tbtq", "tbtr", ... as standard names to refer to the belief states: the belief state that p, the belief state that q, the belief state that r, ...

When we shall consider sets of tbtps, I shall assume that there is some one individual[14] whose belief states we are considering. The individual needn't be a person, it could be one of Ramsey's wonderful creations – a chicken who believes that a certain sort of caterpillar is poisonous.[15]

We assume that for these mental states the belief state that p is identical with the belief state that p, and similarly for the belief state that q, and so on. That is, tbtp = tbtp, tbtq = tbtq, tbtr = tbtr, etc. Furthermore, if we have a set of belief states {tbtp, tbtq, ... , tbtr} we shall assume that they are all different: that is, tbtp ≠ tbtq, tbtp ≠ tbtr, tbtq ≠ tbtr, etc.

We shall not assume any special way in which the states of belief such as tbtp, are related to their propositional references. For example we shall see that

---

[13] "The Nature of Truth" in Rescher and Majer (1991, p. 7). The notion of propositional reference was explicitly described by Ramsey as primitive. However those beliefs that have propositional reference are those beliefs which are that something or other is a so-and-so. The simplest way to understand his view is that he is restricting his discussion to "belief that ..."; only he thinks that all beliefs are beliefs that. It should be said that there is some deliberate vagueness as well in Ramsey's use of the notion of "state", but he does not include sentences, statements or propositions as states. I have accordingly taken the notion of a belief state to be those belief states that something or other is a so-and-so.

[14] The extension of the present theory to finite sets of states that might involve several individuals is fairly straightforward.

[15] In Mellor (1990, p. 40). There's another but similar story that could be told about the caterpillar. But the chicken came first.

if tbtp is identical to tbtq, it will follow from our account of the truth and falsity of belief states, that p and q have to be equivalent. But the converse may be false. Our theory leaves that issue open, as we think it should.[16]

The assumptions in the Tarskian construction were not substantive. The corresponding assumptions for these belief states may be more controversial. The assumption of identities such as tbtp = tbtp require that the term tbtp refers to a mental state that exists. I also assume that although a person can be in many different belief states at one time, that there is only one belief state that p for a person at a time. As for the conditions of non-identity that are assumed, they too seem to come at some price. Parallel to the idea that if p and q are different, then so are their canonical names "|p|" and "|q|", we assume that if two belief states are different , then the canonical terms that refer to those states are different – i.e. tbtp ≠ tbtq.[17]

The theory so far, is silent about what the ps and qs may be. They could be sentences, statements, or propositions. We certainly don't want to say that they have to be truth bearers. That would doom Ramsey's project from the start. Ramsey aim was to give an account of truth and falsity of belief states such as the belief state that p (tbtp). To assume the notion of "truth bearer" would use the concept of truth for (say) propositions to elucidate the truth or falsity of certain belief states. That would run the analysis in precisely the wrong direction. Fortunately, for the limited purposes at hand, we do not have to resolve these issues.

We shall assume that we have a finite set of belief states $\Delta_n = \{tbtp_1, tbtp_2, \dots, tbtp_n\}$ for which it is assumed that $tbtp_i = tbtp_i$ (for $1 \leq i \leq n$), and $tbtp_i \neq tbtp_j$, for $p_i \neq p_j$ (for $1 \leq i,j \leq n$).

Let us begin with the simplest case – when the task is to define a truth predicate of belief states for a set of belief states that has only one member, say tbtp. Recall Ramsey's stress on the centrality of the idea that "... a belief is true if it is a belief that p, and p". We have to make an adjustment of this "truism" about beliefs (his description) to take into account that now the problem is to define truth and falsity for belief states. The natural expression of the insight for belief states is then given by:

$$\mathrm{Tr}\,(x) : (x = tbtp) \wedge p.$$

This tells us, when a set of belief states has only one member, say tbtp, then any

[16] That is, there could be p and q of the same truth value, and only slightly different in what they say, but the corresponding belief states might be identical. It's a kind of doxastic belief threshold phenomenon.

[17] It would be interesting to determine whether these states of belief should be taken as tokens or types. I would prefer types since I think a person can be in the same state twice, and different people can be in the same state. I don't think that the construction of a truth predicate for finite sets of states of beliefs forces the issue one way or the other.

belief state x is true just in case it is the belief state that p, and p. If the set of belief states Δ, has more than one member, then the truth predicate for that set of states is defined by the disjunction of clauses for the states of Δ, as if they were each considered as in the single case. More precisely, if $\Delta = \{tbtp, tbtq, ..., tbtr\}$, then the truth predicate for the set of states Δ is defined this way:

$$Tr_{\Delta}(x) : [(x = tbtp) \wedge p] \vee [(x = tbtq) \wedge q] \vee ... \vee [(x = tbtr) \wedge r],$$

where "$Tr_{\Delta}(x)$" says that x is a true belief state of Δ.[18]

Earlier we noted the passage in *Facts and Propositions* in which Ramsey expressed his belief that the truth or falsity of the mental factor of a belief (judgement) depends only on what proposition it is that is judged. Ramsey states what that dependence is for an illustrative special case: the belief that aRb is true, if aRb, and false if not.

If we make an adjustment in this passage to reflect that it is the truth and falsity of belief states that has to be elucidated, then Ramsey's observation is that the truth or falsity of the mental factor (e.g. belief state) depends only on what proposition it is that is believed. Transposed to belief states rather than beliefs, this would become: the belief state that aRb is true if aRb, and false if not. More generally (and schematically expressed), the belief state that p is true if and only if p. I shall call this ***Ramsey's Belief Schema***:

(RBS): For any belief state tbtp, $Tr(tbtp) \leftrightarrow p$.

I think that every instance of the Ramsey Belief-Schema is provable. Consider first the simplest case: the truth predicate for a set of belief states for which we seek a predicate "true" has only one member. In this very simple case, the Ramsey "truism" can be expressed as

(1) $Tr(x) : (x = tbtp) \wedge p$.

In this case then, the argument for (**RBS**) is simple: To show that $Tr(tbtp) \rightarrow p$, suppose that Tr(tbtp). Then by our definition, $(tbtp = tbtp) \wedge p$, and consequently, p. In short we have one half of the belief schema: It is surprising that the converse also holds: that is $p \rightarrow Tr(tbtp)$.

The proof goes this way: suppose that p. We have assumed that the tbtps under discussion all refer to belief states of an individual. In that case we have the identity statement (tbtp = tbtp) . So we have $(tbtp = tbtp) \wedge p$. Consequently, Tr(tbtp). Thus in the single case, we have proved that $Tr(tbtp) \leftrightarrow p$.

18 "is a true belief state" is ambiguous, between "is a true (belief state)" and is a (true belief) state". We mean the former of course – we want to pursue Ramsey's idea of ascribing truth or falsity to belief states.

The idea of defining a truth predicate for belief states, one-at-a-time, proves unsatisfactory for the very same reasons that defining a truth predicate for sentences, one at a time is unsatisfactory. The obvious thing to do is to abandon the one-at-a-time procedure, and define the truth predicate for arbitrarily large finite sets of belief states.

Let $\Delta_n = \{tbtp_1, tbtp_{2, \ldots,} tbtp_n\}$ be a finite set of belief states. The truth predicate for Δ_n (suppressing the notation for the set), is

(2) $Tr(x): [(x = tbtp_1) \wedge p_1] \vee [(x = tbtp_2 \wedge p_2] \vee \ldots \vee [(x = tbtp_n \wedge p_n]$.

In this case for each of the p_i in $\{p_1, p_{2, \ldots,} p_n\}$, we can prove that $Tr(tbtp_i) \leftrightarrow p_{i\,.}$

The proof is again straightforward. There are only n members of Δ_n, and it is assumed that all the $tbtp_i$ s are distinct from each other. Then for any p_k,

$$Tr(tbtp_k): [(tbtp_k = tbtp_1) \wedge p_1] \vee [(tbtp_k = tbtp_2) \wedge p_2] \vee \ldots \vee [tbtp_k = tbtp_n) \wedge p_n].$$

Since the negation of every disjunct other than $[(tbtp_k = tbtp_k) \wedge p_k]$, is provable, we conclude that $Tr(tbtp_k) \rightarrow [(tbtp_k = tbtp_k) \wedge p_k]$, is provable, and consequently, so too is $Tr(tbtp_k) \rightarrow p_k$. Conversely, assume that p_k. Since we have $tbtp_k = tbtp_k$, we have the conjunction $(tbtp_k = tbtp_k) \wedge p_k$. But this conjunction is a disjunct of $Tr(tbtp_k)$, and so we have $Tr(tbtp_k)$. That is, we have proved that $p_k \rightarrow Tr(tbtp_k)$. Consequently, $Tr(tbtp_k) \leftrightarrow p_k$.

In light of this result, the more exact description of (**RBS**) should be this:

(**RBS**)*: Let $\Delta_n = \{tbtp_1, tbtp_{2, \ldots,} tbtp_n\}$ be a finite set of belief states of an individual, and "Tr" be the truth predicate for Δ_n, then for any state $tbtp_i$ in Δ_n, $Tr(tbtp_i) \leftrightarrow p_i$ is provable.

The definition of the truth predicate for sets of belief states does not have any occurrences of "truth" in it, for the very same reason that Tarski's definition of truth doesn't. It therefore satisfies one of the desiderata of Ramsey's program for connecting belief states and truth.

The Ramsey belief schema for belief states is striking, and it lays to rest a possible suspicion that the shift to true belief states is just another way of referring to true beliefs. That is, suppose that tbtp (the belief state that p) is a state of belief of Oscar. One could suppose that that state is true if and only if Oscar believes that p. That is,

Tr(tbtp) is equivalent to "Oscar believes that p."

However this equivalence of the truth of Oscar's state of belief and Oscar's belief is incompatible with the Ramsey Belief Schema (**RBS**)*. "Tr(tbtp)" implies p, but "Oscar believes that p" doesn't. If Oscar's belief state that p is true, then it follows (by the Ramsey Belief Schema) that p, but p doesn't follow simply from the statement that Oscar believes it. This seems to me to be correct. If Oscar believes that p, then Oscar may very well be in the belief state that p. Even if he is, even if he believes that p is true, it doesn't follow that that belief state is true, nor that p is true. The shift to states of belief and their truth seems to yield more than simply adhering to the truth of belief statements.

5.

Some Possible Reservations. In the course of proving the various instances of the Ramsey Belief Schema we made several assumptions that might be thought to be incompatible with Ramsey's overall views, or might be thought to be just plain wrong. Here are several of interest.

(1) **Finitism**. It is worth noting that although the restriction of truth predicates to finite sets of sentences might seem too stringent to all but someone who denied the existence of any actual infinite, the restriction to finite sets of belief states may be acceptable. After all, the set of belief states can be as large as one wants or needs, and it is a bit of a stretch to insist that there is a real need to allow for beings with infinitely many beliefs states. Furthermore, there is some reason to believe that Ramsey in 1929 endorsed a finitist view which rejected the existence of any actual infinite.[19] Thus the restriction to finite sets would be entirely compatible with Ramsey's general philosophical commitments held in 1929.

(2) **One Truth Predicate of Sets of Belief States, or Many.** When the notion of truth is elucidated for arbitrarily large finite sets of belief states, it becomes clear that the truth predicates for different sets them will be different. This is also the case when truth predicates are elucidated for various finite sets of sentences. However in both cases the notion of truth is *cummulative*. If one set of belief states is a subset of a second, then all those belief states which are true with respect to the truth predicate of the first set will also be true with respect to the truth predicate for the second. That should ameliorate the systematic ambiguity that is involved.

[19] As reported by R. B. Braithwaite (1931, p. xii).

(3) **Conjunctive Closure and Negation Completeness**. According to another possible objection, the Ramsey Belief Schema has consequences that are controversial and may not be correct. Two examples:

(i) (**Conjunction**) $\mathrm{Tr}(\mathrm{tbtp}) \wedge \mathrm{Tr}(\mathrm{tbtq}) \leftrightarrow \mathrm{Tr}(\mathrm{tbt}[p \wedge q])$,

and

(ii) (**Negation**) $\mathrm{Tr}(\mathrm{tbtp}) \leftrightarrow \neg\mathrm{Tr}(\mathrm{tbt}[\neg p])$.

Consider (i) first. This condition on belief states looks like the familiar controversial conjunctive condition on beliefs: a person believes a conjunction if and only if they believe each of the conjuncts. However, this result is quite different. One has to be careful and not confuse the usual conjunctive rule for beliefs with the present result about the way the truths of the states tbtp, tbtq, and tbt[$p \wedge q$] are related. A proof of (i) might run this way: By (**RBS***), we have $\mathrm{Tr}(\mathrm{tbtp}) \leftrightarrow p$, $\mathrm{Tr}(\mathrm{tbtq}) \leftrightarrow q$, and $\mathrm{Tr}(\mathrm{tbt}[p \wedge q]) \leftrightarrow p \wedge q$. Therefore by classical sentential logic, we have $\mathrm{Tr}(\mathrm{tbtp}) \wedge \mathrm{Tr}(\mathrm{tbtq}) \leftrightarrow \mathrm{Tr}(\mathrm{tbt}[p \wedge q])$.

The proof is not faulty, but one has to pay careful attention to what it assumes. It is assumed that all three states of belief , tbtp, tbtq, and tpt [$p \wedge q$], are in some set Δ, and "Tr" is the truth predicate for Δ. The result holds if all three states are in the set. However, if tbtp and tbtq are in some set, but the belief state tbt[$p \wedge q$] is not, then (i) will not be provable.

In other words, it is not guaranteed by this theory that if someone is in the belief state tbtp, and also in the state tbtq, then they are also in the conjunctive state tbt[$p \wedge q$]. The theory so far leaves that open –as it should.

Condition (ii) on negation also appears to be controversial. In the case when the belief states are of some particular individual, it seems to imply that for any p, that the individual is either in the belief state that p, or else in the belief state that $\neg$p. One would have to be in either one belief state or the other. There's no room for a kind of indeterminacy, and that would be a drawback of the theory. However the present result is quite different. It says that of the two states of some individual, tbtp, and tbt[$\neg$p], one or the other of them is true: $\mathrm{Tr}(\mathrm{tbtp}) \vee \mathrm{Tr}(\mathrm{tbt}[\neg p])$.

The proof of (ii) is simple enough. Suppose that Δ is some set of belief states, and that for some p, both tbtp and tbt[$\neg$p] are in Δ. Then if "Tr" is the truth predicate for Δ, we have both $\mathrm{Tr}(\mathrm{tbtp}) \leftrightarrow p$, and $\mathrm{Tr}(\mathrm{tbt}[\neg p\ ]) \leftrightarrow \neg p$. Consequently $\mathrm{Tr}(\mathrm{tbtp}) \leftrightarrow \neg\mathrm{Tr}(\mathrm{tbt}[\neg p])$.[20] The proof is fine, but it does depend

[20] The assumptions under which (ii) and the claim that $\mathrm{Tr}(\mathrm{tbtp}) \vee \mathrm{Tr}(\mathrm{tbt}[\neg p])$ are provable also show that they are equivalent to $p \leftrightarrow \neg\neg\, p$, and $p \vee \neg p$ respectively. That isn't a problem, unless intuitionist scruples are at issue.

upon both states, tbtp and tbt[¬p] being members of the set for which the truth predicate is defined. If one of them is in the set but the other not, then (ii) will not be provable.

(4) **Incoherent States of Belief.** Some of the preceding observations may suggest some further misgivings about the theory we have been explaining. We have considered various sets of belief states and the truth predicates for them. Some of those sets might seem to be strange, perhaps incoherent sets of belief states, if they are supposed to be states of some one individual. The suggestion is that there needs to be further work to delimit or restrict the membership of collections of states, before we define truth predicates for them. For example, in the discussion of (ii) above, we considered certain sets of belief states that contained the two states of belief tbtp (the belief state that p) and tbt[¬p] (the belief state that ¬p). Could they both be states of belief of some individual? The theory developed thus far imposes no constraints on the belief states that might be collected into a set, and have a truth predicate for them.[21] Thus, for all that we have said so far, tbtp and tbt[¬p] could be states of one individual. That has not been ruled out. What has been ruled out, however, is that those two states cannot both be true states. The reason is simply that the conjunction of Tr(tbtp) and Tr(tbt[¬p]) is inconsistent since it implies $p \wedge \neg p$. One of the two states is not true –say it is tbt[¬p]. Although it is a false state of belief, it is nevertheless still a state of belief of the individual. It may be the understatement of the year, but not all states of belief of an individual have to be true.

(5) **Two Kinds of Truth Predicates (for "Sentences" and for States of Belief) or One?**

There is a temptation to think that despite our emphasis on the advantages of shifting to states of belief rather than belief sentences, there isn't much difference or advantage in defining truth predicates for finite sets of belief states. If that were so, then the point of trying to give an account of the truth of belief states that did not rely on an account of truth for sentences, statements, or even propositions, would be pointless. The reason involves a mistaken inference from the two kinds of schemata. Assume that $\Delta = \{p, q, ..., r\}$ is some finite set of "sentences", and that $\Delta^* = \{tbtp, tbtq, ... , tbtr\}$ is the corresponding finite set of states of belief. This is just an assumption. We do not assume that for every set of type Δ, there will always exist a corresponding Δ^*.

If we form the corresponding truth predicates for the two sets, we will have proofs of the following instances of each schema:

[21] In this way, the treatment for finite sets of states of belief and finite sets of sentences are on a par.

(i) Tr*(tbtp) $\leftrightarrow$ p,

and

(ii) Tr(| p |) $\leftrightarrow$ p.

From the two schemata we could infer the equivalence

(iii) For the sets Δ and Δ^*, Tr*(tbtp) $\leftrightarrow$ Tr(p).

This however does not entail that the two truth predicates are the same; they are not even coextensional for one is defined on belief states, and the other is defined on "sentences".

There may be a deeper problem with (iii) that has to do with additional assumptions about belief states other than those we made in deriving the equivalence. Here is a made-up example of what we have in mind. Suppose that someone has a theory of belief states according to which belief states are maximal in this sense: For any disjunction, the belief state that (A $\vee$ B) is either identical to the belief state that A or it is identical to the belief state that B (i.e. tbt[A $\vee$ B] = tbtA, or tbt[A $\vee$ B] = tbtB). We might call such a view "dedicated belief state intuitionism"

We can now see that (iii) will fail for certain disjunctions: Let the disjunction (A $\vee$ B) and its disjuncts (neither of which implies the other), be in the set Δ, and the belief state tbt[A $\vee$ B] as well as the belief state that A, and the belief state that B, be in the corresponding set Δ^* of belief states. Suppose too that (iii),

Tr(A $\vee$ B) $\leftrightarrow$ Tr*[tbt(A $\vee$ B)],

where "Tr" is the truth predicate defined for the set of sentences Δ, and "Tr*" is the truth predicate defined for the set of corresponding belief states Δ^*. "Tr(A $\vee$ B)" is equivalent of course to "(A $\vee$ B)". Consider Tr*[tbt(A $\vee$ B)]. Either tbt(A $\vee$ B) is identical to tbtA (the first case) or it is identical to tbtB (the second case). In the first case, Tr*[tbt(A $\vee$ B)] is equivalent to Tr*[tbtA], which is equivalent to A. Consequently (A $\vee$ B) implies A, and so B implies A. But that is impossible. In the second case, Tr*[tbt(A $\vee$ B)] is equivalent to Tr*[tbtB] which is equivalent to B, so that (A $\vee$ B) implies B. Consequently A implies B. But that too is impossible. Therefore (iii) fails, and it fails because in this imagined example of a theory of belief states, there are no disjunctive states of belief where neither disjunct implies the other.

The moral of this imaginary extension of our simple theory of belief states is that one has to be careful in how a theory of belief states and their truth are developed beyond the elementary assumptions in our account. This is especially so, if the theory of belief states says, as our made up example does, that certain kinds of belief states do not exist.

6.

The Unfinished Theory and Pragmatism. The theory as presented thus far, yields certain results that were characteristic of Ramsey's thoughts about the relation of belief states and truth. By focusing on the elucidation of an account of the truth and falsity of mental states we were able to construct arguments for certain theses like the Ramsey Belief Schema* which Ramsey advocated. However, in our proof that the instances of the Ramsey Schema were provable, we used certain assumptions that in turn need some explanation. We assumed the existence of finite sets of belief states tbtp etc., since we assumed that they were self identical (tbtp =tbtp, etc.), and they were different from each other (tbtp ≠ tbtq, etc.). Just to fix our ideas a little more definitely, consider the truth predicate for a set of two belief states tbtp and tbtq of an individual. That truth predicate was given by $\mathrm{Tr}(x) : [(x = \mathrm{tbtp}) \wedge p] \vee [(x = \mathrm{tbtq}) \wedge q]$. Consequently if we want to determine whether $\mathrm{Tr}(\mathrm{tbtp}) \leftrightarrow p$, and $\mathrm{Tr}(\mathrm{tbtq}) \leftrightarrow q$ hold or not, we need information as to whether tbtp = tbtp, tbtq = tbtq, and tbtp ≠ tbtq. How are we supposed to obtain that information? Something should be said about when we attribute a belief state to an agent, and when the beliefs states that we attribute are different.

Ramsey had already addressed a similar problem in *Facts and Propositions*. The answer he thought lay in a so-called pragmatist view that was part of his story of the celebrated chicken and the caterpillar. From that story and a hint at the kind of pragmatism he had in mind, we can see one likely way in which the ascription of beliefs to individuals, no matter where they are in the pecking order, can be carried over with some adjustment, to the case of the attribution of belief states to them.

In order to focus on what role Ramsey assigned to the "pragmatist view", it is best to begin with two passages from "*Facts and Propositions*". The first is the celebrated passage on that famous chicken, and the second implicates Russell as the source of his pragmatism. Each is worth quoting in full:

> ".... It is, for instance, possible to say that a chicken believes a certain sort of caterpillar to be poisonous, and mean by that merely that it abstains from eating such caterpillars on account of unpleasant experiences connected with them. The mental factors in such a belief would be parts of the chicken's behaviour, which are somehow related to the objective factors, viz. the kind of caterpillar and poisonousness. An exact analysis of this relation would be very difficult. But it might well be held that in regard to this kind of belief the pragmatist view was correct, i.e. that the

relation between the actions were such as to be useful if, and only if, the caterpillars were actually poisonous. Thus any set of actions for whose utility *p* is a necessary and sufficient condition might be called a belief that *p,* and so would be true if *p*, i.e. if they are useful." [22]

Three things are worth noting about the first passage. It attributes a definite belief to the chicken: a certain sort of caterpillar is poisonous. Second, there is mention of the mental factor and the objective factors of that belief. We take the mental factor to be a reference to what he later called the belief state (tbtp), and the objective factors concern the kind of caterpillar and its poisonousness (p). Third there is the idea that a pragmatic view will provide an account of why that particular belief is properly attributed to the chicken. In the second passage, Ramsey is explicit about the Russellian origin of his pragmatism:

> "My pragmatism is derived from Mr. Russell; and is of course, very vague and undeveloped. The essence of pragmatism I take to be this, that the meaning of a sentence is to be defined by reference to the actions to which asserting it would lead, or, more vaguely still, by its possible causes and effects. Of this I feel certain, but of nothing more definite."[23]

Of course we cannot discount Russell's influence, but as we shall see, Ramsey's view seems closer to a Peircean brand of pragmatism.[24] It is worth a try to include caterpillars as having beliefs, but the Russellian suggestion to define the meaning of sentences by reference to actions that lead to their assertions is off the mark. However, the relevant actions needn't be limited to actions that lead to assertions, and Ramsey considers a more inclusive kind of action . He says that what we mean by attributing to the chicken a belief that those caterpillars are poisonous is that it avoids eating them on account of the untoward experiences that eating them would provide. The reference to certain actions of the chicken provides a way of attributing specific beliefs to the chicken. It is intended as a solution to the problem of belief ascription.

Even if this so-called pragmatic appeal to actions indicates whether the individual has that belief, it does not of course settle the question of whether that belief is true. It might be true that the chicken believed the caterpillar was poisonous, and just be wrong. Lucky caterpillar!

Moreover, the project is, as Ramsey said in his draft of "The Nature of Truth", to elucidate the truth (or falsity) of belief *states.* Consequently prag-

[22] Mellor (1990, p. 40).

[23] Mellor (1900 , p. 51).

[24] The case for the Peircean connection is clearly explained in Sahlin (1990, p. 70–73), together with an interesting relation of it to decision theory. This view of the matter is also supported by similar considerations in Dokic and Engel (2001, p. 22–25).

matist accounts of either the meaning or the truth of sentences, statements, or propositions would seem to be irrelevant. It is belief *states* whose truth needs discussion.[25] There is, nevertheless, a very appropriate place in this account, where Peircean views help to move things along, by saying something about the attribution of belief states, and the way in which one belief state might be distinguished from another. The idea is to continue to deploy a pragmatist view even if we shift from a discussion of sentences such as "Oscar believes that caterpillars are poisonous" to the attribution of a belief state (the belief state that caterpillars are poisonous) to Oscar.

There are some things that Peirce wrote, in "How to make our Ideas Clear"[26] that are helpful. One idea is that

> "The essence of belief is the establishment of a habit, and different beliefs are to be distinguished by the different modes of action to which they give rise" (pp. 129–130),

and

> "... what a thing means is simply what habits it involves. Now, the identity of a habit depends on how it might lead us to act, not merely under such circumstances as are likely to arise, but under such as might possibly occur, no matter how improbable they may be. ... Thus we come down to what is practical and tangible, as the root of every real distinction of thought, no matter how subtile it may be; and there is no distinction of meaning so fine as to consist in anything but a possible difference of practice." (p. 131)

The idea is then to transpose these suggestions to belief states, and propose that the states of belief (of Oscar, say) are connected with certain dispositions of Oscar to act under various circumstances, even improbable ones. Being specific about the kind of connection is of course the big problem. Rather than revert

---

[25] The issue of whether truth is a byproduct of an account of beliefs or the byproduct of an account of psychological mental states like belief states is not easily separated by citing those Ramsey papers in which beliefs are the target, or papers in which beliefs states are the target. In "Facts and Propositions" (1927), Mellor (1990), in the chicken passage, there is mention of both beliefs and their mental factors, and in the draft of "The Nature of Truth" (1927 – 28, or perhaps 29), Rescher and Majer (1991), the discussion is clearly intended to be about mental states, though there too he also talks about all beliefs as necessarily having a *propositional reference* ("that they are all beliefs that something is so-and-so", p. 7). It is possible that he could be also be understood, as also saying that all belief states are belief states that something is so and so, which is the way we understand him here.

[26] Hauser and Kloesel (1992).

to meanings, for the present let us just assume a weak Peirce-like proposal: the actions under various circumstances are *evidence* that Oscar is in a certain state of belief (that the caterpillar is poisonous).[27] So appeal to evidence of a practical, tangible sort would be evidence that the chicken was in a certain belief state. And evidence that one state of belief is different from another would be provided by the possible difference of practice associated with each. At root the difference, Peirce says, comes down to what is practical and tangible. That emphasis on the practical and tangible would indicate the utility these belief states have for the chicken, and us, and might be what lay behind Ramsey's use of the notions of utility and usefulness at the end of the caterpillar passage. The nice feature of Ramsey's appeal to pragmatic considerations is that it allows us to secure evidence for the claim that a certain state of belief is true, i.e. Tr(tbtp), without relying on some account of truth already in place. To my mind the present theory is coherent. With a little luck, it may even be the beginning of a theory that is correct.

Arnold Koslow
The Graduate Center
The City University of New York
365 Fifth Avenue, New York
akoslow@mindspring.com

References

Braithwaite, R. B., (Ed.) (1931). *The Foundations of Mathematics and other Logical Essays, by F. P. Ramsey*. London: Routledge.

Dokic, H., & Engel, P. (2001). *Frank Ramsey, Truth and Success*. London: Routledge.

Hause, N., & Kloesel, C. (Eds.) (1992). The Essential Peirce. *Selected Philosophical Writings* (vol. I, 1867–1893). Bloomington: Indiana University Press.

Koslow, A. (1990). Ramsey on Truth and Simplicity. In Nils-Eric Sahlin (Ed.). Ramsey's Ontology. *Metaphysica,* 3, 89–108.

[27] The usual connection by pragmatists is in terms of meanings. However we are concerned with belief states, and their connection to various acts is not a matter of meaning. Ramsey's appeal to causes would lose the connection with what is tangible and practical –i.e. to utility. The use of an evidential connection however, has its drawbacks as well.

Mellor, D. H. (Ed.) (1990). *F.P. Ramsey, Philosophical Papers*. Cambridge University Press.

Majer, U. (1991). Ramsey's Theory of Truth and the Truth of Theories: A Synthesis of Pragmatism and Intuitionism in Ramsey's Last Philosophy. *Theoria,* 57(3), 162–195.

Rescher, N., & Majer, U. (Eds.) (1991). On Truth. *Original Manuscript Materials (1927–1929) from the Ramsey Collection at the University of Pittsburgh*. Kluwer Academic Publishers.

Sahlin, N.-E. (1990). *The Philosophy of F.P.Ramsey.* Cambridge University Press.

Sahlin, N.-E. (Ed.) (2005). Ramsey's Ontology. *Metaphysica, 3*.

Tarski, A. (1969). Truth and Proof. *Scientific American*, *6,* 63–77. The revised version can be found in S. R. Givant and R. N. McKenzi (1986).

Givant, S. R., & McKenzi, R. N. (Eds.) (1986). *Tarski's Collected Works.* Basel: Birkh*ä*user.

Hause, N., & Kloesel, C. (Eds.) (1992). The Essential Peirce. *Selected Philosophical Writings* (vol. I, 1867–1893). Bloomington: Indiana University Press.

Pluralism and the Logical Basis of Metaphysics

Nils Kurbis

1. Introduction. Proof-Theory, negation and metaphysics

According to Dummett and Prawitz, the meanings of the logical constants may be given completely by their introduction and elimination rules in a system of natural deduction. Negation is the crucial constant when it comes to the question which the proof-theoretic justification of deduction has been purpose-built to decide: which of the two metaphysical positions realism and anti-realism is the correct one? Dummett reconstructs the realism/anti-realism debate as one about whether a certain logical principle holds: the principle of bivalence. Realism is equated with adopting classical logic, which keeps the principle, anti-realism with intuitionistic logic, which rejects it. The core idea is that the proof-theoretic justification of deduction enables us to solve the dispute from metaphysically as well as logically neutral grounds. It is independent of semantic assumptions, like the principle of bivalence, and thus independent of metaphysical assumptions, given the Dummettian reconstruction of the debate. Dummett argues that it is settled depending on which logic turns out to be the justified one: proof-theory is the logical basis of metaphysics. It is common knowledge that Dummett and Prawitz think that intuitionistic logic emerges as the proof-theoretically justified one and accordingly that anti-realism is the metaphysics to be favoured.[1]

I have argued elsewhere[2] that the definition of the meaning of intuitionistic negation given by Dummett and Prawitz is not workable, because the rule *ex*

[1] This gloss of the debate skirts the question whether the dispute is rather one about the verification transcendence of truth and whether there could be an anti-realist justification of classical logic. I take it, however, that at least at an initial stage – in Dummett's development of his ideas as well as in how he envisages the problem is to be tackled – this equation is the moving force behind the project, as guaranteeing metaphysical as well as epistemological neutrality. Anti-realism sets off using only intuitionist logic, as the logic emerging from the proof-theoretic justification of deduction; to establish that classical logic is anti-realistically acceptable arguments at a further stage in the development of a comprehensive theory would be called for.

[2] In my Ph.D. thesis and an extract of it 'Negation: A Problem for the Proof-Theoretic Justification of Deduction', currently in preparation for publication. This paper is also an extract from my thesis.

falso quodlibet does not guarantee that $\perp$ is always false. The symbol '$\sim$' defined in '$\sim A =_{def.} A \supset \perp$' is not negation, or if it is, then only because non-proof-theoretic considerations have implicitly been appealed to. The meaning of negation cannot be defined proof-theoretically, but rather has to be presupposed as given together with the meanings of the atomic sentences.

The purpose of this paper is to investigate into the repercussions of this result for the logical basis of metaphysics. Essentially, it means that the proof-theoretic justification of deduction does not provide for a way of deciding the issue between intuitionist and classical logicians. I shall argue that both logics have to count as unobjectionable from the perspective of proof-theory, as both the intuitionistic as well as the classical treatment of negation constitute legitimate ways of formalising and regimenting our informal, pre-theoretical concept of negation. Negation is underspecified in the sense that 'considered judgements of logicality' do not speak decisively for or against one or other option when we consider the cases which are at issue between classicists and intuitionists. This logical pluralism I argue for raises the question whether accepting two logics is at all coherent. I argue that it is. What needs to be given up however is the idea that proof-theory could be a logical basis for metaphysics.

2. The proof-theoretic justification of deduction should not be rejected

Before going into any details it might be worth reflecting why one shouldn't take the stance that, as the programme of the proof-theoretic justification of deduction has failed to meet its main objective – i.e. to decide between classical and intuitionistic logic –, it should be rejected as being a failed approach to the justification of deduction. This response should be particularly attractive to philosophers – the majority, I presume – who hold that intuitionistic and classical logic are in some sense 'rivals' for the title of the correct logic. In this light, the outcome that the proof-theoretic justification of deduction leaves us with (at least) *two*[3] acceptable logics rather than just one may be perceived as rather problematic. The reason why I should not recommend this way with the proof-theoretic justification of deduction is straightforward. There is much to be said in favour of Dummett's and Prawitz' programme. It is arguably the only workable systematic proposal for a justification of deduction. Semantic approaches presuppose a notion of truth and run the danger of circularity: the logical laws, like *tertium non datur*, that are to be established are implicitly assumed through properties of truth. So the choice is between living with a justification of de-

[3] In fact, I argue that there is a whole range of acceptable logics in addition to classical and intuitionistic logic which are unobjectionable from the proof-theoretic justification of deduction, in particular relevance logic and some of its relatives.

duction which fails to decide between classical and intuitionistic logic, and an approach which hardly deserves this name.[4] Proof-theory provides the most powerful method for justifying deduction ever proposed. If it fails to make a decision between classical and intuitionistic logic, then this is *prima facie* reason to accept them both as correct. Thus it is mandatory to investigate whether a logical pluralism is possible which accepts that classical and intuitionistic logic are equally good logics. To explore this is the purpose of this paper. But first, let's have a look at whether there might be some other way of deciding which logic to accepted. After all, if negation has to be presupposed as an undefined primitive in proof-theory then it might be thought that our previously given understanding of negation as used in ordinary discourse provides the means for deciding which negation rules to use, as it is this which informs our choice of them. I shall argue in the next section that this understanding is as indecisive when it comes to the question which of the two options for negation rules are the correct ones as is proof-theory. This consolidates the pluralist conclusion drawn earlier, as both, classical and intuitionistic negation rules may be backed up by reflection on the use of negation in ordinary discourse.

When in the following I talk about 'intuitions' and 'evidence' these are not to be understood as 'untutored', but rather as the basis of 'considered judgements of logicality' in the spirit of Mark Sainsbury and Michael Resnik: they are pre-theoretical logical insights on which formal logical theorising builds.[5] I shall call the negation of ordinary discourse 'informal negation', in contrast to its formal analysis as classical or intuitionistic negation. The aim of the next section is to argue that informal negation can intelligibly be used in either classical or intuitionistic fashion.

3. Indecisive intuitions

Intuitions concerning our pre-theoretical, informal concept of negation and its use would appear to open up a way of deciding which rules for formalised negation are the correct ones if it was possible to single out by means of them which set of rules matches them best. This however is unlikely to succeed if the choice is between classical and intuitionistic logic. Both logics agree in a large class of cases in their treatment of negation and these cases provide for the core of the use of negation in ordinary discourse, namely where sentences are used which may with some right be called *decidable*. These are the only cases where we can expect to have strong and decisive intuitions concerning the correct use of

---

[4] There are of course other approaches, but typically these are not *systematic* ones. *Cf.* the literature cited in footnote 5.

[5] Sainsbury (2002), Resnik (1996).

negation, but they are precisely cases on which a decision between classical and intuitionistic negation cannot be built. Classical and intuitionistic logic diverge only in circumstances quite arcane relative to common discourse, namely where undecidable sentences are used, e.g. involving quantification over an infinite domain. It is unlikely that evidence is forthcoming which could be strong enough to decide which logic to use here. There are no paradigm cases of discourse which could be cited to back up a claim that negation behaves classically or intuitionistically when the domain of quantification is infinitely large. Quite to the contrary, the mere fact that intuitionistic mathematics has been developed seems to speak for the thesis that there are two reasonable ways of treating negation in such cases. Thus for mathematics at least, no decision is forthcoming. Surely there are other regions of discourse where Dummettian realists and anti-realists disagree whether negation satisfies *tertium non datur* $A \vee \sim A$ in particular discourse about the future and subjunctive conditionals. To substantiate the claim that a decision between classical and intuitionistic negation based on evidence from reflecting on ordinary discourse is not possible in a wider class of cases either, let's have a closer look at two examples.

First, the future. Consider the statement that next week, I'll drink that bottle of Sancerre that's been sitting on my shelf for days now and that I haven't managed to drink yet. Are we to say that *tertium non datur* holds for 'I'll drink that bottle of Sancerre next week' or not?[6]

Pro

During the course of the week either I drink the bottle at some point or I don't. These two cases exhaust the possible options there are, *tertium non datur*. Thus either I'll drink the bottle or I won't. That *tertium non datur* holds for the future tense may be based on the fact that *tertium non datur* undoubtedly holds for the corresponding present tense sentence 'I drink that bottle' at some time during the course of next week, as at some point during the next week either it or its negation is bound to be true.

Contra

There is not yet a moment in time lying in the next week which would make either of the present tense sentences 'I drink the bottle' and 'I don't drink the bottle' true. Whether or not I drink it next week also depends on factors which are unpredictable now: other social events might come up which force me to give up my hopes that I'll drink it. We cannot base its truth on present intention that I'll drink it. Thus it is not determinate whether I drink the bottle or not and thus *tertium non datur* should be rejected.

---

[6] To see what is going to happen, we could just wait until the week has passed, so the example is reasonably far removed from cases of undecidable sentences of mathematics.

Both options constitute reasonable views on the behaviour of negation in future tense sentences, but neither argument is conclusive. Both views have their rationale in the light of the evidence. To accept *tertium non datur* for statements about the future focuses on the intuition that the two cases – either I drink the bottle or not – exhaust the possibilities and one of them has to materialise in the course of the next week. Rejecting *tertium non datur* is to do justice to the "openness" of the future. Both views focus on different *aspects* of informal negation, one might say. In the absence of a principled way of excluding one or other view, informal negation as used in discourse about the future has to count as underspecified.

Some philosophers may of course have views about the nature of the future that provide them with grounds for rejecting one or other option. Such a philosopher would have to show that the reasoning goes astray in one of the cases. But whatever reasons one could give to support such a claim, they would be of a rather different nature than the evidence appealed to above. They would be metaphysical reasons and thus we may exclude them from consideration, as the aim is to base metaphysics on logic rather than the other way round.

Similar considerations may be made in the case of counterfactuals. Suppose someone starts writing a Ph.D. and at some point during his course he drops out and takes to bee-keeping instead. Then we may ask ourselves whether the conditionalised instance of *tertium non datur* holds for 'Had he continued working on it, he either would have written an excellent Ph.D. or not.' is true.[7] Again we can give two lines reasoning. First, *pro*: writing an excellent Ph.D. or not doing so exhaust the possible options, *tertium non datur*. Hence either had he continued working on it, he would have written an excellent Ph.D. or he wouldn't. Secondly, *contra*: as in fact he hasn't continued working on it, there is no fact of the matter whether his Ph.D. would have been excellent or not had he continued working on it. Hence the conditionalised instance of *tertium non datur* should be rejected. There are more robust cases of counterfactuals where a conditional *tertium non datur* may be beyond reasonable doubt: for instance, had he completed his thesis and handed it in, then either he would have passed or he wouldn't (excluding unfortunate events that prevent preconditions for passing or failing to obtain). But one may doubt that all counterfactuals are of this kind, as the forgoing example shows. Thus counterfactuals provide further examples that show informal, pre-theoretical negation to be underspecified. I should argue that Dummett's example 'Jones was brave' is another case where informal, pre-theoretical negation does not decide whether or not *tertium*

---

[7] This may be considered to be a more genuine case of undecidability, as we do not have *scientia media*, but it still is notably different from mathematical examples: in the case of counterfactuals, there is not much we can do to decide what is the case, whereas in the case of undecided sentences of mathematics, at least we might hit on a proof one day.

non datur holds and so, I take it, is fictional discourse. But there's no space to go into any more details here.

This discussion underpins the unintended result of the proof-theoretic justification of deduction rehearsed in section 1 and provides independent support for the conclusion I draw from it. There is more than one option for formalising informal negation. Intuitionist and classical logic both have their rationale. Each logic captures different aspects of informal negation and focuses on different intuitions. Each regiments these aspects, but leaves out other aspects. If a metaphor may be allowed, formalising informal negation is like the straightening of a river: there are constraints on doing it properly, but there are several viable options of doing so, and you'll always leave some cut-off meanders. The discussion also gives independent support to the claim made in section 2 that the fact that the proof-theoretic justification of deduction fails to be decisive shouldn't lead us to reject it. That both classical and intuitionistic logic are proof-theoretically acceptable mirrors our pre-theoretical intuitions. Informal negation may thus be said to be underspecified relative to formalisation: it does not determine one of the two options of negation rules as the only correct ones. It is not determinate whether classical or intuitionistic principles should be applied. Informal negation is neither classical nor intuitionistic.

4. Is it incoherent to have two logics?

I have argued that we have two equally acceptable options of formalising negation. It might be objected that while it may very well be true that negation in natural language is neither quite classical nor quite intuitionistic, we'd better change this as it is questionable whether *both* logics could possibly be correct. In other words, it might be objected that this indecisiveness merely points to an inadequacy in our pre-theoretical, informal concept of negation. There is a simple argument employing reasoning acceptable to both, intuitionistic and classical logicians that purports to show that accepting two logics is inconsistent. If there are two logics, then it should be the case that there is a set of assumptions Γ and a conclusion A, such that according to one logic, A follows from Γ, but according to the other logic, it does not follow. But then A does and does not follow from Γ. Contradiction. So there cannot be two distinct correct logics. Assuming that there are some correct standards of logical reasoning, it follows that there can be only one correct logic.

Here is another problem one might find in pluralism. Assume all assumptions in Γ are accepted as true. Then either I am or I am not entitled to assert A, one is inclined to say, and the logic that tells me which is the case is the correct one. If there were two logics, in such a situation we would not know whether or not we can rely on the truth of A in our actions. Logic would fail to be a guide of thought. Pluralism is thus incoherent.

The first problem is a logical one, the second a pragmatic one. I'll discuss them in the next two sections and show that they are not really problems for pluralism.

a) Pluralism is not logically inconsistent

The logical argument against pluralism is easy to answer, strong and convincing as it looks at a first glance. A second glance shows that it is simply invalid. Although of course it is possible that a formula A follows from a set of formulas Γ according to classical logic, but not according to intuitionistic logic, no contradiction arises. It is true that, for certain Γ and A, $\Gamma \nvdash_I A$ and $\Gamma \vdash_C A$, were $\vdash_I$ and $\vdash_C$ are the intuitionistic and classical consequence relations. But this is as much a contradiction as the one between aRb and ~aSb. Thus no logical problem arises from accepting both logics as correct.

It might be objected that it nonetheless cannot be the case that both, classical and intuitionistic logic, are correct formalisations of our pre-theoretical notion of consequence, and thus although $\Gamma \nvdash_I A$ and $\Gamma \vdash_C A$ do not formally contradict each other, they cannot both correctly capture this notion. This objection misses the point that if negation is underspecified, so is the pre-theoretic notion of consequence. If there is more than one way of giving rules for negation, it follows that there is more than one way of capturing our pre-theoretic notion of consequence by logical consequence as determined by what counts as a deduction. If negation can be formalised in two different ways, the same counts for our pre-theoretical notion of consequence. Now in any case of well-formed formulae Γ, A where $\Gamma \nvdash_I A$ and $\Gamma \vdash_C A$, some of Γ, A must be undecidable. In other words, the cases where there is a real choice of logics are exactly those discussed in section 3, such that we have no grounds for favouring classical or intuitionistic logic. Hence this objection poses no further problem to what has already been discussed.

An opponent of logical pluralism might wish to strengthen her point: it may well be that no unique formal systems captures all our intuitions about consequence, and that there are two formal system which are equally adequate; nevertheless there *ought* to be only one logic, and hence we are under an obligation to make a decision which logic is the correct one and to declare some intuitions to be fallacious. If this course is taken we are back where we started: there is no reasonable means of making such a decision. Any decision would either beg the question – e.g. you chose the logic you assumed right from the start to be your favourite one – or it is based on grounds too feeble to support a choice as important as the choice of logic – e.g. you chose the one you've been trained to use in your undergraduate years of studying philosophy. A meaningful 'ought' should imply a 'can', should it not? Here we have a case where we can *not* do what allegedly we ought to do. If my arguments are correct, then the claim that there ought to be only one logic is pointless. There is no adequate rationale on which to base the decision which logic it would be.

I conclude that there is no logical problem with logical pluralism. So let's move on to the pragmatic problem.

b) Pluralism is not pragmatically incoherent

Here is again the pragmatic objection to logical pluralism. Given $\Gamma\nvdash_I A$ and $\Gamma\vdash_C A$ and you accept the premises Γ, should you go on to assert A or not: which logic are you to apply? If there are two logics, then you seem to have a choice, but we are inclined to say that it is not upon us to make a decision. Given all premises Γ are true, A either is or it isn't, and logic should tell you that: logic should guide your thought and tell you whether you are entitled to assert A or not. But this is possible only if there is just one logic.[8]

First note that the problem cannot raise a point against logical pluralism: there is no reason to believe that the question which logic to apply in a case of reasoning has a *general* solution with *one* answer that covers all cases. This practical issue does not force one to narrow down the range of acceptable logics to one system.

A monist might advance the following reasoning. Given $\Gamma\nvdash_I A$ and $\Gamma\vdash_C A$ and we accept all of Γ, we could always go on asserting A: classical logic provides us with a sufficient justification for asserting the conclusion. Thus the problem which logic to use has a simple solution: always use classical logic, as it is the stronger logic. Now it may very well be true that one could always use classical logic. However, this does not address the question whether this is always the right way of looking at a given case. The proof-theoretic justification of deduction shows that the classical analysis of arguments is not all there is to logic. So although it may be possible to treat every argument classically, this does not show that this treatment is always adequate, let alone that it is the only possible treatment. That this is so should be obvious in the case of conditionals. What I am arguing here is that the same phenomenon extends to negation. The examples of statements about the future and counterfactual situations discussed earlier show that negation may intelligibly be treated in a classical as well as in a non-classical way.

Much of the force of the pragmatic problem stems from the way it has been stated. A closer look at how such a problematic case could arise shows that it

[8] The situation is in some ways similar to a familiar one in mathematics. After the invention of Non-Euclidean geometries the question arose whether Euclidean or Non-Euclidean geometry should be used to describe the world. This is not a question for mathematics to decide; rather it depends on observations and experiments in physics. One might object that the case of alternative logics is inherently different from the case of alternative geometries, as there are no *experimenta crucum* which could decide which logic is the correct one: logic has no subject matter; it is 'topic neutral'. But this is not quite right. There are such experiments: our pre-theoretical logical intuitions provide the relevant data. However, the problem with them, as argued earlier, is that they do single out a unique logic as the right one.

does not in fact introduce any new problems. Let's assume for simplicity's sake that the same formal language is used for intuitionistic and classical logic, and let's write Γ^{o} and A^{o} for the ordinary language sentences we are formalising by Γ and A and $\Vdash$ for our intuitive notion of consequence. Applying the resources of formal logic, we discover that we have two ways of regimenting the informal argument for A^{o} from Γ^{o}, a classical and an intuitionistic case, and $\Gamma \vdash_{C} A$ and $\Gamma \nvdash_{I} A$. Then, accepting all of Γ^{o}, we ask ourselves: should we assert A^{o} or not? In other words, should we take $\Gamma^{o} \Vdash A^{o}$ to hold or not? Well, under which conditions can this question arise? That both $\Gamma \vdash_{C} A$ and $\Gamma \nvdash_{I} A$ happens only in very uncommon situations, namely if undecidable sentences are involved. Thus Γ^{o} and A^{o} will be sentences similar to the ones discussed in section 3. Thus we may recycle what has been said there. Our informal concept of negation is neither classical nor intuitionistic in the sense that neither of the two logics can claim to capture this concept either "entirely" or better than the other logic. Both logics give reasonable, well motivated ways of regimenting the informal concept. Extrapolating to the present case, whether or not you should consider $\Gamma^{o} \Vdash A^{o}$ to hold and go on asserting A^{o} depends on whether you intend to focus on the classical or the intuitionistic aspect of informal negation. There is no absolute answer to the question, no answer, that is, which would be independent of the formalisations.

I conclude that there is no pragmatic problem for pluralism either. I'll say a little more connected to this in the conclusion.

5. Conclusion

The problem of how to formalise natural language sentences to some degree always arises. It is a problem that everyone faces who thinks that formal logic may serve in the analysis of informal arguments. It can hardly be denied that in formalising natural language for the purposes of logic there is a bunch of options one can choose from. For instance, shall I treat 'or' as a primitive, or shall I analyse it in terms of conjunction and negation? Shall formalise a phrase 'the F' as a complete expression (a term) or as an incomplete one (a Russellian description)? Shall I formalise a conditional as a material one, a strict one, a variably strict one or a relevant one? What I am arguing for is more of this kind, only in a more radical case, as it does not seem to have been suggested very often in the case of negation. If, in analysing a natural language argument, natural language sentences are represented by formulas, a decision has to be made not only concerning how to represent the structure of the sentences in question, but also concerning which machinery the 'logical words' in them are subject to, this way making them precise. Formalisation, in other words, involves conceptual analysis. In the case of informal negation, the analysis involves making a decision whether it is to be treated classically or intuitionistically. Due to the underspecification of informal negation, if you analyse an argument in formal logic, you need to make a decision which aspect of negation it is that you are focussing on, the classical one or the intuitionistic one. Informal negation is neutral between the two. Once we've realised that there are two options of analysing negation, we can make explicit which one we focus on in an argument. But neither is "nearer to the truth" or "more fundamental" than the other. We have to live with two options, no absolute decision between them being possible. But this is not incoherent. You just need to make clear which of them you are using. Formal logic helps us making these different aspects precise (or, indeed, helps us noting their existence).

I have argued that neither logical nor pragmatic problems arise from accepting that both, classical and intuitionistic logic are all right. But a problem remains. It can hardly be the case that both realism and anti-realism are correct! For while it is true that no decision needs to be made which of classical and intuitionistic logic is "the right logic", we cannot equally accept both metaphysics that each logic according to Dummett gives rise to. At least one of them has to go. But as we have no basis for deciding which one, given the proof-theoretic justification of deduction fails to decide between classical and intuitionistic logic, we should reject *both* metaphysics. We should give up the thought that proof-theory could provide a logical basis for metaphysics and that using one or other logic commits one to a certain metaphysics. Proof-theory is metaphysically neutral.

This leaves the question what to do about the notion of truth: does it or does it not satisfy the principle of bivalence? I take it that this question can adequately be dealt with by adopting a minimalist or pro-sentential theory of truth, but there is no space to go into this here.

Department of Philosophy
King's College London
The Strand
WC2R 2LS
London
United Kingdom
email: nils.kurbis@kcl.ac.uk

References:

Dummett, M. (1978). The Justification of Deduction. In *Truth and Other Enigmas*. Cambridge, Mass.: Harvard University Press.

Dummett, M. (1993). *The Logical Basis of Metaphysics.* Cambridge, Mass.: Harvard University Press.

Prawitz, D. (1965). *Natural Deduction. A Proof Theoretical Study.* (Acta Universitatis Stockholmiensis, Stockholm Studies in Philosophy 3). Stockholm, Göteborg, Uppsala: Almquist & Wiksell.

Prawitz, D. (1977). Meaning and Proofs: On the Conflict Between Classical and Intuitionistic Logic. *Theoria 43*, 2-40.

Resnik, M. (1996). Ought There to be but One Logic? In B. J. Copeland (Ed.). *Logic & Reality. Essays on the Legacy of Prior* (pp. 489–517). Oxford: Clarendon.

Sainsbury, M. (2002). What logic should we think with? In A. O'Hear (Ed.). *Logic, Thought & Language* (pp. 1–18). Cambridge: Cambridge University Press.

Semantics of the Axiom (Schema) of Comprehension

Pavel Materna

The problem with ϕ

Axioms and axiom schemata of various theories of sets surely reflect some basic intuitions. Mostly the correspondence between the axiom and our intuitions does not involve any serious problem. As an example we can adduce the Axiom of Extensionality:

$$\forall \mathrm{ab}\ (\mathrm{a} = \mathrm{b} \leftrightarrow \forall x\ (x \in \mathrm{a} \leftrightarrow x \in \mathrm{b}))$$

The intuition connected with this axiom consists in our conviction that the sufficient and necessary condition of identifying sets A and B is that A has the same members as B. Some delicate questions can be asked from the viewpoint of Philosophy of Mathematics but the basic intuition is clear.

This is not the case with the Axiom Schema of Comprehension, AC. The specific character of AC in this respect became obvious as soon Russell in his famous letter to Frege discovered the possibility of arriving to a paradox. See Russell (1902).

Let us use the common way of putting down the AC:

AC $\exists \mathrm{X} \forall x\ (x \in \mathrm{X} \leftrightarrow \phi(x))$

or its restricted form

RAC $\forall \mathrm{Y} \exists \mathrm{X} \forall x\ (x \in \mathrm{X} \leftrightarrow x \in \mathrm{Y} \wedge \phi(x))$.

The source of the problem with AC is "ϕ". Our question is:
What is the semantics of "ϕ"?

We must state that the verbal characteristics of AC are at least careless. All in all three proposals of interpreting "ϕ" can be found in various sources:

a) "ϕ" is a *formula* (not containing a free occurrence of "X"),
b) "ϕ" is a *predicate* (in one variable, not using the symbol "X"),
c) ϕ is a *property*.

The proposals a) and b) only postpone the answer: they do not say what "ϕ" *means* or *denotes*, they only define the *kind of expression* that has to be used in the place of a particular occurrence of "ϕ".

What about *property?*

We find a characteristic formulation (FOLDOC):

"An axiom schema of set theory which states: if P(x) is a property then $\{x : \mathrm{P}\}$ is a set. I. e. all things with some property form a set."

Now what is symptomatic is that the entry property has to be explained but instead we read: "No match for property. Sorry, the term property is not in dictionary." This means that the term is considered to be a common term, which is understood by everybody.

Another source justifies the property-interpretation:

"...we have seen sets introduced in three ways.

1. By listing elements: {1, 2, 3} This works only for finite sets of manageable size.
2. *By properties of their elements:* $\{x : x$ *is an even natural number*$\}$ *or* $\{x : x$ *was listed in the 1900 US Census*$\}$. Sets listed in this way are usually infinite ...or unmanageably large.... The first case falls under this case as well: $\{1, 2, 3\} = \{x : x = 1 \mid x = 2 \mid x = 3\}$.
3. By cheating: **N** = {0, 1, 2, ... }."

Yet the examples adduced show that what is called *property* here is not a property in the sense of P(ossible)W(orld)S(emantics): the PWS properties are considered to be functions from possible worlds to sets of objects (or to the chronologies of sets of objects – see Tichý (1988, 2004)) whose values are distinct in at least two possible worlds-times. If properties are defined in this way then "ϕ" in AC cannot denote a property.

We can, of course, classify as properties such functions whose value is the same in all pairs <world, time>, and call them *trivial properties*, but the 'logical behavior' of such properties makes them indiscernible from classes / sets of objects. Can we perhaps interpret ϕ as being a *class* or a *set*?[1]

Let us try. According to **RAC** the set of primes **P** would be defined as follows:

$$\forall x\,(x \in \mathbf{P} \leftrightarrow x \in \mathbf{N} \wedge \mathbf{P}(x)),$$

which is absurd. But perhaps we could introduce the set/class **P** in one of the three ways mentioned above. The 1st way works for finite sets only. The 3rd way (here {2, 3, 5, 7, 11, 13, 17, 19, 23, ...}) is really a kind of cheating. Thus we

[1] For our purposes it is not necessary to distinguish between sets and classes.

choose the 2^{nd} way ("by properties of their elements"), and since such 'properties' cannot be empirical properties they will be again classes. An infinite regress.

The way to the solution

It would seem that the set **P** could be defined according to **RAC** without the above absurd consequences. The 'property' ϕ would be *having exactly two factors*. Let it be denoted by **ETF**. Then we have

$$\forall x\,(x \in \mathbf{P} \leftrightarrow x \in \mathbf{N} \wedge \mathbf{ETF}(x)).$$

What is however the semantics of **ETF**? It evidently denotes a class, the class of primes, that is. Thus the result of accepting this definition is extremely meager: Instead of **ETF** we can write **P**.

All the same, there is obviously a sound idea behind this last proposal. We are used to interpret *having exactly two factors* as a property: we don't care whether it is an empirical or a 'non-empirical' property, we simply feel that it is something what can characterize some objects (here: numbers) and distinguish them from other objects. What is essential for this naïve interpretation is the fact that we seriously respect the particular components of the verbal characteristics, in particular the expressions *factor*, *exactly two*, *have,* and the way they are connected due to the grammar of the given language. Thus we can say that what is of interest for us is not the class as a simple mapping but a *structured way* leading to the class. In other words, we take into account the fact that beside the 'flat', i. e., set-theoretical objects there are *complexes*, *structured entities, abstract procedures.*[2]

Preliminarily we can say that the ϕ from AC has to be considered as a *presentation* of some class, so it cannot be a set-theoretical object because, as rightly says Zalta in his (1988):

"Although sets may be useful for describing certain structural relationships, they are not the kind of thing that would help us to understand the nature of presentation. There is nothing about a set in virtue of which it may be said to present something to us."

So we have to go over to *complexes* in contrast to set-theoretical objects like classes.

[2] It was Bernard Bolzano, who was aware of the fact that *concept*, unlike its content, has to be a complex. See (Bolzano, 1837, p. 244).

Complexes

In 1968 the young Czech logician Pavel Tichý (see Tichý, 2004, p. 79) stated that

> "[t]he notion of an effective procedure plays an almost negligible role in current logical semantics."

and decided that this situation has to be changed because

> "[t]he relation between sentences and procedures is of a semantic nature; for sentences are used to record the results of performing particular procedures." (*Ibidem*, p. 80)

This idea infiltrated all the work by Tichý, who has founded T(ransparent) I(ntensional) L(ogic) (see Tichý, 1988, 2004). Since we will use TIL when solving our problem we have to adduce some basic information about TIL.

Intermezzo: *Basic notions of TIL.*

A. Informal characteristics

TIL is a typed system, where types are sets of functions (this orientation is shared with Montague's IL).

Basic atomic types (sets of nullary functions) are chosen dependently on the kind of problems to be solved. If natural language has to be analyzed then the four atomic types are ι (individuals, cf. Montague's e), o (truth-values, cf. Montague's t), τ (time moments, also real numbers), ω (possible worlds, cf. Montague's 'non-type' s). For analyzing extensional systems (our case of AC) we do not need possible worlds and, properly speaking, we do not need individuals either: we can use numbers (i.e., type τ for real numbers or ν for natural numbers).

Functional types are sets of partial functions. The *type* of a function whose values are in a type α and arguments are tuples of objects of the types $\beta_1,\ldots,\beta_m$, respectively, i. e., the *set* of all such functions, is recorded as $(\alpha\beta_1\ldots\beta_m)$. Classes of objects of the type α are identified with characteristic functions, so their type is $(o\alpha)$.

Constructions are abstract procedures. We will need four of them[3]. The obvious inspiration by λ-calculi (shared by Montague) is understandable: Church's

[3] Tichý defines still other two constructions; further constructions can be added if necessary.

ingenious idea of an essential reduction of procedures to *constructing functions* and *applying functions to arguments* proved to be most fruitful, in particular in the environment of computers. All constructions are thought to be *abstract procedures*. They are:

a) *variables*, countably infinitely many for each type; they are incomplete in that they construct objects dependently on valuations (they "*v*-construct" with *v* the parameter of valuations;
b) *trivializations*, which construct an object immediately, not using another construction;
c) *compositions*, which (*v*-)construct the value of a function on the given argument;
d) *closures*, which (*v*-)construct a function *via* an 'abstraction'.

Higher-order types. Types of order 1 are basic atomic types and functional types over them. The ramified hierarchy defines higher-order types. Roughly: Constructions of order *n* are defined (they construct objects of lower order types) and the set of all constructions of order *n,* denoted by $*_n$, is the type of order $n + 1$. Thus let a variable *x* be a numerical variable, i.e., a variable that *v*-constructs (say, real) numbers. Since the latter are objects of the type of order 1, *x* is a construction of order 1, i.e., its type is $*_1$, which means that its *type* is of order 2.

(Observe: constructions construct objects of a type α but the type of the constructions themselves is distinct. To distinguish both we use other symbols. In our example we would write $x \to \tau$, $x / *_1$.)

B. Notation

The exact definitions can be found in TIL literature, in particular in Tichý (1988, 2004) or, e.g., Duží & Materna (2005). Thus we adduce here only the way in which the particular constructions will be written. Remember that *the record of a construction is not the construction: the latter is – unlike the former – an abstract procedure*, so it cannot contain letters, brackets etc.

Trivialization of an object (including constructions) X ... ^{0}X
Composition: where $X \to (\alpha\beta_1...\beta_m)$, $X_i \to \beta_i$ ... $[XX_1...X_m]$
Closure: where $x_1 \to \beta_1, ..., x_m \to \beta_m$, and $X \to \alpha$ ... $[\lambda x_1...x_m X]$

C. Types of some logical objects

Connectives: negation ¬ / (οο)
binary connectives ∧, ∨, ⊃, ↔ / (οοο)
Quantifiers: ∀, ∃ / (ο (οα)) (schema; a case of type-theoretical polymorphism)

End of Intermezzo.

That component of the semantics of an expression E which makes us understand E (and corresponds to what Frege intuitively meant by his *sense*) will be called *meaning* here; it serves to presentation of the object (if any) denoted by E. We have already suggested (exploiting Zalta's right observation) that meaning cannot be a set-theoretical object: it has to be a complex. Now we have suggested an explication of the term "complex" so that we can claim that *the meaning of an expression is a construction.*

In Jespersen (2004) the author has shown that AC can be semantically analyzed from this viewpoint: the meaning of AC is a construction (i. e., a procedure) whose particular steps are visible from inspecting the record of this construction. A subconstruction of this construction is $[\phi x]$. The respective step is in Jespersen (2004) analyzed into three steps:

"[1] Execute ϕ to obtain a set.

[2] Execute x to obtain an α-object.

[3] Apply the set obtained in [1] to the α-object obtained in [2] to obtain a truth-value."

My question + comment concerns the step [1]. Now it is evident that ϕ is a *procedure*, and since (abstract) procedures are handled as *constructions* in TIL we can ask: *Which kind of construction should be ϕ?*

Our answer will be derived from the following points:

a) Let C be such a construction. The object (if any) constructed by C is obviously a set whose members are some objects of a type α. Thus

$$C \rightarrow (o\ \alpha),$$

while, of course,

$$C\ /\ *_n.$$

b) Further, C cannot contain any free variable. It is a *closed construction.* Elsewhere we have proposed (see Materna, 1998, 2004) an explication of the term ***concept*** according to which concept would be just a *closed construction*.[4]

c) Also, C cannot be an empirical concept: the constructed object is a class, it is no property. Thus any naïve examples (to be found, e.g., under the head *The axioms of Naïve Set Theory* (see http://home.sprynet.com/~ow/not244c.pdf)) like

"There is a set of all cats $\exists s\ \forall x\ (x \in s\ \leftrightarrow \mathrm{C}(x))$"

[4] Among the assets of this proposal there is one highly important: if concept is a procedure then we can justify the intuition according to which we can use various distinct concepts to identify one and the same object; an intuition, that is, that has been defended by the great logician Bolzano (see Bolzano, 1837).

are naïve but they are not examples of applying AC. (We will return to this point later.)

If however ϕ is a concept, a closed construction, how should we interpret the sloppy notation "$\phi(x)$"?

The x ranges over (v-constructs) the objects of the type α, the concept C constructs a class of objects of the type α (see **a)**).

Now we can write down a more fine-grained schema of C. It is a construction without any occurrence of a free variable (see **b)**) that constructs a class of some objects (see **a), c)**), so the schema of C can be written as

$$\lambda y\ [...y...],$$

where $y \rightarrow \alpha$ and $[...y...]$ v-constructs a truth-value. Thus instead of $\phi(x)$ we have

$$[[\lambda y\ [...y...]]\ x],$$

i.e., a composition that constructs the value of the function constructed by λy $[...y...]$ on the argument provided by the variable x.

Thus the set X from **RAC** is constructed as the subset of Y such that its members 'fall under' the concept λy $[...y...]$. The first part of our answer to the question about the semantics of ϕ is thus:

ϕ *is a concept.*

Comparing this answer with the alternatives *formula, predicate, property* can we say that it does not share the problems connected with the former?

One possible objection to our proposal can be formulated as follows:

According to this proposal, $\phi(x)$ would have the form $[[\lambda y\ [...y...]]\ x]$, but then the objection we have formulated above to **P** simply instead of **ETF** could be repeated: the class X that should be 'created' by AC is now simply denoted by $[\lambda y\ [...y...]]$, i.e., in another way.

To meet this objection we must recapitulate the distinction between semantics of *meaning* and semantics of *denotation*.

To 'interpret' or 'logically analyze' an expression E means to find the meaning of E rather than its denotation. Meanings are complexes, denotations are often 'flat' objects like sets, functions. Thus a logical analysis of the sentence *Three times two is greater than three plus two* consists in recording the respective procedure (in TIL: $[^0\!> [^0\!*\ ^0 3\ ^0 2]\ [^0\!+\ ^0 3\ ^0 2]]$) rather than stating its truth-value. Another example:

Let $\lambda x\ (x + 1)$ be interpreted a) as a λ-term, b) as expressing a construction.

In the case of a) we say that the expression denotes the function *successor*, in the case of b) we say that the meaning of the expression is the procedure given by the construction $\lambda x\ [{}^0{+}\ x\ {}^01]$.[5]

As Jespersen has shown in his (2004), AC can be analyzed as a procedure; the meaning of any instance[6] of AC should be (not a truth-value but) a procedure. The meanings of the particular instances of AC, as well as of their particular subexpressions, are concepts. Thus $\lambda y\ [...y...]$ is analyzed as a concept, a procedure rather than a name of some class: the class to be 'created' must be first *constructed*.

Russell's paradox

Which concept plays the role of ϕ in the case of Russell's paradox?

Curiously enough, none, at least from the viewpoint of TIL.

Let y v-construct sets. Let $\in$ be the membership relation, so $\in$ / $(o\ \alpha\ (o\alpha))$. The source of the paradox is the "set of normal sets", i. e., "the set of those sets that are not members of themselves". Let us try to record the concept that should correspond to this verbal definition. We have to write down

$$\lambda y\ [{}^0\neg\ [{}^0{\in}\ y\ y]]$$

but we immediately see that *there is no construction, and so no concept here* (see **B.** in *Intermezzo* above*,* type-theoretical checking).

Indeed, one can demur that this result is dependent on accepting hierarchy of types; Russell's way of avoiding the paradox is just hierarchy of types, TIL has defined a (functional) hierarchy of types.

Axiomatic set theories such as ZF do not accept types, but they also do not admit – in a way distinct from Russell's – definitions containing or presupposing self-membership. So it seems that any sound intuition would refuse possibility of a procedure that could realize self-membership.[7]

Thus we can say that *Russell discovered (or 'invented', if you like) the hierarchy of types to demonstrate that existence of a concept (in our sense) that would realize self-membership is incompatible with our intuition.*

---

[5] A (here innocuous) simplification: Which procedure / construction it is depends on which 'primitive concepts' are at our disposal. See Materna (2004).

[6] I. e., for RAC, an instance is the formula that arises from RAC by omitting $\forall$Y, substituting a name of a particular set for Y and a name of a particular condition for ϕ.

[7] Semantics of type-free λ-calculus (Scott's domains) does not correspond to intuitions connected with the 'self-membership'.

Concepts for ϕ in AC

Recapitulating that the concepts that can play the role of ϕ in AC are of the form

$$\lambda y\ [...y...],$$

where [...y...] v-constructs truth-values, we can ask whether any concept of this form can be accepted.

To begin with unimportant banal cases, evidently concepts that identify the universal or the empty class, e.g.,

$$\lambda y\ [{}^0{=}\ y\ y],\ \lambda y\ [{}^0 \neq y\ y],$$

are surely not concepts for sake of which AC has been introduced. Does it mean that any concept that constructs a non-empty proper subset of the respective universal set is admissible?

A. Once more, *empirical concepts are not admissible.*

Empirical concepts do not denote classes: they always denote intensions, i.e., objects whose type is $\alpha_{\tau\omega}$ for some type α. To return to the example with cats, i.e.,

"There is a set of all cats $\exists s\ \forall x\ (x \in s\ \leftrightarrow \mathrm{C}(x))$",

we see that "C" is supposed to denote a class, whereas it denotes a property, type $(o\iota)_{\tau\omega}$. Again, no concept corresponds to the deceptive verbal formulation.

B. Let g be a bijection, type $(\tau *_1)$ that enumerates constructions that correspond to well-formed formulas of the 1st order predicate logic. Let Th / $(o *_1)$ be the class of the constructions that underlie theorems of the 1st order predicate logic. Consider the following concept:

$$\lambda y\ [{}^0\exists\ \lambda z\ [{}^0\wedge\ [{}^0{=}\ y\ [\mathrm{g}\ z]][{}^0\neg[\mathrm{Th}\ z]]]]$$

Does this concept construct the class of Gödel numbers of the non-theorems (of their respective constructions) of the 1st order predicate logic? Yes, it does.

Can this concept be used as ϕ in AC? No, it cannot. The concept is not an algorithm.

Let us adduce an important quotation from Tichý's article "Constructions" (1986), see Tichý (2004, p. 613):

> "[n]ot every construction is an algorithmic computation. An algorithmic computation is a sequence of *effective* steps, steps which consist in subjecting a manageable object (usually a symbol or a finite string of symbols) to a feasible operation. A construction, on the other hand, may involve steps which are not of this sort. The application of any function to any argument, for example, counts as a legitimate constructional step; it is not required that the argument be finite or the function effective. Neither is it required that the function constructed by a closure have a finite domain or be effective. As distinct from an algorithmic computation, a construction is an *ideal* procedure, not necessarily a mechanical routine for a clerk or a computing machine."

From this viewpoint we can distinguish two kinds of concepts, see Duží, Materna (2004): 1) the concepts that are *effective procedures* and 2) the other concepts. The concepts of the first kind we have called *analytic concepts,* the members of the second group *synthetic concepts* (a hint at Kant).

All objects we can talk about are modeled in TIL as functions (see the definition of types). The majority of them are, of course, non-recursive functions. They can be conceptually identified but the respective concepts are synthetic, non-algorithmic. As for the recursive functions they can be identified either by synthetic, non-algorithmic concepts, or by the analytic concepts (see Kleene, 1952, p. 317). Obviously, using AC as a means of discovering ('creating', if you like) classes we have to choose only analytic concepts.

Conclusion

As soon as we are aware of the fact that the ϕ in AC is a concept, i.e., a procedure, all restrictions of using AC reduce to the requirement that the respective concept were an effective procedure, an algorithm. Russell's paradox can be seen then as a consequence of the simple fact that there is no concept at all which would correspond to the verbal / symbolic articulation of the 'set of normal sets'.

Pavel Materna
Institute of Philosophy
Academy of Sciences of the Czech Republic
110 00, Praha 1, Jilská 1, Czech Republic.
materna@lorien.site.cas.cz

References

Bolzano, B. (1837). *Wissenschaftslehre I.* Sulzbach.

Duží, M., & Materna, P. (2004). A procedural Theory of Concepts and the Problem of Synthetic a priori. *Korean Journal of Logic, 7*(1), 1–22.

Duží, M., & Materna, P. (2005). Logical Form. In G. Sica (Ed.). *Essays on the Foundations of Mathematics and Logic* (pp. 115–153). Milan: Polimetrica.

Jespersen, B. (2004). Aussonderung as Procedure. In L. Běhounek (Ed.). *The Logica Yearbook 2003* (pp. 133–144). Praha: Filosofia.

Kleene, S. C. (1952). *Introduction to Metamathematics*. New York, Toronto: D. van Nostrand Company, Inc.

Materna, P. (1998). Concepts and Objects. *Acta Philosophica Fennica, 63*. Helsinky: Societas Philosophica Fennica.

Materna, P. (2004). *Conceptual Systems*. Berlin: Logos Verlag.

Russell, B. (1902). Letter to Frege. In J. van Heijenoort (Ed.). *From Frege to Gödel* (p. 125). Cambridge, Mass., London, New York: HUP.

Tichý, P. (1988). *The Foundations of Frege's Logic*. Berlin: De Gruyter.

Tichý, P. (2004). *Pavel Tichý's Collected Papers in Logic and Philosophy.* Edited by V. Svoboda, & B. Jespersen, & C. Cheyne (Eds.). Praha: Filosofia, Dunedin: University of Otago Press.

Zalta, E. N. (1988). *Intensional Logic and the Metaphysics of Intentionality.* Cambridge, Mass., London: MIT Press.

Paraconsistency and Logical Hypocrisy

Casey McGinnis

1. Introduction

A paraconsistent logic is one in which inconsistent theories do not necessarily "explode" into triviality. A paraconsistent logician (or "paraconsistentist") is, of course, a logician who advocates some form of paraconsistent logic. An accusation one sometimes hears is that paraconsistent logicians are "hypocrites," in that they routinely make use of logical rules or principles that are, by their own standards, incorrect. After stating this accusation as convincingly as I can, I will consider two ways, based on approaches in the literature, in which the paraconsistent logician might attempt to show that he or she is not really hypocritical in this sense. I will argue that neither of these approaches is wholly satisfactory. Finally, I will consider a different approach to paraconsistency that is, I believe, truly non-hypocritical (though it has some rather radical and perhaps unpleasant consequences).

2. Paraconsistent logic

To repeat, a paraconsistent logic is one in which inconsistent theories are not always trivial. What does this mean? A *theory* (as I will use the term) is a set of sentences (or well-formed formulas) that is closed under logical consequence, i.e., that contains all of its logical consequences. A theory is *inconsistent* just in case it contains some sentence and its negation. And a theory is *trivial* just in case it contains every sentence of the language.

The key feature of a paraconsistent logic is that, unlike classical logic (and many nonclassical logics, such as intuitionistic logic), it invalidates the principle of "Explosion," according to which anything whatsoever follows from a contradiction: $\alpha, \neg\alpha \vdash \beta$. (Here '$\vdash$' expresses a generic logical consequence relation.)

As is well known, any paraconsistent logic must reject at least one of the following principles, as they jointly entail Explosion:

Disjunctive Addition: $\alpha \vdash \alpha \vee \beta$

Disjunctive Syllogism: $\alpha \vee \beta, \neg\alpha \vdash \beta$
Cut: if $\Gamma \vdash \alpha$ and $\Delta, \alpha \vdash \beta$ then $\Gamma, \Delta \vdash \beta$

The most common approach to paraconsistent logic is to allow models in which α and $\neg\alpha$ are both true but β is not. (There are, of course, other approaches.) Such systems typically invalidate Disjunctive Syllogism but preserve Disjunctive Addition, Cut, and most other classically correct principles.

3. J'accuse!

The accusation I want to consider in this paper runs as follows:

> "Paraconsistent logicians do not really *believe* in the logics they advocate. For consider how they reason when, e.g., proving metatheorems about their favored paraconsistent logics. They routinely rely on principles that are, by their own lights, incorrect. In this sense, paraconsistent logicians are *logical hypocrites*."

What, exactly, are the paraconsistently unacceptable inference patterns or reasoning techniques that the paraconsistentist allegedly uses? As we have seen, the most notable inference schemas that are paraconsistently invalid are Explosion and Disjunctive Syllogism. It is quite hard, however, to find an *explicit* instantiation of either of these schemas in the (written or spoken) reasoning of paraconsistent (or any other) logicians. To be sure, these patterns often seem to be *implicitly* invoked in some sense, but such apparent invocations can easily be "explained away" in paraconsistently legitimate terms. For example, logicians often say things like "It's not α, so it must be β." This appears to be an enthymemic application of Disjunctive Syllogism: $[\alpha \vee \beta], \neg\alpha \vdash \beta$ (Square brackets indicate suppressed premises.) However, it can also be characterized as an innocuous application of Modus Ponens: $[\neg\alpha \rightarrow \beta], \neg\alpha \vdash \beta.$

One of the most common proof techniques used by logicians, including paraconsistent logicians, is *Reductio ad Absurdum*, i.e., showing that α is true by deriving a contradiction, $\beta \wedge \neg\beta$, from $\neg\alpha$. Can this be a paraconsistently valid form of reasoning? It seems not; for once one allows that β and $\neg\beta$ may be jointly true, it is hard to see why one should reject $\neg\alpha$ merely because it implies $\beta \wedge \neg\beta$.

In fact, many paraconsistent logics *do* validate the inference from $\neg\alpha \rightarrow (\beta \wedge \neg\beta)$ to α. Consider, for example, the three-valued logic RM3 (Anderson & Belnap, 1975), which can be defined as follows. An RM3 interpretation (or model) is a relation R that relates each atomic formula to *at least* one truth value, 1 (true) or 0 (false): $\alpha R1$ means that α is (at least) true, and $\alpha R0$ means that α is (at least) false. The relation is extended to all formulas as follows:

$\neg\alpha R1$ iff $\alpha R0$
$\neg\alpha R0$ iff $\alpha R1$
$\alpha \wedge \beta R1$ iff $\alpha R1$ and $\beta R1$
$\alpha \wedge \beta R0$ iff $\alpha R0$ or $\beta R0$
$\alpha \vee \beta R1$ iff $\alpha R1$ or $\beta R1$
$\alpha \vee \beta R0$ iff $\alpha R0$ and $\beta R0$
$\alpha \rightarrow \beta R1$ iff (not($\alpha R1$) or $\beta R1$) and (not($\beta R0$) or $\alpha R0$)
$\alpha \rightarrow \beta R0$ iff $\alpha R1$ and $\beta R0$

Thus, for example, $\alpha \wedge \beta$ is true just in case α and β are both true, and false just in case either α or β is false. The sentence α is a semantic consequence of Γ in RM3 (in symbols, $\Gamma \vDash_{\text{RM3}} \alpha$) iff every interpretation such that $\beta R1$ for all $\beta \in \Gamma$ is such that $\alpha R1$. That is, α is a semantic consequence of Γ just in case α is (at least) true whenever every element of Γ is (at least) true. It is easy to check that $\alpha, \neg\alpha \nvDash_{\text{RM3}} \beta$ but $\neg\alpha \rightarrow (\beta \wedge \neg\beta) \vDash_{\text{RM3}} \alpha$.[1]

This leads to another objection: "It is true that in your paraconsistent logic, deriving a contradiction from $\neg\alpha$ allows you to conclude that α is *at least* true. But surely you mean to prove something stronger than that, namely that α is true *only* – i.e., true *and not false*. Your paraconsistent logic does not license drawing this conclusion."

The paraconsistent logician may respond as follows: "I am assuming that α is 'well-behaved', i.e. not both true and false. (Surely you, the traditionalist, will not challenge *that* assumption!) And it can easily be shown that on my semantics, if α is well-behaved and $\neg\alpha$ implies a contradiction, then α is uniquely true."

The paraconsistentist's metatheoretical claim about his semantics is easily verified (assuming he has set up the semantics right – as in, e.g., RM3). But how, exactly, is the inference in question to be represented *in* a paraconsistent logic? In particular, how are we to represent the notion that α is well-behaved? $\neg(\alpha \wedge \neg\alpha)$ will not do, for this schema holds for all formulas α, even ill-behaved ones. In many paraconsistent logics, such as RM3, there is simply no way to say that a sentence is well-behaved. With respect to RM3, this is easy to see by considering the interpretation that assigns both 1 and 0 to all atomic formulas. A simple proof by induction shows that all formulas are assigned both 1 and 0 under this interpretation. Thus, in general, there is no formula $\Phi(\alpha)$ such that $\Phi(\alpha)$ is true just in case α is well-behaved.

Perhaps more worrisome is the fact that in a logic like RM3 there appears to be no way to truly deny, reject, or express disagreement with a formula. (Cf. Batens, 1990; Parsons, 1990.) One can of course assert $\neg\alpha$, but this is perfectly

[1] One way to verify the latter is to observe that both Modus Tollens ($\alpha \rightarrow \beta, \neg\beta \vdash \neg\alpha$) and (a version of) the Law of Non-Contradiction ($\vdash \neg(\alpha \wedge \neg\alpha)$) both hold in RM3.

compatible with α's being true! (Dually, one cannot express genuine *agreement* with a formula, since asserting α does not rule out the truth of $\neg\alpha$.) Prima facie, this is an unacceptable limitation on expressibility. The paraconsistentist has got some explaining to do if she is going to make a convincing case that she is not a logical hypocrite.

4. Two ways of attempting to avoid hypocrisy

I now consider two ways, based on approaches in the literature, that the paraconsistentist might attempt to tell a story showing that he is not, after all, logically hypocritical – that he does not reason with principles that are, by his own standards, incorrect.

4.1 Internalizing consistency

Suppose we add a new unary connective, $\circ$, to RM3. (Call the resulting system RM3°.) This new connective has the following truth and falsity conditions:

$\circ\alpha R1$ iff not($\alpha R1$ and $\alpha R0$)
$\circ\alpha R0$ iff $\alpha R1$ and $\alpha R0$

The formula α can thus be read as "α is not both true and false" or "α is well-behaved." In the terminology of Carnielli et al. (2005), RM3° is a "logic of formal inconsistency" (perhaps better named 'logic of formal consistency' or 'logic of formal well-behavedness'), since the semantic consistency (i.e. well-behavedness) of a formula can be expressed in the object language.[2] (Note that the *ill*-behavedness of α *can* be expressed in RM3, using $\alpha \wedge \neg\alpha$, which is true just in case α is both true and false.)

The operators Δ and ∇, meaning *true only* and *false only*, can be defined as $\Delta\alpha =_{df} \alpha \wedge \circ\alpha$, and $\nabla\alpha =_{df} \neg\alpha \wedge \circ\alpha$. (Note that ∇ is just Boolean negation.) It is easy to check that $\nabla\alpha$ is true just in case α is *false and not true*. Thus we can now express genuine disagreement with α simply by asserting $\nabla\alpha$. Similarly, we can express genuine agreement with α by asserting $\Delta\alpha$.

RM3° validates the following form of Reductio: $\neg\alpha \rightarrow (\beta \wedge \neg\beta), \circ\alpha \vdash \Delta\alpha$. One advocating this kind of paraconsistent logic can claim that *this* is the form of Reductio she uses. However, by introducing a notion of consistency or "well-behavedness" (and hence Boolean negation) into her logic, this paraconsistent

2 Reading $\circ\alpha$ as 'α is consistent' is somewhat infelicitous, since $\circ(\alpha \wedge \neg\alpha)$ could very well be true, and $\alpha \wedge \neg\alpha$ is normally seen as the paradigm of an inconsistent formula.

logician introduces a new – or, perhaps, the old – form of explosion, namely A, $\nabla A \vdash B$ (as well as its dual, $\neg A$, $\Delta A \vdash B$). Thus the claim that her logic is capable of formalizing inconsistent but nontrivial theories must be qualified: her logic *cannot* formalize nontrivial theories that are inconsistent with respect to *Boolean* negation; such theories explode, as in classical logic. Now, perhaps this feature is defensible. But if the aim of paraconsistent logic is to allow sensible, nontrivial inference from inconsistent theories, one would think that the *type* of negation with respect to which a theory is inconsistent should not matter. If Explosion is counterintuitive with respect to *any* kind of negation, it is counterintuitive with respect to *every* kind of negation.

4.2 Distinguishing between denying something and asserting its negation

Another approach that the paraconsistentist might adopt is to reject the commonly held view that to deny, reject, or express disagreement with α is simply to assert, accept, or express agreement with $\neg\alpha$. This is the approach taken by Priest (2006).[3] On Priest's view, the fact that the truth of $\neg\alpha$ does not rule out the truth of α does not show that it is impossible to express disagreement with α. For one expresses disagreement with α not by asserting $\neg\alpha$ but rather by denying α, which is (on Priest's view) a distinct speech act. Thus Priest would likely agree that when he uses Reductio, he shows that α is true but does not rule out its also being false. He can, however, express his disagreement with α simply by denying it.

One apparent problem with this view is that it seems to disallow *logically demonstrating* that something is uniquely false (or uniquely true, or well-behaved). One can express disagreement with α, but cannot provide logical grounds to back up this disagreement. One can only provide logical grounds for agreeing with $\neg\alpha$, which is not the same thing. One who takes Priest's view might respond that it is possible to provide logical support for one's disagreement with α in a way that is precisely dual to the way one would provide logical support for one's agreement with α. Just as one would provide logical support for accepting α by showing that it is entailed by something that ought to be accepted, one can provide logical support for rejecting α by showing that it entails something that ought to be rejected. (For example, I might provide logical grounds for rejecting a theory of physics by showing that it entails that two objects can occupy the same space at the same time.) This, I think, is an adequate response.

There is, however, another problem with the approach under consideration. Can an agent simultaneously accept and reject, or agree with and disagree with,

[3] It should be noted that Priest distinguishes sharply between accepting and rejecting, which he takes to be cognitive states, and asserting and denying, which he takes to be speech acts.

a proposition? According to Priest (2006, pp. 103, 110), the answer is 'no'. And this makes sense, for otherwise rejecting something would no more rule out accepting it than accepting its negation would. Consider this, however: it seems reasonable to treat 'it is accepted (by agent *a*) that' and 'it is rejected (by agent *a*) that' as formalizable modal operators. (Cf. 'it is known (by agent *a*) that' as formalized in epistemic logic.) In a paraconsistent logic of acceptance and rejection that conforms to Priest's views, we would have $\mathsf{Accept}(\alpha)$, $\mathsf{Reject}(\alpha) \vdash \beta$, which seems no less counterintuitive than (the standard version of) Explosion. (The inference would be valid, of course, because there would be no models in which an agent both accepts and rejects the same proposition.) It seems that the current approach, like the one considered before it, has evaded Explosion in letter but not in spirit.

5. Truth-preservation reconsidered

As we have seen, neither of the above defenses against the charge of hypocrisy is entirely adequate. I now want to consider a different approach to paraconsistency that is, I believe, not vulnerable to the accusation of hypocrisy. What I will propose here is similar to approaches entertained by Strawson (1948), Smiley (1959), Priest (1999), and Bremer (2005, p. 236).

The standard definition of (semantic) logical consequence is as follows: $\Gamma \vDash \alpha$ just in case every model of Γ is a model of α, i.e., there is no model of Γ that is not also a model of α. Thus, whenever there are no models of Γ, there are, *a fortiori*, no models of Γ that are not also models of α, whence $\Gamma \vDash \alpha$. What if we were to modify this definition as follows?:

> $\Gamma \vDash \alpha$ just in case (i) *every* model of Γ is a model of α; and
> (ii) *some* model of Γ is a model of α.[4]

Let us apply this definition to classical (propositional) logic, leaving everything else as usual. Call the resulting system C^*. It is easy to provide a tableau-style proof theory for C^*. Start with the usual definition of tableaus. To check the validity of Γ / α, construct *two* tableaus: one for $\Gamma \cup \{\neg\alpha\}$ and another for $\Gamma \cup \{\alpha\}$ (or just Γ). The inference is valid just in case the first tableau closes and the second does not.

Clearly $p, \neg p \vDash q$ (p, q atomic) fails in C^*. Thus we have defined a paraconsistent logic. Yet both $p \vDash p \vee q$ and $p \vee q, \neg p \vDash q$ hold in C^*, as does $\neg p \rightarrow$

[4] Equivalently, we could say: $\Gamma \vDash \alpha$ just in case (i) every model of Γ is a model of α; and (ii) there is a model of Γ. (For if there is a model of Γ and every model of Γ is a model of α, then obviously there is a model of $\Gamma \cup \{\alpha\}$.)

$(q \wedge \neg q) \vDash p$. Moreover, it is obvious that the valid sentences of C* are exactly those of classical logic. So far, C* looks like a very attractive system. However, the *general* (or schematic) forms of the following inferences all fail:

Disjunctive Addition: $\alpha \vDash \alpha \vee \beta$ [let $\alpha = p \wedge \neg p$]
Disjunctive Syllogism: $\alpha \vee \beta, \neg\alpha \vDash \beta$ [let $\alpha = \beta$]
Reductio ad Absurdum: $\neg\alpha \rightarrow (\beta \wedge \neg\beta) \vDash \alpha$ [let $\alpha = p \wedge \neg p$]

By 'formula schema' I mean: any schema built up from the schematic letters α, β, etc. via the usual connectives. The following fact will come in handy:

Lemma. If a formula schema is contingent (i.e. neither tautological nor contradictory), then it has an instance that is tautological and an instance that is contradictory.

Proof. An induction on the length of the schema. The result holds for atomic α, since $p \vee \neg p$ and $p \wedge \neg p$ are both instances of it. We need to show that the result holds for $\neg\alpha$ and $\alpha \vee \beta$. (The other connectives can be defined as usual.) Consider $\neg\alpha$. By the inductive hypothesis, α has an instance that is tautological and an instance that is contradictory. Call these $\alpha_\top$ and $\alpha_\bot$, respectively. Thus $\neg\alpha_\bot$ is tautological and $\neg\alpha_\top$ is contradictory. But $\neg\alpha_\bot$ and $\neg\alpha_\top$ are both instances of $\neg\alpha$. Thus the result holds for $\neg\alpha$. In a similar manner, it can be shown that the result holds for $\alpha \vee \beta$. ■

We now show that:

Theorem. An inference schema is valid in C* only if each element of its premise set is a tautology.

Proof. Suppose, for Reductio, that $\beta_1 ,..., \beta_n \vDash_{C^*} \alpha$ and that some β_i is not tautological. Then β_i is either contingent or contradictory. If it is contingent, then, by the Lemma, it has an instance that is contradictory. If it is contradictory, then obviously it has an instance that is contradictory. Either way, $\{\beta_1 ,..., \beta_n , \alpha\}$ has an instance that is unsatisfiable. Thus $\beta_1 ,..., \beta_n \nvDash_{C^*} \alpha$. (Contradiction.) ■

C* enjoys none of the following standard or "Tarskian" features of a logical consequence operator ($Cn(\Gamma) =_{df} \{\alpha : \Gamma \vDash \alpha\}$):

Reflexivity: $\Gamma \subseteq Cn(\Gamma)$ [let $\Gamma = \{p, \neg p\}$]
Transitivity: $Cn(Cn(\Gamma)) \subseteq Cn(\Gamma)$ [let $\Gamma = \{p, \neg p\}$]
Monotonicity: if $\Gamma \subseteq \Delta$ then $Cn(\Gamma) \subseteq Cn(\Delta)$ [let $\Gamma = \{p\}$, $\Delta = \{p, \neg p\}$]

Nor does C* verify the Deduction Theorem, which states that if $\Gamma \vDash \alpha \rightarrow \beta$ then $\Gamma, \alpha \vDash \beta$. (Let $\Gamma = \varnothing$, $\alpha = p \wedge \neg p$, $\beta = q$.) It should also be noted that while C* avoids *some* of the so-called "fallacies of relevance," such as Explosion, it retains others, such as $\neg p \vDash p \rightarrow q$ and $p \vDash q \vee \neg q$.

Despite the somewhat unhappy features listed above, the paraconsistentist who advocates C* (or another paraconsistent logic that is similar in the relevant respects) cannot be accused of logical hypocrisy. Suppose she applies Reductio ad Absurdum, for example. While the general form of this inference is not valid on her account, she can always find a specific instance of it that properly captures her reasoning and is valid, provided her premises are consistent with her conclusion. Moreover, she is not vulnerable to the objection that her system validates certain counterintuitive Explosion-like principles. The only objection to which she is vulnerable is that her system lacks some of the fundamental features that we normally take for granted in a logical consequence operator. But this may just show that non-hypocritical paraconsistency comes at the cost of considerable inconvenience, and a need to view logical validity in a much more fine-grained manner than that to which we are accustomed.[5]

Casey McGinnis
Department of Philosophy
University of Minnesota, Twin Cities
831 Heller Hall
Minneapolis, MN 55455
USA
mcgin017@umn.edu

References

Anderson, A. R., & Belnap, N. D. (1975). *Entailment: The logic of relevance and necessity*. Princeton: Princeton University Press.

Batens, D. (1990). Against global paraconsistency. *Studies in Soviet Thought*, *39*, 209–229.

[5] I would like to thank Manuel Bremer, Alexandre Costa-Leite, Nils Kurbis, and Graham Priest for helpful comments on earlier drafts of this paper. I have also benefited from discussions with Lou Goble. All (and probably some) mistakes herein are mine alone.

Bremer, M. (2005). *An introduction to paraconsistent logics*. Frankfurt: Peter Lang.

Carnielli, W. A., & Coniglio, M. E., & Marcos, J. (2005). Logics of formal inconsistency. In D. Gabbay, & F. Guenthner (Eds.). *Handbook of philosophical logic* (vol. 12, 2nd ed.). Dordrecht: Kluwer Academic Publishers.

Parsons, T. (1990). True contradictions. *Canadian Journal of Philosophy*, *20* (3), 335–354.

Priest, G. (1999). Negation as cancellation, and connexive logic. *Topoi*, *18*, 141–148.

Priest, G. (2006). *Doubt truth to be a liar*. Oxford: Oxford University Press.

Smiley, T. (1959). Entailment and deducibility. *Proceedings of the Aristotelian Society*, *59*, 233–254.

Strawson, P. F. (1948). Necessary propositions and entailment-statements. *Mind, 57*, 184–200.

On Building Abstract Terms in Type Systems*

Giuseppe Primiero

1. Abstraction and predication

This paper offers some historical and conceptual remarks on the philosophical and logical procedures of abstraction, based on an account of the notions of concept and function. In order to provide a complete analyis, one should start by considering Plato's theory of Ideas, which provides the first interpretation of "abstract terms" in the history of philosophy[1]. The nature of the most general Forms, the related problem of the knowledge thereof, their connection to existing (concrete) objects, are the essential features of the Platonic theory of knowledge and of his metaphysics. The Platonic approach is grounded on the principle of conceptual priority of Ideas over their partecipations, the Forms existing separeted from all the particulars: the former are interpreted as *standard* particulars to which other particulars conform. Nonetheless, my investigation will start rather by Aristotle, who held first the relation of predication to be the basis for defining abstraction: from this I will try to consider some important ideas for the notion of abstraction in Type Systems.

By rejecting the Platonic understanding of general Forms, Aristotle maintains that the logical relation of predication is the starting point for any account of the Categories: these are intended as the forms of both what there is, and of what can be said[2]. Consequently, to speak about the existence of a certain category means properly to make a predication about a certain substance (as the first category, to which the others refer). To analyse predicative expressions (in order to explain the related predicables) will thus amount to a proper *abstraction procedure*. The Aristotelian understanding of abstract terms is in turn given by two related aspects, expressing the underlying distinction between abstract methods and abstract terms:

---

* This research has been made possible thanks to a post-doc research fellowship founded by the University of Palermo (Italy) and spent at the Faculty of Philosophy at the Universiteit Leiden (the Netherlands).

[1] Different passages in the dialogues can be referred to for his theory of abstraction, among them: *Phaedo*, 100a–101e; *Theaetetus* 201d–202c; *Republic*, 476a, 596a; and the entire discussion in the *Sophist*.

[2] Aristotle, *Categories*, par.2.

- the notion of *universal* (καθὸλυ);
- the notion of object produced by (a method of) abstraction (τὰ εξ αφαίρεσεος λεγόμενα, εν αφαίρεσει, δί αφαίρεσεος, αφαίρειν).

Starting by the classical form of judgement "P belongs to S" (with P and S used as schematic letters respectively for predicate and subject), one says that P *belongs universally* to S if:

- the predicate P belongs to every element S;
- that P belongs to S is due to S itself in such that it is an S (in virtue of S and qua S), i.e. not by accidens.

A universal is therefore identified with a predicable satisfying the two previous conditions[3]. On the other hand, the idea of (a method of) abstraction (by which a certain abstract term is produced) corresponds to a *removal operation*[4]. This idea is formulated in the general logical context of the *Analytics* in terms of the operation of removing particular predicates, namely those not falling under the previously given description of the universal. The removal operation preserves only the definitional predicates for the subject (its defining categories), and this corresponds to proceeding from the particular to the more general of the categories[5]. This procedure amounts to concepts formation by abstraction, in terms of the classification of the properties belonging to objects or entities, thus providing their hierarchy of universality. The determination of universals in terms of the particulars defines moreover the peculiar aspect of demonstration[6].

The distinction between formal procedures of abstraction and abstract entities (concepts) seems thus to be already developed in the Aristotelian logic and metaphysics, where "concept" must be intended as the abstract entity resulting from progressive universalization of predications. In the following I will consider further the relation between abstraction and concepts, to show the development of the logical notions involved.

2. Universality and meaning

The relation between general names and related particulars, first questioned by Plato at length for example in the *Parmenides*, is one of the greatest heredities the Middle Age received from antiquity. Porphyry in his *Isagoge*, or introduction

[3] For this explanation cf. e.g. *Metaphysics B*, 4, 1000a1; Γ, 9, 1017b 35; *Z*, 13, 1038b11–13; *Posterior Analytics* I, 4, 73b 26–74a 3.

[4] The standard example is given in the *Physics*, in terms of perceptibles defined as physical magnitudes, something which by nature can be added or removed. Cf. Aristotle, *Physics*, Book 3.

[5] See e.g. *Posterior Analytics*, I, 18, 81b 3–7 and *Metaphysics*, M 2, 1077b 10.

[6] See Aristotle, *Posterior Analytics*, I, 11, 77a 5–9.

to Aristotle's *Categories*, considers the problem of universals explicitely. The nature of Platonic Forms (or of Aristotelian categories) became then a ground problem for later philosophers in the Middle Age, among them Boethius in his *Commentaries on Porphyry's Introduction* discussed it, and Abelard maintained that Ideas preexist the creation as patterns which determine divine Providence in creating the best of the possible worlds (a thesis known as exemplarism).

The theory of universals, which is clearly connected to the nature of abstract entities and therefore to abstraction, was later deeply influenced by the semantic theory of *suppositio*. One of the most relevant interpretations was the thereon based notion of *generality* developed by Ockham, and the derived theory of abstract entities. The theory of supposition furnishes notoriously a semantic treatment for the properties of terms in a sentence: it actually consists in determining the context of semantic validity of a categorematic term in a proposition. Ockham uses the powerful theory of supposition in connection with the theory of universals, by considering the relation between categorematic and syncategorematic terms (i.e. respectively the counterparts of Aristotelian categories and of modern logical constants), in order to answer the question whether universal terms have proper signification. He maintains that a term which supposits generally in a proposition (i.e. under the specification of the syncategorematic "all"), supposits for every term contained in the appellative domain determined by the general noun. In paragraphs 6–8 of his *Summa Logicae*[7], he states that each of two names which are respectively the abstract and the concrete of the same concept (e.g. humanity-man, animality-animal, hotness-heat) are not synonymous, i.e. they do not stay in relation of supposition for the same thing, and what is predicated of one of them cannot be also predicated of the other one. This amounts, in turn, to explicate the problem of predication holding for general abstract and concrete terms in relation to meaning, and identity of meaning is established in terms of their definitions, therefore appealing to their *supposita simplex*. Thus according to Ockham *"every universal is one particular thing and [...] is not a universal except in its signification, in its signifying many things"*[8]. According to this quotation (which expresses Ockham's nominalism), the referents of terms render the distinction between universal and particulars, whereas it exists a common *suppositum*, a meaning determined as an affection of the soul. This is also confirmed by the thesis that a concrete term, being a predicate in a proposition, supposits for a form, as "white" for "whiteness" in "Socrates is white"[9]. The interesting thesis held by Ockham is therefore that the term in the universal form introduces the context of semantic validity of any

[7] Ockham (1349), these paragraphs are titled respectively: "On concrete and abstract names that are synonyms", "The correct account of abstract and concrete names" and "On the third mode of concrete and abstract names".

[8] Ockham (1349, § 14, p. 78)

[9] Ockham (1349, § 63, p. 189).

case of predication in which the same name is used in the concrete form: this means also that universals are applicable to concrete things insofar the latter resemble each other and the concept resembles each of them.

According to Aristotle the nature of abstract terms was obtained by a procedure of removal from the concrete predications, a position which finds many variants in Averroes, Aquinas, Scotus. By introducing explicitely the semantic relation of supposition in this context, Ockham describes an abstract term as allowing those concrete predications to be formulated, by displaying the context of semantic validity in which they can be performed, and thus working as their logical presupposition. The two approaches differ in the understanding of the conceptual priority of concepts and the hierarchy of predications.

3. From functions to types

After this debate, here exemplified by the theories of Aristotle and Ockham, an essential change is provided by the new Fregean paradigm: on the one hand, it preserves the essential role of predication in the definition of concepts; on the other hand, it provides a completely new (logical) form for such a notion, notoriously in terms of functions. The most relevant consequence of this theoretical shift is the connection to the problem of impredicativity and the development of the hierarchy of predications, which will lead to the invention of logical types.

The Fregean approach is a direct critique of the Aristotelian understanding of abstraction as the determination of a unity among many separeted enitites. Frege presents abstraction in relation to the notion of function in the *Begriffschrift*, in terms of the so-called *Abstraction Principle*: if in an expression a simple or compound sign has one or more occurrences and that sign is recognised as replaceable in all of its occurrences by some other sign, the invariant part is then a function and the replaceable part is its argument[10]. On the basis of this general principle, Frege develops the related notion of concept: a concept is not the result of a removal operation in the Aristotelian sense, rather it is the reference (*Bedeutung*) of a predicative expression[11]. This is obtained by explaining a predicate as an unsaturated object, whereas an argument for it is the instantiation of a concept saturating the former[12]. The classical judgemental form "S is P/P belongs to S" is thus changed to a new functional structure $F(x)$, where F plays the role of the predicate, to be evaluated for a certain object (the variable

---

[10] Frege (1879, § 9).

[11] Frege (1892). Notoriously, he explains moreover the reference of singular terms as the objects they stand for, and that of indicative sentences as their courses of values. See also his (1892b, p.193). Cf. Primiero (2004). .

[12] Frege (1891, p. 6).

representing a place-holder for it). To build up a judgement means therefore essentially to *evaluate* a concept for an argument, and the distinction between concept and object determines the former as an abstract term. According to Frege therefore, a concept cannot play the role of the reference of the grammatical subject: it has to be converted into (or represented by) an object, which allows for its evaluation. According to this theory of *concepts as predicates* it is not even necessary for the predicate to be logically possible; the existence of an object instantiating a self-contradictory predicate is obviously a completely different matter.

The formal definition by abstraction of the old-fashioned notion of concept, establishing a class of given objects which satisfy an equivalence relation *R* such that reflexivity, identity and transitivity hold, can be provided also for functional expressions, for which one has to specify their course of values only by means of terms for which they can be evaluated: this makes the notion of function corresponding to its extension, namely the correlation of its arguments and its values[13]. A concept intended as an abstract term (represented by a function) has a range corresponding to the usual logical extension, i.e. the set of all objects falling under it. In *Funktion und Objekt* Frege avoids the essential paradox of the *Grundgesetze* by considering a first-level and a second-level form of abstraction, producing different kinds of functions[14]. Frege had essentially obtained the same result of the later Russellian Ramified Type Theory (RTT)[15], but treating *Wertverläufe* as ordinary objects, he allows a function to be applied on its very same course-of-values (corresponding to a function applied to its own graph), thus he cannot avoid the possibility of impredicative definitions. In the development of the notion of function for RTT, based on the distinction between the hierarchy of types of propositions and that of propositional functions, an important conceptual change occurs: the connection between predication and concepts as abstract entities is forgotten, and it is only partially recovered in terms of the notion of function as the stable part of the abstraction procedure. Correspondingly, the abstraction procedure in the formation of propositional functions is the basis of RTT, where it holds the double hierarchy of simple types and of orders, which also allows to abstract by universal quantification on all propositional functions of a certain order. The thesis that abstraction has now a rather different meaning is confirmed by the understanding of generality for functions as interpreted by Russell. In RTT, this property either means the assertion of any value of a propositional function, or the assertion that the

---

[13] The latter is then an ordered pair, corresponding to the notion of function as a graph; conceptually different is the simple dependent object considered before. In the next section, the notion of function involved in the analysis of type systems will be yet a different one. The clarification of the distinction of these three notions is due to B. G. Sundholm.

[14] Frege (1891, pp. 26–27).

[15] Russell (1908).

function is always true: in the first case, one refers to real variables (any value of the function is asserted); in the second, to apparent variables (the function is always true). The notion of abstraction which leads to the function as independent object (and in turn to contradiction whenever the appropriate hierarchy of predications is not considered) acts on the values (real variables), what Russell called a proper *propositional function*[16]. Thus the type of functions does not only depend on the type of arguments, rather also on the type of apparent variables (place-holders)[17].

Typing procedures introduced by Russell are not just the solution to the problem of impredicativity: they present a new interpretation of the notion of function and a different approach to abstraction. In particular, abstract terms require to be explained on the basis of the logical notion of type, as objects of an higher level.

4. Abstraction from type-free to typed λ-calculus

The traditional interpretation of the notion of function due to Leibniz, Bernoulli and Euler, is that of an analytic expression in one or more variables. Frege and Russell formulated the logical notion based on the relation of predication, which refers to substitution and evaluation as its defining operations. In the Fregean interpretation a function stands also by itself as an independent object of individual type, defined by the correlation to its course of values. By the Russellian theory of types, functions as formal structures of predication are admissible on the basis of the order of their objects. This evolution led to a different model of function and to the notion of type: it restores the old-fashioned notion of function as rule rather than as graph, i.e. it consists of an operation from an argument to a value.

The formalization of abstraction in terms of functional expressions à la Russell substituting the notion of Fregean concepts, where real variables take the place of the abstraction procedure, is further exemplified in the simple typed λ-calculus developed by Church (1940). It combines the Russellian calculus and the operation of deramification, which removes all the orders on types. In a type-free structure the objects of study are both functions and arguments; the alphabet of such a calculus is formed by λ-terms, which are formal expressions for functions and for applications of functions. In this kind of calculus, actually

[16] Russell (1908, p. 157).

[17] A condition which notoriously Russell restricted by formulating the Axiom of Reducibility (AR): for each formula f there is a formula g with a predicative type such that f and g are logically equivalent, where a type is predicative if none of its objects is of a higher order than the order of the elements of the class to which this object should belong.

everything is or is meant to represent a function, based on a composed process of *abstraction* and *application*[18]. Their combination is essential to the formulation of functions, and application is in fact the main operation, whereas abstraction is complementary. The system of operations is completed by *reduction*, consisting in the process of computing from a λ-abstracted term to its value. The model of abstraction here at hand is of a different kind than the Fregean notion of function: the result of abstraction is performed by an operator, and it produces a function rather than a predicate or a concept. Consider the numerical expression 2 + 3, and its transformation into a function, by which one takes into account first the λ-term $(\lambda x.x + 3)2$ which is the β-expansion of the given numerical expression: in this transformation we have a certain argument (2) which is substituted by the argument-variable (x) via the *lambda*-abstractor. What is peculiar in this operation is that one has already the abstracted term (which in turn performs the role of an abstractor operator on values) without having necessarily the starting term from which one abstracts. The β-expansion in the λ-calculus is thus joined to the operation of removing an argument and it corresponds to the function construction. Respectively, instantiation of the latter corresponds to application plus β-reduction of the former[19]. The conceptual identity between the predicative part and the argument is relevant to the understanding of the notion of abstraction involved: the very same expression can perform the role of the operator and of its object. In this sense no term represents the result of an abstraction procedure, nor the context-determining term of predication (as in the case of the relation of supposition). The abstractor operators are such that they can be applied to functions without considering the order of progressively higher types: the abstraction is a pure operation, not a complete process and functions are first-class citizens. The resulting notion of function for these calculi is therefore considered in terms of evalutation (function values), rather than in terms of objects (abstract function).

The typed version of λ-calculi affects then the simple version in an essential way: every term of the calculus has now a normal form, i.e. every possible β-and η-reductions terminate, which makes the set of typable λ-terms entirely recursive. The introduction of types in order to describe the functional behaviour of the terms is relevant in two ways: first, it transforms the idea of abstraction connected to these calculi; second, it provides a bridge between the notion of function and the one of type. On the basis of the Brouwer-Heyting-Kolmogorov interpretation of propositions-as-proofs, in the typed version of this calculus a proof of an implication is a construction, and accordingly a construction of an implication is a function (from the proof of the antecedent to the proof of the consequent). The intrepretation of all operations in terms of their types, due to

---

[18] See e.g. Laan (1997, pp. 4–5).

[19] Laan (1997, p. 43).

the Curry-Howard isomorphism, leads to a different consideration of abstraction procedures. A useful way to explain this, which holds for all kind of typed systems, is to consider the *information*-content of the expressions: the explicit formulation in the syntax of constructions and bound variables in the environments represents the formal structure for which operations of abstraction can be defined[20]. Typed systems can in fact be intended as such that types for all variables and terms are fixed, and expressions contain *full type information* (whereas a so-called type-assignment system would not have such a full information in the basic syntax). The procedure which allow to transform a fully built term into a "core one", i.e. one which provide only the necessary type information, can be seen as an operation of erasing the domains of abstraction; conversely, the typability of terms consists in filling in a proof-trace with the missing elements. According to this relation between informational content of terms and procedures of abstraction, one needs now to distinguish between two different uses of "abstraction": abstraction as $\lambda a.M$ consists in a proof by generalization; an abstract type, will be instead the term obtained by means of an existential type. This distinction is of the greatest importance to understand the nature of abstraction for type systems: it is clearly based on the logical nature of types and on the essential connection which seems to hold between abstraction and information.

5. Other examples of typed systems

Other examples of abstraction procedures in typed systems show interesting properties connected to the explanations provided in the previous sections.

A *prototype proof*, whose notion was first formulated by Herbrand[21], is the proof of a universally quantified statement, whose verification is applicable to each specific instance of the quantified variable. It is executed by assuming a certain generic element of the set the quantification ranges over, in this way making the proof independent from that specific element. This shows the formulation of a *model* to be applied in different cases[22]. A second case in which the same notion of model is clearly involved in connection to abstraction is that of *abstract data types* (ADT). ADT are defined as a set of data values (*abstract data structure*) and associated operations (*interface*) that are considered inde-

[20] This results clearly in the different formulations known as Curry-style and Church-style typed λ-calculi. See e.g. Sørensen, Urzyczyn (2006, Ch. 3).

[21] See Longo (2000, p. 2).

[22] Longo (2000) provides a nice interpretation of the notion of prototypic proof in type systems under the propositions-as-types interpretation, where one can consider this kind of proofs as λ-terms: this is done by considering a simple type called *generic*, i.e. such that it can be assumed as a variable, and whose proof is provided by a specific instance called "parametric", i.e. which can be uniformly substituted.

pendently from any specific implementation. These data represent the result of an operation of abstraction intended as the process of deleting unnecessary details from necessary characteristics (i.e. again to formulate a model by removing information) in order to solve a problem (i.e. to provide the correct operations on a certain set of data). Clearly, the process of obtaining the relevant data is exemplified by the *abstract predicates* entering into the solution of the problem; as in the case of the application function, this process of abstraction is never taken separately from the determination of the operations which are to be performed on the empty schema. To mention a last example: by *polytypic abstraction* one understands instead formalizations and verifications abstracted in respect to a large class of datatypes, which is especially relevant in functional programming. A simple and nice example is that of the function *map* in the Hindler-Miller type system (map: $\forall A,B.(A \to B) \to (list(A) \to list(B))$) which provides a structure of transformation of data lists in other kind of data, with an untouched schema or model, irrelevant to the kind of data instantiated as object of that function[23]. Also in this case we are treating with a procedure of abstraction by which an empty model is obtained, able to implement all the different data of a certain range of (distinct) equivalent types, and to be used in terms of application[24].

6. Types, abstraction and information

The connection between types and information in the light of the abstraction procedures can be reconsidered under the syntactic-semantic method of Martin-Löf's Type Theory. The relevant notion of type in such a system is related to abstraction both from the philosophical and the purely formal viewpoints: its interpretation provides also interesting comments on the analysis done up to now. The main argument is that the syntactic procedures (rule-based operations) are not the unique way to account for abstraction: on the one hand one has to consider the removal operation by which the notion of emtpy (polymorphic) model is obtained; on the other hand, the notion of abstract (meaning-determining and predicative-component) object is involved by the definition of types themselves. In both cases, the notion of information plays a key-role.

At the syntactic predicative level, the notion of abstraction is satisfied in terms of rules. Abstraction and application rules concern the informative con-

[23] See Pfeifer, Ruess (1998).

[24] This obviously requires that the basic distinction between monomorphic and polymorphic languages holds; its better application is given by the so called *parametric polymorphism*, according to which the same object or function can be used uniformly in different type contexts without changes (provided all data are represented). See Cardelli, Wegner (1985, p. 477).

tent of expressions in terms of the specific constructions. The rule of Π-introduction defines an independent object of the lowest individual type

$$\frac{\begin{array}{c}[x : A]\\ b(x) : B(x)\end{array}}{\lambda((x)b(x)) : (\Pi x : A)B(x)}$$

with the related Π-elimination or Application rule; on the other hand, abstraction consists in functions formation of the higher type, i.e. if x is a variable of type A and b is a term of type B, then $(x)b$ is a term of type $A \rightarrow B$:

$$\frac{\begin{array}{c}[x : A]\\ b : B\end{array}}{(x)b : A \rightarrow B}$$

explained by the ordinary β-rule, expressing what does it mean to apply an abstraction to an object in A:

$$\frac{a : A \quad b : B[x : A]}{((x)b)(a) = b[x/a] : B[x/a]}$$

The distinction between universalization and abstraction on contents is therefore clearly stated in terms of syntactic operations. The object of abstraction is here the informative content of judgements, as stated by the *Forget-restore Principle*[25]: the principle says that to build up an abstract concept from a raw flow of data, one must disregard some information, and an abstraction is constructive when the information forgotten can be restored at will. Under this principle, abstraction corresponds to an operation of forgetting from irrelevant computational information, whereas instantiation is the restoring of such information. In particular, by a procedure of abstraction one obtains the transition from the monomorphic to the polymorphic versions of the theory.

Operational abstraction leads to consider higher types as abstract terms themselves. By insisting on the procedure of abstraction in terms of removing the informational content of the constructions, one formulates the judgement declaring the truth of the involved types[26]:

$$\frac{a : A}{A \; true}$$

[25] Valentini (1998).

[26] This formulation simply expresses the propositions-as-sets interpretation, see Nordström, Petersson, Smith (1990, p. 37).

A multi-level typed λ-calculus can be provided for rigorous treatment of judgements of the form "*A* true", on the basis of canonical expressions of the form $a : A$. In this sense, abstraction procedures allow for type-expressions of the form *A type*, provided that abstraction applies as follows:

$$\frac{A \quad set(/prop)}{A \; type}$$

This procedure leads to the analysis of types as independent objects of predication; they are presuppositions for judgements in which those types are used[27]. The explanation of this abstraction procedure is given accordingly to the syntactic-semantic method of Martin-Löf's Type Theory, in which types come conceptually before their objects. This means that the conceptual order between types and their instantiation goes from the former to the latter (i.e. types are abstract terms in respect to their constructions), whereas in the order of knowledge one proceeds from existential predications to their types by means of an abstraction procedure. Higher types, i.e. of the monomorphic kind, should therefore be explained as abstract terms, in connection to predication and semantic context: they recover essential features of abstraction lost in the functional interpretation.

Giuseppe Primiero
Centre for Logic and Philosophy of Science
Universiteit Gent
Blandijnberg 2
B-9000 Gent, Belgium
Giuseppe.Primiero@UGent.be

References

Aristotle. *Works*. The Loeb Classical Library.

Cardelli, L., & Wegner, P. (1985). On Understanding Types, Data Abstraction and Polymorphism. *Computing Surveys, 17* (4), 471–522.

Church, A. (1940). A formulation of the simple theory of types. *The Journal of Symbolic Logic, 5*, 56–68.

[27] See Primiero (forth).

Frege, G. (1891). Funktion und Begriff. Lecture given on 1891, at Jenaischen Gesellschaft für Medizin und Naturwissenschaft. In M. Textor (Ed.) (2002). *Funktion – Begriff – Bedeutung* (pp. 2–22). Vandenhoeck & Ruprecht.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, NF 100, 25–50. In M. Textor (Ed.) (2002). *Funktion – Begriff – Bedeutung* (pp. 23–46). Vandenhoeck & Ruprecht.

Frege, G. (1892b). Über Begriff und Gegenstand. *Vrijschrift für wissenschaftliche Philosophie, 16*, 192–205. In M. Textor (Ed.) (2002). *Funktion – Begriff – Bedeutung* (pp. 47–60). Vandenhoeck & Ruprecht.

Laan, T. (1997). *The Evolution of Type Theory in Logic and Mathematics*. PhD Dissertation Thesis. Technische Universiteit Eindhoven. Enschede: Print Partners Ipskamp.

Longo, G. (2000). Prototype Proofs in Type Theory. *Mathematical Logic Quarterly, 46* (3).

Martin-Löf, P. (1993). *Philosophical Aspects of Intuitionistic Type Theory*. Unpublished notes of lectures given at the Faculteit Wijsbegeerte Leiden.

Nordström, B., & Petersson, K., & Smith, J. (1990). *Programming in Martin-Löf 's Type Theory – An Introduction*. Oxford: Clarendon Press.

Ockham, W. (1349). *Summa Logicae – Part One: Logic of Terms*. Translated and introduced by M. J. Loux. Notre Dame/London: University of Notre Dame Press.

Pfeifer, H., & Ruess, H. (1998). Polytipic Abstraction in Type Theory. *Informal Proceedings of Workshop on Generic Programming (WGP98)*. Marstrand, Sweden.

Primiero, G. (2004). The Determination of Reference in a Constructive Setting. *Giornale di Metafisica, 26* (3), 483–502.

Primiero, G. (forth). Presuppositions, Assumptions, Premises. Forthcoming.

Russell, B. (1908). Mathematical Logic as based on the Theory of Types. In J. van Heijenoort (Ed.) (1999). *From Frege to Gödel – A source book in mathematical logic, 1879–1931* (pp. 150–182). ToExcel Press.

Sørensen, M. H., & Urzyczyn, P. (2006). *Lectures on the Curry-Howard Isomorphism*. Studies in Logic and the Foundations of Mathematics, vol. 149. Amsterdam: Elsevier.

Valentini, S. (1998). The forget-restore principle: a paradigmatic example. In Sambin, Smith (Eds.), *Twenty-five years of Construcitve Type Theory* (pp. 275–283). Oxford: Clarendon Press.

Realizing Possibilities: *Prior Analytics I.15**

Adriane Rini

There is a famous old puzzle about the following piece of logic:

Every *B* is *A*
Every *C* is possibly-*B*
Every *C* is possibly-*A*

In the jargon of Aristotelian syllogistic, this is known as Barbara XQM.[1] According to Aristotle, Barbara XQM is a *valid* syllogism – a valid deductive schema. But of course in modern logic Barbara XQM is *invalid*. The following is a counter-example:

(1)	Everything in the paddock is a horse	T
(2)	Every man could be in the paddock	T
(3)	Every man could be a horse	F

What makes Barbara XQM especially interesting is that Aristotle gives us a formal proof to establish its *validity*, and then, straight after this, he follows the proof with a counter-example to illustrate its *invalidity*. *Prior Analytics* I.15 contains both Aristotle's proof of its validity and his proof of its invalidity.

This is the kind of puzzle scholars like to cite as evidence that Aristotle's logic is really not very good: 'Sure, Aristotle *invented* logic, but he really isn't

* This research is supported by the Marsden Fund Council from Government funding, administered by the Royal Society of New Zealand.

[1] The name Barbara is a mnemonic devised by medieval scholars to encode the premise/conclusion pairs in Aristotle's logic. The 'B' in Barbara indicates what Aristotle calls a first-figure syllogism. The three occurrences of the letter 'a' indicate that the two premises and the conclusion are universal and affirmative. The letters XQM are a modern way of indicating the 'modality' of propositions. X indicates that the first premise is non-modal, or, as it is sometimes called, an 'assertoric' proposition. Q indicates that the second premise is about one kind of possibility. Aristotle uses one word for two different senses of possibility, which he takes pains to carefully distinguish for us. The kind of possibility in a Q proposition is ordinary contingency, i.e., the Q premise is about what is neither necessary nor impossible. M indicates that the conclusion is about Aristotle's other sense of possibility – i.e., what is not necessarily not the case.

particularly *good* at logic – especially where modals like necessity and possibility are involved.' Unless we are happy to say he is a bad logician, then we need a good account of Barbara XQM. Scholars have certainly tried to explain Barbara XQM. Recent work includes: Tredennick (1938), Ross (1949), Hintikka (1973), Smith (1989), Thom (1995), Patterson (1996), and my own Rini (2003). But no interpretation has so far proved to be entirely successful. What makes Barbara XQM important is that Aristotle treats it as an axiom in his system of modal syllogistic – that is, he uses Barbara XQM to establish the validity of several other syllogisms that involve possibility. So, clearly, we need an explanation.

Let's first look at the formal proof Aristotle gives to establish the validity of Barbara XQM. It is a *reductio* proof, and it depends upon a principle stipulated in *An.Pr.* I.15:

34a25–27 ...when something false but not impossible is assumed, then what results through that assumption will also be false but not impossible.[2]

I have numbered the steps in his proof:

34a34–b2 Now, with these determinations made, (4) let *A* belong to every *B* and (5) let it be possible for *B* to belong to every *C*. Then (6) it is necessary for it to be possible for *A* to belong to every *C*. (7) For let it not be possible, and (8) put *B* as belonging to every *C* (this is false although not impossible). Therefore, if (7) it is not possible for *A* to belong to every *C* and (8) *B* belongs to every *C*, then (9) it will not be possible for *A* to belong to every *B* (for a deduction comes about through the third figure). But it was assumed that it is possible for *A* to belong to every *B*. Therefore, it is necessary for it to be possible for *A* to belong to every *C* (for when something false but not impossible was supposed, the result is impossible).

Barbara XQM is (4)(5)(6), and when we replace the term variables *A*, *B*, and *C*, with 'horse', 'in the paddock', and 'man', then we get the counter-example that we started with (1)(2)(3):

(4)	Every *B* is *A*	(1)	Everything in the paddock is a horse
(5)	Every *C* is possibly-*B*	(2)	Every man could be in the paddock
(6)	Every *C* is possibly-*A*	(3)	Every man could be a horse

[2] Throughout this discussion I use Robin Smith's (1989) translation of the *Prior Analytics*.

Suppose

(7) Some *C* is not possibly-*A*	Some man could not be a horse
(8) Every *C* is *B*	Every man is in the paddock

Then

(9) Some *B* is not possibly-*A*	Something in the paddock could not be a horse

Step (7) is the reductio hypothesis. Step (8) is *the realization of the possibility* in (5). That is, in the move from step (5) to step (8) we are in effect actualizing the possibility which is described in step (5). The example using terms makes this easy to see: We go from the premise about possibility – 'every man *could be* in the paddock' – to supposing the possibility is actual, i.e., to supposing 'every man *is* in the paddock.' We are relying on the principle set out, at 34a25–27, about assuming 'something false but not impossible'. We know that assuming that a possibility is actual might lead to falsity, but it will not lead to impossibility – the truth of premise (5) guarantees that our assumption is at least possible.

(9) and (4), however, cannot both be true. So we have a reductio proof to show that (4)(5)(6) is valid. But of course it isn't valid, and the problem is easy to see when we put terms in place of variables, as in the right-hand examples above. In moving from (5) to (8) we are realizing or actualizing a possibility. In supposing that (5)/(2) is actualized, Aristotle seems to have forgotten that that changes the truth value of the initial premise (4)/(1). In supposing that (8) every man *is* in the paddock, we are *denying* that (4)/(1) everything in the paddock is a horse. Lindsay Judson (1983) calls this the 'insulated realization manoeuvre' or *IR*-manoeuvre. Judson finds Aristotle making this same mistake in *De Caelo* I.12. And, as Judson makes abundantly clear, it is a terrible error for Aristotle, or for any logician, to make. But Aristotle himself is worried. He gives his own counter-examples to Barbara XQM. They come in a famous passage:

34b7–18 One must take 'belonging to every' without limiting it with respect to time, e.g., 'now' or 'at this time', but rather without qualification. For it is also by means of these sorts of premises that we produce deductions, since there will not be a deduction if the premise is taken as holding only at a moment. For perhaps nothing prevents man from belonging to everything in motion at some time (for example, if nothing else should be moving), and it is possible for moving to belong to every horse, but yet it is not possible for man to belong to any horse. Next, let the first term be animal, the middle term moving, the last term man. The premises will be in the same relationship, then, but the conclusion will be necessary not possible (for a man is of necessity an animal). It is evident, then, that the universal should be taken as holding without qualification, and not as determined with respect to time.

Consider the effect of the first terms Aristotle gives. When we follow his instructions we get:

All moving things are men
<u>All horses are possibly moving</u>
All horses are possibly men.

Clearly, this is a counter-example and Aristotle means it to be a counter-example. It works just the same as our (1)(2)(3).

Interpreters offer all sorts of solutions. Some want to excise the passage.[3] Some say Aristotle's discussion and this counter-example in particular indicate an explicit rejection of Barbara XQM – i.e., Aristotle, here, gives us proof that Barbara XQM is invalid and not strictly a syllogism.[4] Some scholars offer principles of formal logic to try to make sense of the passage.[5]

Barbara XQM is probably the most important of all of Aristotle's syllogisms about possibility because of the role it plays as an axiom in his system. But we all come unstuck here: Interpreters trying to make sense of Aristotle's discussion seem to need to pull rabbits out of hats to make it work. The puzzle about Aristotle's treatment of Barbara XQM comes from realizing a possibility. So let's consider some of what Aristotle has to say about possibility. The following passages are from *Prior Analytics* I.13, the emphasis added is mine:

32a19 I use the expressions 'to be possible' and 'what is possible' in application to something if it is not necessary but nothing impossible will result if it is put as being the case (for it is only equivocally that we say that what is necessary is possible).

32b4 Having made these distinctions, let us next explain that 'to be possible' has two meanings.

32b5 One meaning is what happens *for the most part and falls short of*

---

[3] For example, Patterson (1996). See pages 166-176, and especially page 174.

[4] See Tredennick (1938).

[5] See Mignucci (1972) and Hintikka (1973). Both attribute to Aristotle the following principle: If it is possible that *p*, then at some time it is the case that *p*. This is known as the *principle of plenitude*, and it has the effect of turning a modal premise into a temporal premise. In Rini (2003) I try to show some of what happens when we try to put the principle of plenitude to work in the modal syllogistic.

necessity, as for a man to turn gray or grow or shrink, or *in general what is natural to belong* (for this does not have continuous necessity because a man does not always exist; however, when there is a man, it is either of necessity or for the most part).

32b11 The other meaning is *the indefinite*, which is *capable of being thus as well as not thus*, as, for instance, for an animal to walk or for there to be an earthquake while one is walking, or, *in general, what comes about by chance* (for it is no more natural for this to happen in one way than in the opposite).

32b14 Now, each of these kinds of possible premises also converts in relation to its opposite premise, but not, however, in the same way. A premise concerning what is natural converts because it does not belong of necessity (for it is in this way that it is possible for a man not to turn gray), whereas a premise concerning what is indefinite converts because it is no more this way than that.

32b18 Science and demonstrative deduction are not possible concerning *indefinite things* because the middle term is disorderly; they are possible concerning *what is natural*, however, *and arguments and inquiries would likely be about what is possible in this sense*. A deduction might possibly arise about the former, but it is, at any rate, not usually an object of inquiry.

These passages contain crucial clues to help interpret Aristotle's discussion of Barbara XQM. Notice first that the possibility described in 32a19 is clearly contingency – the logician's $Q\phi =_{df} \sim L\phi \wedge \sim L\sim\phi$. Notice also the distinction between the two ways of being possible: the first way is by being natural (32b5); the second way is by being indefinite or chance (32b11). But there are no syllogisms about chance (32b18). *Science* is about *natural* possibilities. Aristotle's natural possibilities are natural *capacities* or *potencies*. For example, it is *natural* for an acorn to become an oak tree – an acorn is in this sense a 'possible oak'. It is a Q-oak.

Now consider:

(10)	Every oak is a deciduous tree	T
(11)	Every acorn could be an oak	T
(12)	Every acorn could be a deciduous tree	T

(10)(11)(12) has the form of Barbara XQM. (11) is a premise about natural possibility – an acorn could become an oak. When we realize the possibility in premise (11), we do not have the problems we had in the earlier example (1)(2)(3). When we actualize the second premise ‘every *C* is possibly-*B*’, we get ‘every *C* is *B*’. But look closely at what this actualizing does to the *first* premise – that is, to the BA premise. In (1)(2)(3), when *B* is ‘in the paddock’ – i.e., when *B* is indefinite, an accidental term – then premise (1) *only happens to be true at some moment*. In (10)(11)(12), *B* names something with an essence – *B* is ‘oak’. In (10)(11)(12) when the acorn’s potential is actualized, then in premise (10) we have a necessary truth about an essential subject, and the truth value of such a premise is not susceptible to change. Realizing the possibility in premise (11) cannot alter a *necessary truth*. But realizing the possibility in premise (2) *does* alter the value of premise (1) – and this is precisely the mistake Judson calls the *IR*-manoeuvre. (10)(11)(12), however, is not susceptible to the *IR*- manoeuvre. We cannot change the value of a necessary proposition by changing the contingent facts.

If that is right, then why has Barbara XQM been a problem for so long? The answer would seem to have a lot to do with modern views about logic. We have tended to approach Aristotle’s logic as though it is a formal logic, where validity is determined purely by form. When we approach Aristotle’s logic that way, then we cannot capture his distinction between different kinds of syllogistic terms. Some terms name accidents – things that are merely by chance or that are indefinite. Some terms name things that have essential natures. Which kind of term we use affects our ability to syllogize, as *An.Pr.* 32b18 makes clear. So we cannot answer the puzzle posed by Barbara XQM with purely formal principles. Aristotle’s logic is a term logic – in the modal syllogistic validity is restricted to certain choices of terms.

School of History, Philosophy and Classics
Massey University
Private Bag 11222
Palmerston North
New Zealand
A.Rini@massey.ac.nz

References

Hintikka, J. (1973). *Time and Necessity: Studies in Aristlotle's Theory of Modality*. Oxford: Clarendon Press.

Judson, L. (1983). Eternity and Necessity in *de Caelo* I.12. *Oxford Studies in Ancient Philosophy*, *1*, 217–255.

Mignucci, M. (1972). *On a Controversial Demonstration of Aristotle's Modal Syllogistic: An Enquiry on Prior Analytics A.15*. Padova: Ed. Antenore.

Patterson, R. (1996). *Aristotle's Modal Logic: Essence and Entailment in the Organon*. New York: Cambridge University Press.

Rini, A. (2003). When Time is of the Essence: *Prior Analytics* I.15 and *de Caelo* I.12. *Logique et Analyse, 183-184*, 419–440.

Ross, W. D. (1949). *Aristotle's Prior and Posterior Analytics. A revised text with Introduction and Commentary*. Oxford: Clarendon Press.

Smith, R. (1989). *Aristotle's Prior Analytics.* Indianapolis: Hackett.

Thom, P. (1995). *The Logic of Essentialism: An Interpretation of Aristotle's Modal Syllogistic*. Dordrecht: Kluwer.

Tredennick, H. (1938). *The Organon I: The Categories, On Interpretation, Prior Analytics*. [Translation of *Prior Analytics* is by Tredennick.] Cambridge, Mass.: Harvard University Press (Loeb Classical Library).

Bolzano on Judgement and Error

Maria van der Schaar

1. The problem of error

Error has a place in modern philosophy at most as a principle of selection for theories, in accordance with a naturalist, Darwinian paradigm (Mach, Popper). Inconsistencies and defects may show that an error has been made, but what error is, is left unexplained. Within logic and philosophy of language, the problem of error has been reduced to questions concerning false sentences. But the notion of false sentence differs from that of error. By pronouncing a false sentence, one may lie, suppose, or give an example without erring. If one does not judge the false sentence to be true, one does not err. Judgement is essential to error. Recently, within Martin-Löf's Constructive Type Theory, the notions of judgement and error have gained some importance.

The problem of error is the problem how incorrect judgement is possible. The problem of error thus presupposes an answer to the question what judgement is. We may distinguish two types of answer to the latter question:

I. One may stress the parallel between judgement and knowledge.
II. One may stress the parallel between correct and incorrect judgement.

In accordance with this distinction, two types of answer to the problem of error can be distinguished:

I'. What explains error is not an independent force. Error is a privation; the incorrect judgement is not as a judgement should be.
II'. Error or incorrect judgement is explained in complete parallel to the explanation of correct judgement. Correct and incorrect judgement are explained, for example, in terms of the proposition which forms the content of the judgement. If the proposition is true, the judgement is correct; if the proposition is false, the judgement is incorrect. Propositions are true or false, just as some roses are red, and others are white, as Russell once said (Russell 1904, p. 75).

Answers of type I' have the advantage that they explain the asymmetry between correct and incorrect judgement. Correct judgement is related to what is (Reality); incorrect judgement to what is not (Appearance). But if one stresses the asymmetry too much, error becomes impossible. As Socrates says in the *Theaetetus* (189a), to judge what is not is to judge nothing, which is not to judge at all.

The advantage of answers of type II' is that error is clearly made possible; it is as real as correct judgement. The disadvantage is that the asymmetry between correct and incorrect judgement threatens to disappear. Such a theory has to answer the question: what is it that makes some propositions true, which is absent in the case of false propositions?

Keeler (1934) ends his history of the problem of error with Kant, and Balduin Schwarz, in his article on 'Irrtum' in the *Historisches Wörterbuch der Philosophie,* only mentions 'the important analysis' of error given by Bolzano. In the less known third part of the *Wissenschaftslehre* (1837), the 'Erkenntnislehre', there are several chapters on judgement, knowledge and truth, with a special section on error. Besides the logical / conceptual question how error is possible, Bolzano also asks the epistemological / psychological question what the causes of error are, how error arises in us.

With respect to the concept of error, one has to distinguish between act and product. 'Error' and the German term 'Irrtum' stand for the product, resulting from an act of erring ('das Irren'). The distinction is a special case of the distinction between the act of judgement and the judgement product. Both act and product need to be distinguished from the proposition, which Bolzano also calls an error, if it is false but held true.

Because Bolzano explains error primarily as incorrect judgement (WL, I, § 36), the question what judgement is comes first (section 2). To understand the concept of error, one also needs to understand what knowledge is (section 3). In my analysis of Bolzano's notions of judgement and knowledge I have profited from Mark Siebel's two recent articles on these topics (Siebel, 1999 and 2004). In section 4 Bolzano's concept of error will be dealt with.

2. Bolzano on judgement and belief

In modern philosophy the notion of judgement has been replaced by that of belief, which replacement creates at least two problems. Judgement is an act, whereas belief is a state, and the replacement has thus led to a neglect of the act of judging. Bolzano, though, explains the notions of opinion, knowledge, error and inference in terms of the act of judging, and rightly so, I think. The other problem is that the term 'belief' is highly ambiguous. It may stand for: *opinion*, the state of mind in which one holds a proposition to be true; the degree of *confidence* with which one holds a proposition to be true; or *faith*, that is, trust in God, a person or a doctrine. Bolzano was fully conscious of the distinction between these notions, introduced an appropriate terminology, and also gives an explanation of *holding true a proposition*.

In the first part of the *Wissenschaftlehre* (WL, I, § 34), Bolzano takes the act of judgement ('eine Handlung unseres Geistes') to be what is common to assertion (as act), opinion and faith. An act of judgement has a *Satz an sich*, a proposition, as its matter (*Stoff*; the term '*Inhalt*' has a different meaning in Bolzano's work; for systematic reasons I will use the term 'content' instead of 'matter'), which is independent of the act of judging, of language and of space and time. In contrast to a *Satz an sich*, the act of judgement is dependent upon the judging mind, in space and time, and it may stand in causal relations, because it is real (*wirklich*). According to Bolzano, the *Satz an sich* is not the result of an *act of setzen*. The proposition is not to be identified with the judgement product, for the latter is dependent upon the act of judging (WL, I, § 20, p. 82).

Bolzano opposes a theory of judgement of type I, namely Kant's theory, in which judgement is explained in terms of knowledge: "Das Urtheil ist die mittelbare *Erkenntnis* eines Gegenstandes" (KdrV, A 68; cf. WL, I, § 35, see further Schaar 2003). According to Bolzano, such an explanation applies only to correct judgements; besides, it contains a circle, because knowledge, according to Bolzano, has to be explained in terms of judgement. According to Bolzano, a judgement is correct, if the proposition that is its content is true; if the proposition is false, the judgement is incorrect. Bolzano's theory of judgement is thus of type II: the explanation of the incorrect judgement is parallel to that of the correct judgement. Unlike Russell (1904), Bolzano gives an account of what makes a proposition true: "our judgements are true if we connect with our presentation of a certain object the presentation of such a property [the object] really (*wirklich*) has." (WL, I, § 42, p. 180). One may add, the judgement or proposition is false, if the object does not have the relevant property. Bolzano is thus able to give an explanation of the asymmetry between truth and falsity. This explanation of error is preliminary to the more extended explanation given in the third part of the *Wissenschaftlehre*.

All notions for which we now use the term 'belief' indifferently, Bolzano explains in terms of the act of judgement. We *hold true* ('are consistently committed to', *sind fortdauernd zugethan*) a certain proposition M, consisting of a subject-presentation S and a predicate-presentation P, if we judge that S is P as often as the presentations S and P, or the question whether S is P, come to our mind (WL, III, § 306, p. 200). In order to exclude cases of holding true for vacuous reasons (when the presentations S and P never come to our mind and we never judge that S is P), we need to add a clause. Mark Siebel (1999) proposes to add that one has to have passed the judgement that S is P at least once. This clause is in accordance with the explanation of holding true, for Bolzano presupposes in the explanation that we remember a judgement as one that we have passed before (§ 306, p. 200), and it is also in accordance with Bolzano's explanation given in § 307 (p. 208). The extra clause forces one to say that someone who can calculate does not hold true that 2156 > 1, if he has never passed the judgement 2156 > 1. According to Siebel (1999, p. 71), who follows Robert Audi in this, in such a case one has the capacity to judge ('Disposition zum Erwerb einer Überzeugung',) that 2156 > 1, but one does not hold true ('die Überzeugung haben') that 2156 > 1. The state of holding true has a certain duration: one holds true the proposition M as long as one stands in the above-mentioned relation to its subject-presentation S and predicate-presentation P.

An *opinion* (*Meinung*) of someone is, according to Bolzano, a *Satz an sich* which that person holds true, whether that proposition be true or false, and whether the degree of confidence (*Zuversicht*) be weak or strong (§ 306, p. 200).

The degree of *confidence* pertaining to a judgment made is to be distinguished from its liveliness, that the judgement owes to the liveliness of the presentations that form its parts. We may have a strong degree of confidence in religious truths without having a lively presentation of them (§ 293, pp. 112, 113). If one practically denies the possibility of the opposite of one's judgement, the degree of confidence is called conviction (*Überzeugung*, § 319). The degree of confidence with which one judges is determined by the probability of the proposition, which will be explained below.

We will see in the next section that the concept of justification plays no role in Bolzano's explanation of the concept of knowledge. What is important for him, though, is the question how judgements arise in us, which is preliminary to the question how error arises in us.

Judgements arise by mediation of other judgements or they arise immediately (§ 300). Certain judgements of perception and certain pure conceptual judgements count as immediate judgements. Mediated judgements, or judgements caused by other judgements, are the result of an act of the mind through which one goes over (*übergeht*) from the judgements A, B, C, D to the judgement M, which act is called *an inference* (*ein Schliessen*, *ein Folgern,* § 300, p. 123). The relation of the judgement M to the judgements A to D is a relation of mediation (*Vermittlung*), when no other judgement than A to D is needed in order to judge M. In such a case one says that the judgements A to D are the complete cause of the judgement M (This does not mean that if I have made the judgements A to D, I also will judge M, for there has to be a certain activity of the mind).

According to Bolzano, there are exactly three ways in which a judgement may be mediated by other judgements (§ 300, p. 126). The three relations between these judgements are defined in terms of objective relations between S*ätze an sich*, the contents of any judgement. I will use [A] for the content of the judgement A.

a) The contents of the judgements A to D may form the objective *ground* of the content of M. The relation between the *Sätze an sich* [A] to [D] and [M], in case the former is the objective ground of the latter, Bolzano calls a relation of *Abfolge*, which is always a relation between true *Sätze an sich* (cf. WL, II, § 198). The ground of a true proposition forms the reason why that proposition is true. The proposition that the temperature rises is a ground of the truth that the thermometer rises, but not the other way round.

b) The content of judgement M is deducible *(ableitbar)* with respect to a substitutible part, a presentation (*Vorstellung an sich*), from the contents of the judgements A to D, if every substitution of that presentation in each of the propositions by another presentation that makes the propositions [A] to [D] true, also makes the proposition [M] true. For example, from the proposition [Cajus is a man] the proposition [Cajus is mortal] is deducible with respect to the presentation [Cajus] (WL, II, § 155, p. 120), because every substitution of the presentation [Cajus] in both propositions by another presentation that makes the former proposition true, also makes the latter true. But if we also take the presentation [man] as substitutible part, the relation of deducibility between the two propositions, with respect to the presentations [Cajus] and [man], does not obtain, for [God is almighty] is true, whereas [God is mortal] is false.

c) The content of M has a degree of probability > ½ (*Wahrscheinlichkeit*) relative to the contents of the judgement A to D (with respect to certain substitutable parts). The degree of probability can be determined by taking the number of cases in which one goes over through substitution from true propositions [A] to [D] to a true proposition [M] and divide it by the number of cases in which one obtains through substitution true propositions [A] to [D] (WL, II, 161, p. 172). Deducibility is a special case of the relation of probability, namely when the degree of probability is maximal, that is, 1. As we will see in the last section, the relation of probability is of special importance for the question how error may arise in us.

According to Bolzano, there are only these three ways in which a judgement may arise in us. In those cases where M does not follow from A to D, the judging person has tacitly added as premis the incorrect judgement that from the propositions [A] to [D] one may infer the proposition [M] (WL, III, 300, pp. 129, 130).

The degree of confidence of a judgement M in relation to the judgements A to D is determined by the probability of the proposition [M] in relation to the propositions [A] to [D]. When the probability of a certain proposition is 1, because it may be the content of an immediate judgement or because it may be inferred from such a judgement, the degree of confidence of judgement M is 1, which means that it is *perfect*. If the degree of probability of a proposition [M] in relation to [A] to [D] is μ, then the degree of confidence with which one judges (or would judge) M is 2μ — 1. The degree of confidence in M may be negative or 0, for example, when the probability of [M] is ½ , in which cases we do not judge M (cf. WL, III, §§ 318–320).

3. Bolzano on knowledge

The German language makes a distinction between *Erkenntnis* and *Wissen*. *Erkenntnis* is the product of an act of *Erkennen*, that is, cognition is the product of an act of cognizing. Bolzano's term 'Wissen' may be translated as 'certain knowledge', because of the cognate term 'Gewiss' ('certain'), or it may be translated as 'scientific knowledge', because of the cognate term 'Wissenschaft'. According to Bolzano (WL, I, § 36), cognition is a judgement that contains a true *Satz an sich*. Cognition as product of an act of cognizing does not exist independently of that act, according to Bolzano (WL, III, § 307). In part III of the *Wissenschaftslehre* Bolzano considers the explanation given in part I too narrow: anyone who is at the present moment not judging a certain truth, cannot be attributed cognition according to the explanation given in part I. Cognition that S, Bolzano says in part III, is a state of the mind pertaining to a person P, if (§ 307, p. 207; cf. Siebel, 1999, p. 77):

(a) P has once passed the judgement J;
(b) J has a true proposition as its content;
(c) P is able to remember the judgement J;
(d) P still holds true ('is consistently committed to') the corresponding proposition.

In comparison to the modern explanation of knowledge as justified true belief, there are two important differences. Bolzano does not have a notion of justification as part of the explanation of cognition, and he holds that there is no state of cognition unless it is preceded by an act of judging. In contrast to Siebel, I judge the latter difference an advantage over the modern explanation. Bolzano understands that an explanation of cognition needs to account for the *obtainment* of knowledge. Bolzano has a concept comparable to that of justification, namely that of *cognitive ground* (*Erkenntnisgrund*). The judgement that the thermometer rises is the cognitive ground of the judgement that the temperature rises, if we know that the temperature rises because we have passed the judgement that the thermometer rises. Cognitive grounds are, in contrast to objective or proper grounds, judgements, and they need not be true. If the cognitive ground does not have the objective ground of the conclusion as its content, Bolzano says, it is merely a *subjective* cognitive ground of the conclusion; if it does, it is called its *objective* cognitive ground (§ 313).

The concept of cognitive ground is not part of Bolzano's explanation of cognition, probably because he considers that concept to be too psychological. Neither is the concept of objective ground part of the explanation of cognition. Bolzano explicitly says that a cognition may be the result of pure luck, and that it may be mediated by false judgements (§ 314, pp. 232, 233). No Gettier problems for Bolzano's notion of cognition. Neither does Bolzano use the concept of objective ground to explain the notion of scientific or certain knowledge (*Wissen*). Certain knowledge is a state of the mind with respect to its own judgements. We have certain knowledge of a truth [M], when the confidence which pertains to the judgement M appears to us in such a way that we are not able now to destroy it (§ 321). Bolzano does have a name, though, for those true judgements of which we also know the objective ground: these acts of cognition are called *understanding* (*Begreifen oder Einsehen*), and the corresponding pieces of knowledge are called *clear insights* (*deutliche Einsichten*) (§ 316, pp. 259, 260).

There is an important reason why the notion of objective ground does not form part of Bolzano's explanation of certain or scientific knowledge. This reason can be found in the section on the Kantian question whether there are any limits to our faculty of cognition (§ 314 and 315). If there were such limits, there would be unknowable truths, Bolzano says, but how can we know that there will not be a man in the future that knows such a truth? According to Kant, metaphysics, which deals with questions concerning God, soul, immortality and freedom, is not a science. There is not one truth concerning these topics that is undisputed and not doubtful, and these questions lie beyond the limits of knowledge. According to Bolzano, the propositions that there is a God, that God is immutable, and that no simple substance perishes through time, are part of metaphysics. Although people have disputed what the proper ground is of those propositions, no one doubts that they are true (§ 315, p. 249). Although metaphysics is not perfect regarding the scientific order of its truths, and although we are not able to give the first grounds of this science, that does not imply that we cannot have certain, scientific knowledge concerning the objects of metaphysics. Even the most perfect science such as mathematics is wanting in its first grounds. Which book on geometry is able to give an explanation of the concepts of space, line or body? If the notion of objective ground would be part of the explanation of scientific knowledge, no science would pass the test.

4. Bolzano on error

As we have seen in the first section, Bolzano explains error in the first part of the *Wissenschaftslehre* as incorrect judgement, and an incorrect judgement is a judgement whose content is a false *Satz an sich*. In the third part of the *Wissenschaftslehre*, Bolzano improves upon his explanation of error: An error is every false proposition that a person holds true (§ 307, p. 208). Making a comparison with the explanation of cognition given in the former section: point (a) also holds for error, see (a') below; (b') and (d') have 'false' where (b) and (d) have 'true', and (c) (we are able to remember judgement J) is missing in the explanation of error. Essential to error is thus:

(a’) a person P once has passed the judgement J;
(b’) J has a false proposition as its content;
(d’) P still holds true that false proposition.

There is an important difference between error and cognition: whereas cognition is a mental state, an error is a false proposition held true. Bolzano makes it clear, though, that the false judgement products may also be called errors (§ 307, p. 207). Although Bolzano does not use the verb ‘das Irren’, one may add that we err in the corresponding acts of judgement. Without an act of judgement (a’) one does not obtain the corresponding state of mind of holding true (d’). According to Bolzano, error is possible because there are false *Sätze an sich*, and because we may judge false propositions to be true.

How is it possible that we mistake a false proposition for a true one? According to Bolzano, we cannot err regarding immediate judgements. His argument is that if we doubt one, we have to doubt them all, and everything that is derived from them, because they all arise in a similar way, which is not a particular good argument, for we should doubt disputed immediate judgements, whereas there is no reason to doubt the others. We can also not err, Bolzano says, regarding judgements that are derived from immediate ones, and whose propositions are related by probability 1. If the premises are true, an inference brings us to a new truth (§ 309, p. 212). And the same may be said concerning a judgement M in relation to the judgements A to D, when the propositions [A] to [D] form the objective ground of [M] (§ 301). We may at most say that our faculty of inference (that is, our faculty of cognition) is limited, so that we do not know every true proposition that follows from known premises (§ 308, p. 211). Again, how does error arise in us? What is the *Irrtumsquelle*?

It cannot be a separate faculty besides the faculty of cognition that makes us err. According to Locke, the faculty of knowledge is infallible. He is thus in need of a separate faculty of judgement to explain error. But, according to Bolzano, it cannot be the case that we have a cognitive faculty that God is deprived of: God is not wanting in any sense (§ 301, p. 137). Neither is it man’s will that makes him err, as Descartes thought (§ 310, p. 220). Dependent upon the will is at most our attention, which directs us to certain presentations, but not the judgement itself (§ 291, p. 110). Neither is it true, according to Bolzano, that to err is a form of sinning, as it is thought within the Augustinian tradition, or that it is a form of precipitation (overhaste, § 310, p. 223). The astronomer, who after careful calculations predicts a moon-eclips in a hundred years on a certain day, whereas the eclips holds off because a comet passes by, makes an error, but there is no precipitation.

According to Bolzano, it is the limitation of our faculty of cognition (that is, the faculty of judgement) in which we differ from an omniscient being. In finite beings, the faculty of judgement works in judgements with perfect confidence, and in judgements with imperfect (less than perfect) confidence (§ 301, p. 137).

And it is only with respect to judgements with imperfect confidence that we may err.

Which judgements have only imperfect confidence? Certainly not the immediate judgements. At the end of section 2, we have seen that there are only three ways in which a judgement may be mediated by other judgements: when there is a relation of *Abfolge* between the contents of the mediating and the mediated judgement; when there is a relation of deduction between those contents; and when there is a relation of probability between them. Because we cannot err in the former two cases (unless the chain of inferences becomes very long, so that we have to rely on memory, cf. § 309, p. 214), we can err only with respect to judgements of probability. A proposition [M] that has a high degree of probability (with respect to certain propositions, and with respect to a substitutable part of those propositions) will in a real life situation be judged by us as being in fact true, with a degree of less than perfect confidence corresponding to the degree of probability of the proposition. If the proposition is in fact false, we have made an error. The error is not made when the degree of probability is attributed to the proposition [M]. For example, if there are 99 black balls and one white ball in an urn, the probability of the proposition that a black ball will be drawn is 99/100, and we do not err if we judge the proposition to have that probability. The possibility of error arises, as soon as we expect, and thus judge it to be true, that a black ball will be drawn. To err in such a case, Bolzano says, is a (psychological) necessity; we are not free to withhold our judgement (§ 309, p. 213). All our empirical judgements are judgements of probability. The judgement that the sun will rise tomorrow is the result of a probability inference. As soon as we expect and judge that it is true that the sun will rise tomorrow, the *possibility* of error arises, for the content of that judgement is only very probable (in relation to the content of our former judgements, that is, our experience). "Every error is a proposition, which has, in relation to the other propositions that the erring person holds true, a certain probability." (§ 309, p. 214)

The probability that pertains to every error (false proposition) Bolzano calls the appearance (*der Schein*) of that proposition (the German term for probability is *Wahrscheinlichkeit).* The appearance of a proposition may find its origin in propositions that are all true, in which case the error is called *original*, or it may arise from false propositions. To go back to the exampe of the urn with 99 black balls. If we judge it to be true at t_0 that a black ball will be drawn at t_1; and if at t_1 a white ball is drawn, we clearly have made an error (incorrect judgement product). The probability that the false proposition has, with respect to the propositions that there are 99 black balls in the urn and one white ball, is its appearance (§ 309, p. 214). Because of this probability the proposition appears to us to be true, but it is not really true; it is false. If the probability of a proposition, in relation to certain others, is less than ½, we will not err, because we withhold our judgement (although we may err if we judge the negation of that proposition).

5. Conclusion

Concerning the logical question how error is possible, Bolzano's answer is of type II'. There is a parallel between the explanations of cognition and error, compare (a) to (d) with (a'), (b') and (d'). Concerning the psychological / epistemic question how error arises in us, Bolzano's answer is of a different type. Error arises in finite beings because their faculty of judgement is limited, which means that they have to rely on probability judgements. In order to live, we have to judge [M], which is only probable in relation to the contents of our former judgements, to be true without qualification, with a degree of confidence in proportion to the probability the proposition has in relation to the contents of our other judgements; the degree of confidence is thus less than perfect.

Bolzano's primary answer concerning the problem of error focuses on the question what the *content* is of an incorrect judgement. Error in its objective aspect is a false proposition. Bolzano's answer is thus of type II'. Bolzano's other answer focuses on the question how an incorrect judgemental *act* may arise in us. Error in its subjective aspect is due to our limited faculty of judgement, which means that we often have to judge with a less than perfect degree of confidence. This answer contains elements of type I'; error arises from privation. Finally, the most interesting aspect of Bolzano's answer combines both objective and subjective elements. Probability, which is an objective property of *Sätze an sich*, is essential to the psychological question how error arises in us, because a proposition with a greater degree of probability *appears* to us to be true without qualification.

Faculty of Philosophy
Leiden University
Postbus 9515
2300 RA Leiden
m.v.d.schaar@let.leidenuniv.nl

References

Bolzano, B. (1837). *Wissenschaftslehre*, Leipzig, 1929/1930.
Keeler, L. W. (1934). *The Problem of Error from Plato to Kant*, Rome.

Russell, B. (1904). Meinong's Theory of Complexes and Assumptions. In B. Russell (1973). *Essays in Analysis* (21–76). Edited by D. Lackey. London: Allen and Unwin.

Schaar, M. van der (2003). Opinion, Assertion and Knowledge: Three Epistemic Modalities. In T. Childers, & O. Majer (Eds.). *The Logica Yearbook 2002* (259–268). Prague: Filosofia.

Siebel, M. (1999). Bolzanos Erkenntnistheorie. In E. Morscher (Ed.). *Bernard Bolzanos geistiges Erbe für das 21. Jahrhundert* (59–96). Sankt Augustin: Academia.

Siebel, M. (2004). Bolzanos Urteilslehre. In *Archiv für Geschichte der Philosophie, 86*, 56–87.

Born Again! Anselm in the Person of Charles Hartshorne[1]

Jordan Howard Sobel

The argument Then and Now

Hartshorne derives,

> "There is a perfect being, or perfection exists,"

from the premises that

> "perfection is not impossible,"

and that,

> "perfection could not exist contingently."
>
> (Hartshorne, 1962, pp. 50-1)

These premises are, on certain assumptions, equivalent to corollaries to which Anselm was committed of the premises of the major argument in *Proslogion* 2.

> "Something-than-which-nothing-greater-can-be-thought exists in the mind."

and

> "That-than-which-a-greater-cannot-be-thought cannot exist in the mind alone [and not also in reality]."
>
> (Charlesworth, 1979, p. 117)

[1] From "Born Again! Anselm and Gaunilo in the Persons of Charles Hartshorne and William Rowe," in progress: http://www.scar.utoronto.ca/~sobel/OnL_T/AnselmBornAgain.pdf. With gratitude to Michael Almeida and Martin Tweedale.

1. Then – *Proslogion* 2 – "*That God truly exists*."

> "[1] Well then, Lord, You who give understanding [*intellectum*] to faith, grant me that I may understand, as much as You see fit, that You exist as we believe You to exist, and that You are what we believe You to be. [2] Now we believe that You are something than which nothing greater can be thought [*aliquid quo nihil maius cogitari possit*]. [3] Or can it be that a thing of such a nature does not exist, since 'the Fool has said in his heart, there is no God'? [*Psalms* 14, l. 1, and 53, l. 1.] [4] But surely, when this same Fool hears what I am speaking about, namely, 'something-than-which-nothing-greater-can-be-thought', he understands what he hears, and what he understands [*intelligit*] is in his mind [*intellectu*], even if he does not understand that it actually exists. [5] For it is one thing for an object to exist in the mind, and another thing to understand that an object actually exists. [6] Thus, when a painter plans beforehand what he is going to execute, he has [it] in his mind, but does not yet think that it actually exists because he has not yet executed it. [7] However, when he has actually painted it, then he both has it in his mind and understands that it exists because he has now made it. [8] Even the Fool, then, is forced to agree that **something-than-which-nothing-greater-can-be-thought exists in the mind**, since he understands this when he hears it, and whatever is understood is in the mind. [9] And surely **that-than-which-a-greater-cannot-be-thought cannot exist in the mind alone [and not also in reality]**. [10] For if it exists solely in the mind even, it can be thought to exist in reality also, which is greater. [Peter Millican puts in place of that, Alexander Broadie's 'translation': 'For if it exists solely in the mind, something that is greater can be thought to exist in reality.' Hopkins and Richardson have in (1974): 'For if it were only in the understanding, it could be thought to exist also in reality – which is greater [than existing only in the understanding].'] [11] If then that-than-which-a-greater-cannot-be-thought exists in the mind alone [and not also in reality], this same that-than-which-a-greater-*cannot*-be-thought is that-than-which-a-greater-*can*-be-thought. [12] But this is obviously impossible. [13] Therefore there is absolutely no doubt that **something-than-which-a-greater-cannot-be-thought exists both in the mind and in reality**." (p. 117, bold emphasis and sentence numbers added.)

Charlesworth does not comment on the hyphenated singular terms that his translation of *Proslogion* 2 features. How do they enter this proof? Anselm says that we believe that God is something than which nothing greater can be thought. That he may understand that God exists as he believes, he proceeds in terms of another 'name' for this person in whom he believes, he proceeds in terms of the descriptive name 'something-than-which-nothing-greater-can-be-thought' [4], and says – I now make the best I can of the single-quotation marks in Charlesworth's translation when this name is introduced – that even the Fool who declares that there is no God, understands these words, this hyphenated term, when Anselm speaks to him using them/it. "He understands," Anselm might have spelled out, *using* these words, "*something-than-which-nothing-greater-can-be-thought*. And," Anselm could have added, "*what* he understands is in his mind as it is in my mind." One may gather that Anselm did not need the Fool for his argument which in this part could have been conducted as a Cartesian soliloquy.

2. Detailing the argument

Proslogion 2 features two subsidiary arguments that deliver premises emphasized in sentences [8] and [9] for its major argument, the conclusion of which is drawn from them in sentence [13]. This agrees with the 'take' on its argument with which the able monk Gaunilon begins his response on behalf of the Fool.

> "To one doubting whether there is...something...than which nothing greater can be thought, it is said here...that its existence is proved, **first** because the very one who denies or doubts it already has it in his mind, since when he hears it spoken of he understands what is said; and **further**, because what he understands [this-something-than-which-nothing-greater-can-be-thought] is necessarily such that it exists not only in the mind but also in reality." (*Pro Insipiente* I: p. 157, bold emphasis and bracketed material added.)

2.1 Subsidiary argument [4] through [8].

This is an argument for

> *Premise* I. *Something-than-which-nothing-greater-can-be-thought exists in the mind.*

The Fool understands of what Anselm, with the term 'something-than-which-nothing-greater-can-be thought', speaks. He therefore has not only these words (this term/this indefinite description) in his mind, but this that they (it) designate(s) in his mind. So it is 'in a mind' or 'in the mind'. The curious passage from 'a' to 'the' is unremarked.

***2.2** Subsidiary argument [9] through [12]**.*

2.2.1 This is an argument for

Premise II. *Something-than-which-nothing-greater-can-be-thought* ***cannot*** *exist in the mind alone* (and not also in reality).

or, in other words, for

> *It is not possible that* something-than-which-nothing-greater-can-be-thought *exists* in the mind alone
> (and not also in reality).

which is equivalent to,

> *It is necessary that* it is not the case that
> something-than-which-nothing-greater-can-be-thought exists in the mind alone
> (and not also in reality).

To show this it is sufficient to derive from only necessities that,

> It is not the case that something-than-which-nothing-greater-can-be-thought exists in the mind alone (and not also in reality).

Now comes a derivation for this negation. It is an *indirect* derivation for which we suppose that,

> **(i)**, Something-than-which-a-greater-cannot-be-thought exists in the mind alone (and not also in reality).
> M(A)
> [M: *a* exists in the mind alone (and not also in reality);
> A: something-than-which-a-greater-cannot-be-thought]

It is, however, necessary that:

> **(ii)**, For any kind of thing, a thing of this kind that exists not only in the mind
> but in reality as well is greater than a thing of this kind that exists in the mind alone.

That it is necessary that existence in reality is an 'other-things-equal-greater-making' condition is a plainly implicit premise of *Proslogion* 2.

Therefore, from (i) and (ii),

> "[S]omething that is greater [than something-than-which-a-greater-cannot-be-thought] can be thought" (Broadie's translation),

or in other equivalent words,

> **(iii), Something-than-which-a-greater-cannot-be-thought**
> **is something than which a greater *can* be thought to exist in reality**.

and

> There is something such that it is greater than something-than-which-a-greater-cannot-be-thought,
> and it can be thought to exist in reality.
> **(∃x)[G(xA) & Tx]**
> [A: something-than-which-a-greater-cannot-be-thought; G:
> *a* is greater than *b*; T: *a* can be thought to exist in reality]

How so? Because we can think of something x such that, x is of exactly the same kind as this something-than-which-a-greater-cannot-be-thought, and x exists not only in the mind but in reality as well.

However (now comes a line that is only implicit in Anselm's text),

> **(iv)**, This something-than-which-a-greater-cannot-be-thought
> is something than which a greater *cannot* be thought to exist in reality.

or equivalently

> It is not the case that there is something such that it is greater than this something-than-which-a-greater-cannot-be-thought, and it can be thought to exist in reality.
> **~(∃x)[G(xA) & Tx]**.

The redeployment of the name letter 'A' serves (as usual for repeated terms) to symbolize the explicitly anaphoric phrase '*this* something-than-which-a-greater-cannot-be-thought'.[2] The emphasized contradictory lines – please see their em-

[2] The valid argument – I was talking with John. You met (this) John last week. ∴ I was talking with someone you met last week. – is symbolized (with the quantifier confined to persons) by – M(A) . T(A) ∴ (∃x)(Tx & Mx) – under the scheme – A: John; T: I was talking with *a*; M: you met *a* last week. But not the invalid argument – I was talking with John. You met (a) John last week. ∴ I was talking with someone you met last week. This argument is symbolized by – M(A) . T(B) ∴ (∃x)(Tx & Mx) – under that scheme

phasized symbolizations for their intended interpretations and the contradiction – complete the indirect derivation. According to them, to adapt Anselm's words in [11]: 'This same something-than-which-nothing-greater-can-be-thought, is, (iii), something than which a greater *can* be thought to exist in reality, and, (iv), it is something than which a greater *cannot* be thought to exist in reality. But this is obviously impossible.'

2.2.2 How did the boldly emphasized statement, (iv), of 'self-predication' get into this subsidiary derivation? Perhaps Anselm would say that (iv) is itself necessarily true, and that, *in general*, of any 'a-such-and-so' it *is* a such and so. I propose that Anselm considered (iv) to be a consequence of (i) in which its indefinite description occurs. My suggestion is that his reasoning proceeded in an *unarticulated logic* for indefinite descriptions in which such inferences are all but immediate and can easily go unremarked. It is a very simple and intuitive logic, it is I think *the* intuitive logic, for indefinite descriptions. It can be reached by adding indefinite descriptive terms to a standard quantifier calculus for nonempty domains and denoting terms. For this logic, we may add to the language of the Quantifier Calculus (Kalish, *et. al*, 1980), for variable α, and formula ϕ in which α has a free occurrence, the indefinite description term $\ulcorner @\alpha\phi \urcorner$ in which '@' is a variable-binding operator — literal translation, $\ulcorner$an-α-such-that-$\phi\urcorner$, and add to its deductive system the premiseless inference rule or axiom:

> **Indefinite Descriptions**. For variables α, and formulas ϕ and ψ, and formula $\psi_{@\alpha\phi}$ that comes from ψ by proper substitution of $\ulcorner @\alpha\phi \urcorner$ for α,
> $\psi_{@\alpha\phi} \equiv (\exists\alpha)(\phi \mathbin{\&} \psi)$.
> an-α-such-that-ϕ is an α such that ψ if and only if an α is such that both ϕ and ψ

For example: 'G@xFx ≡ (∃x)(Fx & Gx)' – 'an-x-such-that-Fx is an x such that Gx if and only if an* x is such that both Fx and Gx'. [*Here 'an' has the sense not of 'any' but of 'at least one'.]

Statement (i) of our derivation has in the language of this logic the simple symbolization 'M@xSx' under the abbreviations – M: *a* exists in the mind alone (i.e., *a* exists in the mind, but not in reality); S: *a* is something than which

---

augmented by the abbreviation – B: John. Proper names when repeated in speech or conversation are 'by default' for the same thing or person with '(this)' and '(that)' being understood without statement. Not so for repeated indefinite descriptions which are 'by default' for possibly different things with '(a)' being understood without statement unless 'explicitly overwritten' by 'this' or 'that' as in Charlesworth's translation of *Proslogion* 2.

a greater cannot be thought to exist in reality. Statement (iii) has under this scheme the symbolization '~S@xSx', and statement (iv) has the symbolization 'S@xSx'. The explicitly anaphoric '*this* something-than-which-a-greater-cannot-be-thought' of (iv) is symbolized here by the same symbolic indefinite description's being used in symbolizations of these sentences, as it was in Section 2.2.1 by the same name letter's being used (please see note 2). Sentence (iv), thus symbolized, has the following derivation in the logic for indefinite descriptions just detailed from sentence (i), thus symbolized.

1.	~~*SHOW*~~ (iv) S@xSx	Direct Derivation (8)
2.	M@xSx	(i)
3.	M@xSx ≡ (∃x)(Sx & Mx)	Indefinite Descriptions
4.	(∃x)(Sx & Mx)	3, Biconditional Conditional [left to right], 2, *Modus Ponens*
5.	Sa & Ma	4, Existential Instantiation
6.	Sa & Sa	5, Simplification, Repetition, Adjunction
7.	(∃x)(Sx & Sx)	6, Existential Generalization
8.	S@xSx(∃x) ≡ (∃x)(Sx & Sx)	Indefinite Descriptions
9.	S@xSx	8, Biconditional Conditional [right to left], 7, *Modus Ponens*

2.2.3 The subsidiary derivation in Section 2.2.1 states what I take to have been Anselm's reasoning for *Premise* II. A crucial juncture of the reasoning, namely, the entry into it of line (iv), can be spelled out in a simple and intuitive logic for indefinite descriptions which I take to be Anselm's unstated way with them. There is, however, not a small problem here. This logic is fatally flawed. It is an *inconsistent* logic in which contradictions are derivable! For example, there is in this logic an indirect derivation for,

F@x(Fx & ~Fx) & ~F@x(Fx & ~Fx)

that turns on the case of Indefinite Descriptions,

~[(F@x(Fx & ~Fx) & ~F@x(Fx & ~Fx)] ≡
(∃x)[(Fx & ~Fx) & ~(Fx & ~Fx)].

In this case of Indefinite Descriptions, α is 'x', ϕ is '(Fx & ~Fx)', ψ is ~(Fx & ~Fx), and '~[(F@x(Fx & ~Fx) & ~F@x(Fx & ~Fx)]' is $\psi_{@x\phi}$. It can be seen that '~[(F@x(Fx & ~Fx) & ~F@x(Fx & ~Fx)]' comes from '~(Fx & ~Fx)' by proper substitution of '@x(Fx & ~Fx)' for 'x'.

1.	~~SHOW~~ F@x(Fx & ~Fx) & ~F@x(Fx & ~Fx)	(6, 7, Indirect Derivation)
2.	~[F@x(Fx & ~Fx) & ~F@x(Fx & ~Fx)]	Assumption for Indirect Derivation
3.	~[(F@x(Fx & ~Fx) & ~F@x(Fx & ~Fx)] ≡ (∃x)[(Fx & ~Fx) & ~(Fx & ~Fx)]	Indefinite Descriptions
4.	(∃x)[(Fx & ~Fx) & ~(Fx & ~Fx)]	3, Biconditional Conditional [left to right], 2, *Modus Ponens*
5.	(Fa & ~Fa) & ~(Fa & ~Fa)	4, Existential Instantiation
6.	Fa	5, Simplification, Simplification
7.	~Fa	5, Simplification, Simplification

2.2.4 *A more generous construction of Anselm's reasoning*. Suppose I am right about Anselm's implicit logic for indefinite descriptions. Then, had he articulated it and noticed the fatal flaw of it, he could have found a cure that left his argument for *Premise II* intact. This on the assumption that he would say that the range of his quantifiers was exactly *things that exist in the mind*.

2.2.4.1 *The cure*. Taking into account this range for his quantifiers, Anselm could have seen that the rule Indefinite Descriptions needs to be premised. He could have, (a), revised Indefinite Descriptions to,

Indefinite Descriptions*. $(\exists\beta)\ \beta = @\alpha\phi\ /\therefore\ \psi_{@\alpha\phi} \equiv (\exists\alpha)(\phi\ \&\ \psi)$,

α and β variables, ϕ and ψ formulas, and $\psi_{@\alpha\phi}$ a formula that comes from ψ by proper substitution of ⌜@αϕ⌝ for α; and, (b), letting *M* be a logical predicate for existence in the mind, endorsed the rules,

Existence in The Mind. $M\gamma\ /\therefore\ (\exists\beta)\ \beta = \gamma$; $(\exists\beta)\ \beta = \gamma\ /\therefore\ M\gamma$,

γ a term, β and variable. These rules would reflect the intended range of his quantifiers, within which things that exist in reality would make a proper subclass. Letting *R* be a logical predicate for existence in reality, the latter intent could be reflected by the rule,

Existence in Reality. $R\gamma\ /\therefore\ (\exists\beta)\ \beta = \gamma$.

2.2.4.2 Amending his logic in this manner, Anselm could have argued for *Premise* II much as suggested in Sections 2.2.1 and 2.2.2. Using the new logical predicates '*M*' and '*R*', and 'S' to abbreviate '*a* exists in reality' and '*a* is something than which a greater cannot be thought to exist in reality', the supposition for indirect derivation in Section 2.2.1 could be symbolized thus, (i’) $M@xSx \;\&\; {\sim}R@xSx$,

from which (iv) could be derived thus,

1.	~~*SHOW*~~ (iv) $S@xSx$	(8, Direct Derivation)
2.	$M@xSx \;\&\; {\sim}R@xSx$	(i’)
3.	$(\exists y)\; y = @xSx$	2, Simplification, Existence in The Mind
4.	${\sim}R@xSx \equiv (\exists x)(Sx \;\&\; {\sim}Rx)$	3, Existence in The Mind, Indefinite Descriptions*
5.	$(\exists x)(Sx \;\&\; {\sim}Rx)$	4, Biconditional Conditional (left to right), 2, Simplification, *Modus Ponens*
6.	$Sa \;\&\; {\sim}Ra$	5, Existential Instantiation
7.	$Sa \;\&\; Sa$	6, Simplification, Repetition, Adjunction
8.	$(\exists x)(Sx \;\&\; Sx)$	7, Existential Generalization
9.	$S@xSx \equiv (\exists x)(Sx \;\&\; Sx)$	2, Simplification, Existence in The Mind, Indefinite Descriptions*
10.	$S@xSx$	7, Biconditional Conditional (right to left), 6, *Modus Ponens*

Having added Existence to reflect his intent that the domain of his quantifiers should be exactly ‘things that exist in the mind’, Anselm could wish to ‘free-logic’ the rules of existential generalization and instantiation to agree with that intent, so that lines (6) and (8) should be respectively,

(6’) $(\exists x)\; x = a \;\&\; (Sa \;\&\; {\sim}Ra)$ 5, Existential Instantiation*

and

(8’) $(\exists x)(Sx \;\&\; Sx)$ 6’, Simplification ($\exists x\; x = a$), 7, Existential Generalization*

Existential Instantiation*: for variable α, distinct variable β that is novel to the derivation, formula ϕ and formula ϕ_β that comes from ϕ by proper substitution of δ for α,

$$(\exists\alpha)\phi \;/\therefore\; (\exists\alpha)\; \alpha = \beta \;\&\; \phi_\beta$$

Existential Generalization*: for variable α, term δ, formula ϕ_δ, and formula ϕ that comes from ϕ_δ by proper substitution of α for δ,

$$(\exists\alpha)\; \alpha = \delta,\; \phi_\delta \;/\therefore\; (\exists\alpha)\phi$$

2.2.4.3 Anselm could have maintained that that critique of Indefinite Descriptions in Sections 2.2.3 cannot be adapted to run against Indefinite Descriptions*. That *reductio*, readdressed to Indefinite Descriptions*, would need the premise that *M*@x(Fx & ~Fx). This premise he could have said is not available, since only things of which we can speak and think *without a priori contradiction*, as even the Fool can do of that-than-which-nothing-greater-can-be-thought, lie in the domain on his quantifiers and 'exist in the mind' in the sense of '*M*'. The amendments suggested in 2.2.4.1 to what I think was Anselm's defective implicit logic in *Proslogion* 2 save the subsidiary reasoning in it for *Premise* II, and afford a charitable alternative interpretation of its hidden logic that would place the burden of this chapter's argument squarely on its *Premise* I, and direct critical attention to Anselm's reasoning for it. This direction is taken in "Born Again! Anselm and Gaunilo in the Persons of Charles Hartshorne and William Rowe," in progress: http://www.scar.utoronto.ca/~sobel/OnL_T/AnselmBorn Again.pdf .

3. Now – Hartshorne's modal argument

Hartshorne offers a deduction of the existence of a perfect being from two modalized premises. He dubs his first premise 'Anselm's Principle.'

AP $\Box[Q \supset \Box Q]$,

which under the abbreviation – "'Q' for '$(\exists x)Px$' There is a perfect being" (Hartshorne, 1962, p. 50) – symbolizes,

It is necessary that if there is a perfect being, then it is necessary that there is a perfect being.

Hartshorne provides for ***AP*** the free translation, "*perfection could not exist contingently*" (p. 51, italics added), which idea he gets from *Proslogion* 3: there is in Hartshorne's text for this principle ("The Incompatibility of Perfection and Contingency," pp. 58-68) nothing like Anselm's *reductio* for his *Premise* II. Hartshorne's argument runs in terms of an existential generalization. His reasoning, which is conducted in sentential, not in quantified, modal logic, is well clear of the pitfalls and complications of Anselm's descent for purposes of logical calculation to a particular something. That perfection could not exist contingently has the symbolization,

$\sim\Diamond[Q \& \sim\Box Q]$,

which is logically equivalent to '$\Box[Q \supset \Box Q]$' by a modal-negation interchange followed by several interchanges of sentential equivalents. His other premise comes with the comment, "Intuitive postulate (or conclusion from other theistic arguments)" (op. cit., p. 51): it is the proposition that perfection is possible:

IP $\Diamond$Q.

It follows in modal logic S5 from ***AP*** and ***IP*** that there is a perfect being,

Q.

4. This modal argument 'updates' the major argument of Proslogion 2

Hartshorne's two premises are, on three assumptions, 'philosophic translations' of proximate consequences, to which Anselm was committed, of the premises of the major argument of *Proslogion* 2. *Assumption One* is that Hartshorne's words, 'a perfect being', mean the same as Anselm's words, 'a thing than which nothing greater can be thought'. *Assumption Two* is that to say, in Anselm's mentalistic idiom, that there is something of a kind that it *exists in the mind*, is to say in modal terms that something of this kind *is possible*, or equivalently, that *it is possible that* there is something of this kind. *Assumption Three* is that, 'to exist *in reality*' was, for Anselm, 'to exist *simply*'. On these assumptions, consequences to which Anselm was committed of the stated premises of *Proslogion* 2 are equivalent to Hartshorne's premises.

Premise I of the major argument in *Proslogion* 2 is,

> Something-than-which-nothing-greater-can-be-thought *exists in the mind*.

From this, Anselm would need to say it follows, mainly by Indefinite Descriptions (or Indefinite Descriptions*), that:

> There is something of the kind, thing than which nothing greater can be thought, *that exists in the mind*.

Confirmation. (i) *Premise* I: M@xSx. (ii) S@xSx: a theorem given mainly Indefinite Descriptions, or a consequence of (i), rewritten, '*M*@xSx', mainly by Indefinite Descriptions* and Existence in The Mind. Therefore, **(∃x)(Sx & Mx)** by Existential Generalization, or Existential Generalization*.

Therefore, by Assumption Two, *Premise* I has for Anselm the corollary,

> *It is possible* that there is something than which nothing greater can be thought.

And this, according to Assumption One, is equivalent to,

> ***IP*** It is possible that *there is a perfect being*: $\Diamond Q$

Premise II of the major argument of *Proslogion* 2 is,

> Something-than-which-a-greater-cannot-be-thought *cannot* exist in the mind alone (and not also in reality).

Anselm is committed mainly by Indefinite Descriptions (or by Indefinite Descriptions* and *Premise* I) to,

> Something-than-which-a-greater-cannot-be-thought is a thing than which a greater cannot be thought.

So *Premise* II alone (or with assistance from *Premise* I) has for Anselm the consequence,

> Something than which a greater cannot be thought *cannot* exist in the mind alone (and not also in reality).

or equivalently,

> *It is not possible* that (something than which a greater cannot be thought exists in the mind, though no such thing exists in reality).

which, by my three assumptions, is Anselmian speech for,

> It is not possible that (both it is possible that there is a perfect being, and it is not the case that there is a perfect being).

In particular, 'something than which a greater cannot be thought exists in the mind' is, by Assumption One, synonymous with 'a perfect being exists in the mind', which, by Assumption Two, is synonymous with 'it is possible that there is a perfect being'. That in symbols is,

$$\sim\Diamond[\Diamond Q \mathbin{\&} \sim Q]$$

which is equivalent to

> ***AnP*** $\Box[\Diamond Q \supset Q]$,

and thus to Hartshorne's,

$$\textbf{\textit{AP}} \qquad \Box[Q \supset \Box Q].$$[3]

Each of ***AnP*** and ***AP*** is equivalent to the Leibnizian principle that 'if perfection is possible then it is necessary': $\Diamond Q \supset \Box Q$. ***AnP***: '$\Box[\Diamond Q \supset Q] \equiv \Diamond Q \supset \Box Q$' is an instance of the modal confinement theorem '$\Box[\Diamond P \supset Q] \equiv [\Diamond P \supset \Box Q]$'. ***AP***: '$\Box[Q \supset \Box Q] \equiv \Diamond Q \supset \Box Q$' is an instance of the modal confinement theorem '$\Box[P \supset \Box Q] \equiv [\Diamond P \supset \Box Q]$'.

University of Toronto at Scarborough
sobel@UTSC.utoronto.ca
http://www.scar.utoronto.ca/~sobel/

References

St. Anselm (1979). *Proslogion.* Translated with an introduction and philosophical commentary by M. J. Charlesworth, with Latin facing English, University of Notre Dame Press.

Anselm of Canterbury (four volumes), 1974–6, edited and translated by Jasper Hopkins and Herbert Richardson, Toronto and New York: The Edwin Mellen Press.

Hartshorne, Ch. (1962). *The Logic of Perfection.* LaSalle: Open Court.

Kalish, *et al*. (1980). *Logic: Techniques of Formal Reasoning*. Second Edition. New York: Harcourt Brace Jovanovich.

Millican, P. (2004). The One Fatal Flaw in Anselm's Argument. *Mind, 113.*

[3] Hartshorne missed the proximity of the argument *Proslogion* 2 to his own argument. He wrote that "the famous-notorious Chapter II of the *Proslogium*...is...altogether secondary. [Its] paragraphs represent but a preliminary try, and an unsuccessful one – elliptical and misleading at best – to state the essential point [AP, 'that perfection cannot exist contingently'], which is first explicitly formulated in *Proslogium* 3" (Anselm, 1962, p. 2).

What Is Objective Probability?

Pirmin Stekeler-Weithofer

1. Introduction

Mathematicians do not bother too much about the external, world-related, meaning of their theorems and proofs. For them, mathematical truths are analytic.[1] This means that mathematical truth is defined by merely formal conditions for system of sentences or propositions. Mathematical proofs show that the conditions are fulfilled. If we do not only think of axiomatic deductive systems, the conditions of mathematical truth are laid down by definitions of the mathematical domains consisting of theoretical entities like numbers, sets or geometrical forms (represented by singular terms, which often might be understood in a situation-dependent way, for example when we talk about uncountable domains) and theoretical properties like being prime, being a finite set, or being a rectangular triangle (represented by predicates, i.e. open sentences or open propositions).

Under this view, mathematical theory of probability is combinatorial arithmetics. In its more advanced form, it is a branch of real and abstract analysis,[2] which is, as such, higher arithmetics anyway. But when it comes to the logical status of real probability judgments about events in the real world, things are not so easy. The general situation is the same here as when we ask for the external meaning of the formal truths of Euclidean geometry. This question, too, is all too seldom treated in a serious way, since people seem to be content with the usual stories that tell us that 'real space' is non-Euclidean, as Einstein allegedly has taught us. But to say that Euclidean geometry is no true theory of space is not much more than popular nonsense. The right thing to say would run like this: Euclidean geometry is a formalized theory of our generic talk about spatial forms of bodies. Other geometries are developed for modeling spatial *and* chronological relations *between moved* bodies in a kind of *urbild* U – such that the bodies can be said to produce the arithmetical results of our measurements

---

[1] In his famous early paper "probabilism", de Finetti reminds us, too, that the truths of probability theory are analytic.

[2] This can be shown by a sentence like the following: "Über die Wahrheit von Wahrscheinlichkeiten Betrachtungen anzustellen, kann man getrost Philosophen überlassen. Für uns handelt es sich um eine mathematische Theorie, die durch ihre reiche Begriffswelt relevant ist" (Morgenstern, 1968).

m causally. The structure in U is induced by the measurement mappings f_m with U as domain: the values of the f_m are numbers (or quantities), but they come with dimensions, such that the range (image) of m altogether, viewed as a mapping, is at least four-dimensional.

In a similar way we should distinguish between the *internal*, as such analytic, truths of theorems *in* stochastic theories and their *external* significance. And this means, that any immediate objectivism with respect to probabilities is misleading. To show what this objectivism could consist in, I have chosen Gnedenko's introductory book "Theory of Probability". Gnedenko writes (on p. 47):

"The fact that in a number of instances the relative frequency of random events in a large number of trials is almost constant compels us to presume certain laws."

But it is not clear at all what it means to say that we are 'compelled' to assume certain laws and what the logical status of these (stochastic) laws is. We are rather compelled to ask the following questions:

1. What are *random* events (or *random choices)*?
2. What is a *large* number of trials?
3. What does *almost* constant mean?

The problems are well known. But, astonishingly, there are no satisfying answers available, at least if we do not view traditional positions and dogmatic claims as satisfying. In order to show this, I compare the also well-known positions of frequentism (von Mises)[3], objectivism (Gnedenko), and probabibilism (subjectivism or Baysianism: de Finetti, Carnap et al.) with a constructivist approach to probability, as proposed by Paul Lorenzen quite some time ago.

To begin with, it is fairly easy to see why a mere frequentist account of probability is not sufficient. For it is not at all clear what it means to talk about an *infinite limit* of relative frequencies (like 'heads up' in throwing a coin n times). It does not make sense at all to talk about infinite sequences outside of mathematics. If we assume, therefore, as von Mises does, that a limit should be invariant with respect to random choices of infinite subsequences, the empirical and mathematical domains of discourse already are confounded. In fact, frequentism can be characterized as a position that chooses not to ask the central questions at all, which means to remain content with not understanding the difference between a mathematical domain of discourse and our talk about the real empirical world.

As a seeming way out, Gnedenko defends the usual, i.e. Kolmogorov's, *axiomatic definition* for a probability space, and he adds:

---

[3] A frequentist interpretation of probability is, for example, presupposed in the theories of R.A. Fisher and Neyman and Pearson for statistical inferences.

> "...what is of particular interest, is, that in our definition probability retains its objective meaning, one that is independent of the observer" (Gnedenko, 1968, p. 48).

Once again it is easy to *say* such things about 'objective meaning'. But it is not at all clear what this *means* and if it is *true*. Our next questions therefore are:

4. Is there any objective meaning in assuming probability measures?
5. How do we justify our decisions to posit certain probability measures?
6. And how do we justify the way we calculate in formal probability theory, including rational choice theory and game theory, or rather, how we use such calculations in our real decisions and judgments?
7. How can probabilistic models depend on our empirical observations if, as the subjectivist approach suggests, such models depend on mere decisions of the subject and therefore are a priori?
8. What kinds of reasons do we have for making such decisions that go beyond questions of formal consistency and vague coherence?

Gnedenko seems to think that the case of stochastic laws is not much different from that of other laws of nature and that randomness is something like a natural phenomenon:

> "The laws existed before we came to know them."

With respect to mathematical truth, a sentence like this just says that certain conditions, for example formal consistency, *could* be fulfilled, i.e. that a system of rules with certain properties *was possible*, even before the rules were made explicit. In this reading, Gnedenko's existence claim just refers to the *possibility* of constructing a certain theory with certain properties, like, for that matter, Bolzano's or Frege's claims about the objectivity of logical or mathematical laws. Such possibilities exist always, tautologically, before they are realized. But Gnedenko claims much more, namely that the different forms of the laws of large numbers, which can be seen as generalized versions of Bernoulli's theorem, proved by Tchebychev, Markov and others, provide

> "general sufficient conditions for the statistical stability of the mean".

Paul Lorenzen also says that these laws justify the definition of a 'probability value' to generic outcomes of (physical) generators of random processes (in German: *Zufallsgeneratoren*). But what does this mean? And is it true? De Finetti, at least, does not seem to agree to such an interpretation, even less to a claim like the following:

"The vast experience accumulated by mankind teaches us that a phenomenon with probability very close to one is almost certain to take place" (Gnedenko, 1968, p. 31).

The problem to read the theorems on large numbers in the way Gnedenko and Lorenzen read them can be seen easily if we consider their mathematical content more precisely. For the weak law of large numbers in its standard form writes:

(1) $\lim_{n\to\infty} P[\,|\,D/n\text{-}p\,|\, < \varepsilon] = 1$ (for any ε however small)

and the strong law:

(2) $P[D/n \rightarrow p] = 1$.

D/n is explained as the mean $(\Sigma_{1\leq i\leq n} X_i)/n$ for random variables X_i, which are identically distributed, mathematically independent from each other, and have the same mathematical expectation. P is defined as some kind of 'product measure' on the base of the probability measure p, where p is pre-given as a probability measure on some set W_A of subsets of a set A. In general, W_A is a so-called Borel field or algebra over A, whose elements are called "elementary events". The first theorem says that for any two margins of error ε_1 and ε_2 there is an m, such that for all bigger n the P-probability (whatever measure P is) for the 'fact' (whatever kind of fact that is), that the difference between D/n and p is less than ε_1, does not differ from 1 more than ε_2. The formal condition for this theorem is that the trials are 'independent' and that the probability values for any occurrence of an event a_i in any individual trial are equal, namely p = 1/n. The second theorem says: If P is the product measure on the set of infinite series of possible results, the P-probability that the limit of D/n in such a series is p, is 1. Both theorems seem to say in different forms that 'in the long run' the relative frequency of the occurrence of an a_i is almost equal to p. Since the laws of large numbers are analytic, however, they really only say something about the relation between the stochastic evaluations p and P.

This shows that there is a categorical difference between mathematical sentences as *idealized generic sentences* on one side, *empirical and singular sentences* on the other. I certainly cannot answer all questions here in a satisfactory way how *generic sentences* are to be understood. But a first negative answer is this: Generic sentences are not *universal* sentences. They do not say something about all and every case of a certain class of cases in the real world. They express rather *default cases* or *normality conditions* in the form of *general maxims, principles or rules of thumb*.

2. Sample measures

In order to make our investigation as perspicuous as possible, I shall avoid mathematical generalizations as well as the use of technical terminology like "variance", "distribution", or "random variable". I do this by restricting attention to the prototype case, where theorem (1) reduces to Bernoulli's Theorem. There, D just is the number of occurrences of a generic event a_i out of k possible cases $a_1,\ldots,a_k$ in n different trials (like in dicing). In the Bernoulli case of dicing we denote by A or S(1) the sequence of length 1 or set $a_1,\ldots,a_6$ of *possible* results of throwing a die *once*, by S(n) the set of *possible n-sequences s(n)* or results of throwing a die n times. We denote the i'th elements of S(n) by s_i; it is equal to a_j for some j between 1 and 6. The *real outcome* of n throws of a die is denoted by the (situation-bound!) expression $o_S(n)$. By $P_{os}(n)$ or short P_s I refer to the *situation-bound sample measure*. This is defined as a function, whose arguments are subsets C of the set A. The values of P_s are just the *relative frequencies*

$$\| C \cap s \| \ / \ \| s \|) = c_i/n \text{ (with } \| s \| = n)$$

of the occurence of a_i of the type C in s = $o_s(n)$. We can write, then:

$$P_S(C) := c_i/n = \| C \cap s \| \ / \ \| s \|),$$

if one reads this with a grain of salt. It is clear, that any real sample measure fulfills the conditions of a probability measure on the whole power-set $\mathcal{P}$(A)of the finite *generic* set A.

An abstract probability measure P is, mathematically speaking, any function defined on a subalgebra W_A of the power set $\mathcal{P}$(A) of a given set A (finite or not) into the closed interval [0,1] of real numbers between 0 and 1. The basic conditions for P are well known:

1. The P-value of the whole set A is one, i.e. P(A) = 1, the value of the empty set Ø is zero.
2. Finite additivity for disjoint unions: $P(C_1) + P(C_2) = P(C_1 \cup C_2)$ if $C_1 \cap C_2 = \emptyset$, i.e. if C_1 and C_2 are disjoint subsets in W_A.[4]
3. Σ-additivity for infinite unions: $\Sigma_{1\leq i<\infty}P(C_i) = P(\bigcup_{1\leq i<\infty}C_i)$ if all the C_i are disjoint subsets in W_A. Of course we assume that W_A is closed under finite intersections and such unions, i.e. it is a Σ-algebra, if Σ-additivity applies.[5]

[4] The probability that a generic event A does not happen is 1–P(A). The conditioned probability P(A/B) is defined as the quotient $P(A\cap B)/P(B)$. Generic events A and B are stochastically independent iff the probability $P(A\cap B)$ is equal to the product of P(A) and P(B), i.e. iff P(A/B) = P(A).

[5] Infinity conditions like Σ-additivity are not needed if we restrict ourselves to finite sample mea-

Usually, one says that 1-3 are defining axioms for a probability space W_A. I prefer to turn things around and say: If you want to know the (most basic) properties of a probability measure, just consult real sample measures. This has the advantage that the definition of such a measure is no axiomatic stipulation at all. Moreover, we understand probability theory better if we consider genuine probability measures p (on A resp. W_A) as some kind of *thumb rules*, intended for a *general (generic) approximation of 'most' sample measures resulting in such trials*. This way of looking at things has at least the advantage that the choice of the basic axioms for probability measures *is formally justified from the beginning*. Moreover, as generic thumb rules, probability measures on A resp. W_A are neither 'real limits' of sample measures, nor are they some mysterious things as 'objective dispositions' in a world behind our experience, by which we could 'explain' resulting frequencies. Rather, any talk about 'dispositions' has to be explained by our practice of *ascribing* such dispositions to certain phenomena. *This is crucial because the idea that dispositions or tendencies are just there in the empirical world is deepest dogmatic superstition of modern scientism.*

Whereas de Finetti seems to think that attachments of probability values are subjective attachments of singular persons to singular events, I claim that what we really do is this. We propose to attach a probability value p to *generic* cases. By doing so, we say something like this: If we consider *all possible thumb rules or generic expectations*, *the proposed rule or expectation is 'the best' or 'the true one'*. And this does not mean that in *all* cases of predictions or previsions of this kind (whatever kind or class one has in mind), the predicted frequency gets fulfilled in a certain margin of error is near p. It says that *no other generic prediction is better, as far as we can know*.

I claim, moreover, that attachments of probability values to particular, singular, events are to be understood in view of the generic case. What I do not deny, of course, is that a priori probability measures depend on our decisions.

Gnedenko correctly says (on p. 20f) that in contrast to a merely statistical case, which refers to singular events as singular events, the genuine stochastic case refers to *generic events*, which may occur or even be (re)produced repeatedly. Cases like dicing or throwing a coin are in one respect similar to the statis-

---

sures and Bernoulli's theorem. We also do not have to bother about restricting the domain of the measure on Σ-algebras. In the Bernoulli case, sample measures are always defined on the whole power set in question. In fact, the infinity conditions expressed by Σ-additivity and Σ-algebras only get important in geometrical surroundings: There, we have to distinguish between the elementary events or points of measure zero and the events or sets in a Σ-algebra that have positive measures. The obvious reason is that if we attach to intervals of equal size equal probabilities, finite (and in case of Σ-additivity also countable) sets of discrete points get measure zero. This is the very reason why, mathematically, a probability of measure zero (or one) must be distinguished from impossibility (or necessity, for that matter), and why actual and finite sample measures in continuous domains do not tell much about probabilities of 'continuous' point-sets: Any such sample measure, looked at as a finite set, has probability zero.

tical one, in another they are not. We may be content with a 'good prediction' or 'approximation' of the outcome at the end of an individual dicing game or a poll (1). Or we may be interested in a general rule for such predictions in may cases (2). In the first case, the 'objective probability', which should be approximated by our subjective choice of a measure, just is an objective sample measure (post hoc). In the second case we refer to a potential infinitude of trials or sequences of trials. What is approximated in the second case, is not only unclear, there is nothing i.e. no individual thing or object or event, to be approximated. De Finetti says something like this, but he does not give our explanation.

Obviously, there are 6^n different *generic sequences* s(n), if our game allows us to throw the dice n times. What interests us is the number of these sequences s(n) in S(n), for which the (generic, possible!) sample measures have the following property:

(*) $|p(a_i) - P_s(a_i)| < \varepsilon$.

Let us call this number the Bernoulli number $b(n,\varepsilon)$, ε being a pregiven 'margin' or 'tolerance', however small. It now seems *plausible,* to 'measure' the probability that (*) is satisfied by the number

$$p_{n,\varepsilon} := b(n,\varepsilon)/6^n.$$

Bernoulli's theorem says, then, that this number converges to 1. This just is an arithmetic, analytical, truth. Now, $p_{n,\varepsilon}$ seems to be a frequency or a sample measure. But it is none, as de Finetti knows, of course. Sample measures are defined by *real occurences or tokens* $o_s(n)$. Mere possibilities do not exist in the objective, real, world. You cannot see them or touch them. The numbers $p_{n,\varepsilon}$ and 6^n only count *generic* s(n). And the underlying assumption, that any generic sequence s(n) in S(n) gets the same 'probability value' or weight is crucial for the theorem. This assumption entails that $p(a_i) = 1/6$, since this is just the case of n = 1. It is this way, in which all product measures P used in the laws of large numbers are analytically connected with the probability measure p we started with. How can these laws tell us nevertheless something about the justification, the methodological importance or even objectivity of the a priori probabilities p?

3. Genericity

1. Since probability measures in general are not meant as a prediction of only one real sample measure, it is clear from the beginning that no individual observation (of a frequency) as such can prove or refute it. Hence, if we would strictly

stick to a sense-criterion of verifiability or/and refutability by observation, we would be forced to say that statements about probabilities show a lack of precise objective content, since their truth conditions are somehow foggy. In fact, Logical Empiricism had to accept probability statements as meaningful, but could never give a satisfying answer to the question what meaning they have outside the analytic statements of probability calculus. Of course, such a judgment depends on what criteria of satisfaction are used, especially since most philosophers of mathematics and science until today seem to be more or less content with de Finetti's and Carnap's works on inductive logic and Baysianism.

2. The central point which I want to make here is that these authors forget to mention the following fact: When we fix probability measures P on a set of generic possibilities, we not only have to take real frequencies and empirical knowledge into account, but also the possibilities of articulating *generic* knowledge. To do this in *the best possible way* is not only determined by the world, but by our forms of representing situation-invariant knowledge also.

3. The classical approach to probability, going at least as far back as to the work of Laplace, refers to two things, homogeneity and lack of pre-knowledge. The lack of knowledge argument, if applied to our case, says, roughly, that we do not know at all the next results of dicing – if we assume 'fair play'. The homogeneity argument says that this lack of knowledge is not to be viewed just as a lack, but as knowledge about some objective bounds of possible knowledge, such that it is itself to be seen as a peculiar kind of meta-knowledge. In our case it is knowledge about the usual behavior of good dicing utilities or of other mechanical generators of a real random process.

The argument of homogeneity considers cases like the following: If we notice that the dices we are using do not have equal sides, or that their center of gravity is nearer to one side than to another, we have good reasons not to assume equal probabilities. I.e. there are external grounds *not to assume equal probabilities*, neither for some s_i nor for the whole series s(n), if we want to approximate 'most' of the resulting real sample measures in a reasonable way, i.e. in the best way possible. Therefore we check the geometrical and physical properties of a dicing machine (or of some other mechanical or physical generator of a random process). By this, we judge at least in part independently from mere observation of the results $o_s(n)$ produced by running the machine, if we may expect it to be a good, fair, dicing machine. Such a dicing machine is, as we might say, an *objective representation of randomness*, more precisely, of certain probabilities. The reason is this. There are *indefinite possible results* of such a machine.[6] And we predict not just one sequence of such results when we attach a probability measure to the *generic results*. What we do is measuring possibili-

[6] This was pointed out by Paul Lorenzen, head of the so called Erlangen School, a German group of constructivist philosophers of science.

ties, not realities. This is an absolutely crucial difference to a prediction of a real outcome of an election or a real result in a betting game.

4. The relation between a mechanical or physical generator of a random process to a probability model is similar to the objective representations of abstract or ideal geometrical forms by drawings (together with the description of the constructions). There is also an analogy to the dispositional property of being poisonous. A liquid on a shelf has its poisonous results only if used in a specific way, and even then it might not always work. Our dicing machine realizes its 'dispositional probabilities' in the form of the frequencies and sample measures it produces when we run it. But you can run it any time you wish.

5. We often might want to use observations of real results in running a dicing-machine in order to test homogeneity. But a direct appeal to the laws of large numbers is not very helpful here, since for any number n the following holds: If you look at sufficiently many n-series $o_s(n)$, then 'it is almost sure' that at least some of them define very strange sample measures, for example those, which attach to some a_i a value near O or even near 1. Precisely this fact makes the requirement of von Mises totally obscure that in *any* subsequence, picked out by chance, the relative frequencies should converge somehow.

Nevertheless, the laws of large numbers tell us the following: If it is reasonable to assume that all generic possibilities s(n) in S(n) should get the same stochastic weight attached and if two margins ε_1 and ε_2 are given, then we can compute a number m such that for any larger n the following holds: The difference between 1/6 and $P_{os}(n)(a_i)$ should be 'in most cases' smaller than ε_1. The vague expression "in most cases" means: If one watches quite many n-trials, then the frequency that the prediction just mentioned 'will' (or rather: 'might'!) not be fulfilled (compared, of course, to all *possible* cases) is smaller than ε_2. To make things more precise, we imagine the following frequency test of the quality of our dicing machine: If some first n-trials $o_s(n)$ fulfill the margin-condition for $P_{os}(n)$, we might be content and stop the test. If we get disappointed, we should continue at least some more times than 6^n; it is possible to give good estimations for how long we should try.

De Finetti seems to claim that this procedure does not make sense. But he does not consider the possibility of using the mathematical result not only as a prediction or prevision, but also as a norm: A good, fair, dicing machine should fulfill it. I.e., if our frequency test yields a gross deviation from the theorem of large numbers, we have very good reasons *to check the possible flaws in the dicing machine* or to choose or build another generator for random events. Certainly there are limits of our technical abilities to fulfill such norms. I do not question that. It would be equally wrong to question the fact that there are upper bounds to realize the norms of Euclidean geometry in the physical world. Any mathematical model 'contradicts' by some of its 'ideal infinities' the finiteness of the real, observable, world.

With respect to what we know, we nevertheless would be surprised if there were no well-tested dicing machines, for which it is reasonable to expect really to happen what the theorem of large numbers says about fair dicing machines. Hence, the result of Bernoulli's theorem really gives us a (weak!) *criterion for a test if the attachment of equal probabilities* to the s(n) in the case of a dicing machine was reasonable or not. Insofar the theorem can be used as a *methodological principle*.

6. *Rational expectation* is now to be distinguished from *pre-knowledge* about what will happen. But *rational expectation* and *knowledge about generic probabilities* cannot be distinguished at all. In this sense, de Finetti is right to stress the difference between what he calls (rational) 'prevision' and (true) 'prediction'.

7. But how could we convince a skeptical subjectivist that it is reasonable to attach to all generic outcomes s(n) of our dicing machine equal probabilities, and unreasonable not to do so? A possible argument runs as follows: Since we are not concerned here with particular predictions, but with general thumb-rules based on general knowledge, the assumption of another probability measure than the one with equal probabilities needs special arguments. For the assumption entails, in the sense developed above, that in repeating s(n) in sequences s(2n), s(3n),.. ..,s(mn) and so on, the frequency of some property or subset Q of the set S(n) *will eventually almost always significantly differ from P(Q).* This claim says, in effect, that for sufficiently large m the difference should get more and more significant. But if this a result of a *merely arbitrary* decision to choose a probability evaluation, any other choice would be at least as well or as badly confirmed by now, i.e. it would be as 'unreasonable'.

The point is now, that the assumption that the situation is asymmetric always needs special arguments, whereas the assumption of symmetry does not. This asymmetry of the burden of proof is the deep ground for all symmetry principles in all the sciences. Symmetry assumptions function as default principles in cases where we do not have reasons for non-symmetric rules. They are, in such cases, the best generic principles available.

But, of course, there are well known justifications for the choice of non-symmetric probability measures. For this we may consider so called mixtures (de Finetti) or aggregates (von Mises). As an example, think at the following case: After a random choice between m mugs containing different numbers of black and white balls one randomly picks out one ball.

8. If we say that probabilities allow us to articulate generic expectations (or, as de Finetti says, previsions), does this mean that we can use probability only for *types* of events – such that it would be meaningless to talk of the probability of *singular* event like the event that on Dec. 24 2100 12 o'clock it might rain in Leipzig? In fact, I claim that any probability value we attach to such a singular event is justified by the probability of a *generic event* like raining at Dec 24, 12 o' clock, *whichever year*. Hence all the attempts to talk of *mystifying propensities*

or dispositions in order to attach probability values to singular events and only to them, i.e. not just viewed as instances of some generic events, is misguided from the beginning.

9. Another point refers to the question if there is real chance in nature or if chance is merely a matter of lack of knowledge. The first assumption leads to the picture of a god who is dicing. The second is the picture of Laplace, namely that of determinism. Laplace famously talks of a (possible) god or demon who could know all events in advance (if he cared). But both pictures are to be seen as sweeping generalizations. Laplace generalizes cases sweepingly, in which predictions are available. The image of a dicing god generalizes cases, in which no predictions, but only estimated probabilities are available.

10. But did the 'determinist' Laplace not correctly defend a reading of probability as a kind of measure for *epistemic ignorance* with respect to 'real' causes of events? His principle of indifference says that we should consider two possible events as equally possible *if we do not know* any cause that could tell us why the one should be expected in a higher degree of certainty than the other. It is this idea of subjective probability, which is developed by Bayesians, who speak about 'degrees' of beliefs. They believe that Bayes' theorem can be used in 'rational' calculations with such degrees. The idea is this: The a-posteriori-probability P(H/E) of H given evidence E should be proportional to the product of my subjective a-priori-probability evaluation P(H) of H and the probability P(E/H) of E given H. But it is not enough *to say* or claim that such calculation *are rational* and that a priori or subjective probability evaluations of beliefs are rational only if they fulfill these conditions. What is needed is a 'proof' for this claim, or more precisely, an argument for its 'objective' rationality. It is said that de Finetti's theorems give such a proof. But no mathematical (formal, analytical) theorem can do this, as we have seen at the example of the theorem of large numbers. So we are once again thrown back to the situation that such rules are rational only because we do not know of better rules and that they nicely fit to pre-estimations of sample measures.

This can be made even clearer if we consider the meaning of the product rule in rational choice theory. The rule says that the product of P(A), the probability of A, with the net gain G(A) (measured e.g. by money) gives us a measure for rational choice. But why should it be rational to choose in singular cases A over B in all cases when (*) P(A)G(A)>P(B)G(B), even if the probability of A is high, the net gain low, whereas the probability of B is low, but the net gain high? If we gamble only once, we rather gamble with low probability to win and high gains, than with low gains and high probabilities. And we are totally rational to do so: In *singular or rare cases* there is no sufficient reason for saying that it is more rational to make a choice according to (*) than not. Things get even worse for 'rational' decisions theory when we remember that the choices of subjective expectations P(A) and P(B) are fairly arbitrary if we interpret them in a merely

Baysian setting, i.e. as merely subjective a priori evaluations of subjective belief. No wonder, therefore, that F.P. Ramsey later became skeptical about his own subjective approach to probability.

But it is certainly recommendable to use (*) as a criterion for 'rational' choices *if we play the game sufficiently often* or of we read the advice as an advice to a whole group of persons. Then we probably should use the best generic predictions of sample measures possible – given our knowledge. Multiplication with net results just means, then, that the gains are added and maximized for the group.

11. I do not deny that there are applications of probability calculus, which only deal with the consistency of subjective beliefs and expectations. The estimated probabilities, for example, that one of the worst possible accidents in a nuclear power plants (of a certain type) might occur, are fairly subjective and highly hypothetical. The reason is conceptual: It is meaningless to assume that a singular event without predecessor could 'have' any fixed probability value. Nevertheless we may use probability calculus to estimate a 'probability value' of the security or danger of a nuclear power plant. In doing this, we rely, however, on observed frequencies of human mistakes of a certain type or of the occurrence of an error in a computer-program of some kind or of the malfunctioning of some material parts. By doing so, we consider the 'singular case' as a generic case in a set of generic cases. I.e. we relate a possible accident of a certain kind to the numbers of nuclear plants, their parts, their workers, to the time they are running and so on. In the end we talk as if there were certain 'possible' series of non-accidents and accidents, which would give us a hypothetical sample measure for the frequencies of accidents of a certain type. Talking that way may not be totally senseless, even though only somehow foggy estimations are available. But it is good to know about this fogginess, for it remains foggy as long as we do not clarify the generic events, to which our probability measure refers as a thumb-rule-prevision of sample measures.

Pirmin Stekeler-Weithofer
Institute of Philosophy
University of Leipzig
Beethovenstr. 15
D 04109 Germany
stekeler@uni-leipzig.de

References

Carnap, R. (1962). *The Logical Foundations of Probability*. Chicago: Univ. of Chicago Press.

Finetti, B., de (1989). Probabilism. *Erkenntnis, 31*, 165-223.

Finetti, B., de (1974/1975). *Theory of probability. A critical introductory treatment*. 2 vols. London (Wiley) 1974/75 [= *Teoria delle probabilità = Wahrscheinlichkeitstheorie: einführende Synthesis mit kritischem Anhang*. Wien, München: Oldenbourg. 1981].

Fetzer, J. (1981). *Scientific Knowledge*, Dordrecht: Reidel.

Fetzer, J. (Ed.) (1988). *Probability and Causality*, Dordrecht: Reidel.

Gnedenko, B. W. (1968). *Theory of Probability*. New York: Chelsea Publ. [= *Kurs teorii verojatnostej = Lehrbuch der Wahrscheinlichkeitsrechnung*. Berlin: Akademie. 1958].

Hacking, I. (1975). *The Emergence of Probability*. Cambridge: CUP.

Hintikka, J., & P. Suppes (1966). *Aspects of Inductive Logic*. Amsterdam: North Holland.

Jeffrey, R. (1989). Reading 'Probabilismo'. *Erkenntnis, 31*, 225-237.

Keynes, J. M. (1921). *A Treatise on Probability*. London: AMS.

Kyburg, H. E., & Smokler, H. (Eds.) (1964). *Studies in Subjective Probability*. New York: Krieger.

Laplace, P. S. (1951). *A Philosophical Essay on Probabilities*. New York: Dover.

Levi, I. (1989). Possibility and Probability. *Erkenntnis, 31*, 365-386.

Lorenzen, P. (1974). Zur Definition von "Wahrscheinlichkeit. In P. Lorenzen. *Konstruktive Wissenschaftstheorie* (pp. 209 - 218). Frankfurt/M: Suhrkamp.

Morgenstern, D. (1968). Einführung in die Wahrscheinlichkeitsrechnung und mathematische Statistik. Berlin: Springer.

Niiniluoto, I., *Truthlikeness*, Dordrecht 1987.

Popper, K. R. (1963). *Conjectures and Refutations*. London: Routledge.

Salmon, W. (1966). *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Dover.

Stegmüller, W. (1973). *Personelle und statistische Wahrscheinlichkeit.* Berlin: Springer.

Suppes, P. (1984). *Probabilistic Metaphysics*, Oxford: OUP.

Plato, J., von (1994). *Creating Modern Probability*. Cambridge: CUP.

Gödel-Dummett Predicate Logics and Axioms of Prenexability*

Vítězslav Švejdar

1. Introduction

Gödel-Dummett logic in general is a multi-valued logic where a truth value of a formula can be any number from the real interval [0, 1] and where implication is evaluated via the Gödel implication function. As to *truth values,* 0 (falsity) and 1 (truth) are the extremal truth values whereas the remaining truth values are called intermediate. *Gödel implication function* $\Rightarrow$ is defined as follows: $a \Rightarrow b = 1$ if $a \leq b$, and $a \Rightarrow b = b$ otherwise. The truth functions of the remaining propositional symbols conjunction & and disjunction $\vee$ are the functions min and max respectively. Negation $\neg A$ of a formula A is in Gödel-Dummett logic understood as $A \to \perp$ where $\perp$ is a constant for falsity with a truth value equal 0. Thus truth function of negation is the function $a \mapsto (a \Rightarrow 0)$; speaking exactly, $a \Rightarrow 0 = 1$ if $a = 0$ and $a \Rightarrow 0 = 0$ for all $a > 0$.

A particular Gödel-Dummett logic is obtained by restricting the range of possible truth values, i.e. by specifying a truth value set. More exactly, a *logic T* is *based on* a truth value set V where $\{0, 1\} \subseteq V \subseteq [0,1]$ if only the elements of V can be chosen as truth values of propositional atoms. Then a propositional formula A is a *tautology of* that *logic T* or a *tautology of* the *set V* if $v(A) = 1$ for each truth evaluation v based on V, i.e. for each truth evaluation v (a function defined on all propositional atoms and extendible uniquely to all propositional formulas) whose range is a subset of V. One can easily verify that (i) each truth value set V such that $\{0, 1\} \subseteq V \subseteq [0,1]$ is closed under all truth functions $\Rightarrow$, min, and max, (ii) if $V_1 \subseteq V_2$ then all tautologies of the Gödel-Dummett logic based on V_2 are simultaneously tautologies of the logic based on V_1, and (iii) if two truth value sets are order isomorphic then the logics based on them are the same (equivalent). Also, to show that a particular propositional formula A is not a tautology of a logic T, a finite number of truth values is always sufficient. Since in many considerations truth value sets correspond to Kripke frames, we call this simple fact a *finite model property* and denote FMP. As a result, (iv) all

---

* This work is a part of the research plan MSM 0021620839 that is financed by the Ministry of Education of the Czech Republic.

propositional Gödel-Dummett logics based on an infinite truth value set are equivalent. Thus we can define BG, the basic Gödel-Dummett logic, as the logic based on the full real interval $[0,1]$ (or as the logic based on any infinite truth value set V). Furthermore, we can define the logic G_m as the logic based on (any) m-element truth value set, containing the two extremal values 0 and 1 and $m - 2$ intermediate values. We have $BG \subseteq ... \subseteq G_4 \subseteq G_3 \subseteq G_2$, where inclusion $T_1 \subseteq T_2$ between logics indicates that each tautology of T_1 is simultaneously a tautology of T_2. It is evident that Gödel implication function restricted to two-element truth value set is exactly the classical truth function of implication, so G2 is the classical logic.

An elegant *axiomatization* of the logic BG is obtained by adding the *pre-linearity schema* $(A \to B) \vee (B \to A)$ to a Hilbert-style calculus for intuitionistic logic. So BG as well as all the logics G_m are extensions of intuitionistic logic. An example of a formula (schema) which is a tautology of BG is $\neg A \vee \neg\neg A$, while $A \vee \neg A$, the principle of excluded middle, is in general not a tautology either of BG or of any of the logics G_m for $m \geq 3$.

Gödel-Dummett logic is sometimes called *Gödel logic* or *Gödel fuzzy logic.* It was originally invented by Gödel in connection with the question whether a finitely valued semantics can be developed for intuitionistic logic; nowadays it is mostly studied as one of the fuzzy logics, see e.g. Hájek (1998). *Dummett's* important *contribution* is the result that $A \vee B$ is in the logic BG equivalent to $((A \to B) \to B) \ \& \ ((B \to A) \to A)$, so disjunction is in Gödel-Dummett logic expressible in terms of the remaining connectives. Canonical references for Gödel-Dummett logic are the papers Gödel (1932) and Dummett (1959). My motivation to study these logics is probably close to Gödel's: they are interesting extensions of intuitionistic logic.

In this paper we consider Gödel-Dummett *predicate* logics with an emphasis on properties like prenexability and inter-expressibility of quantifiers. The paper overlaps with Kozlíková and Švejdar (2006) co-authored by my former student Blanka Kozlíková. In comparison with Kozlíková and Švejdar (2006), in the present paper we skip some results and most proofs, but we introduce the notion of characteristic class of a logic and we add some semantical considerations. We also borrow a lot of notions and ideas from Baaz, Preining, and Zach (2003).

2. Gödel-Dummett predicate logics

In Gödel-Dummett predicat*e logi*c we consider the same formula*s* as in classical logic, built up from atomic formulas using the propositional symbols $\to$, &, $\vee$, and $\neg$, and quantifiers $\forall$ and $\exists$. As to *omitting parentheses*, we accept the more or less usual convention that implication $\to$ has higher priority than equivalence $\equiv$, but lower than & and $\vee$.

A *multi-valued structure* $\mathcal{J}$ *based on a truth value set* V, or a *multi-valued model based on* V, has a non-empty domain and a truth assignment that associates a truth value $\mathcal{J}(\varphi)[e]$ with every pair φ, e where φ is an atomic formula and e an evaluation of (free) variables. The truth assignment extends uniquely to all formulas using the truth functions of logical connectives defined above, and using the conditions $\mathcal{J}(\forall x\varphi)[e] = \inf_{a\in D}\mathcal{J}(\varphi)[e(x/a)]$ and $\mathcal{J}(\exists x\varphi)[e] = \sup_{a\in D}\mathcal{J}(\varphi)[e(x/a)]$, where D is the domain of the structure $\mathcal{J}$, inf and sup denote the least upper bound (infimum) and greatest lower bound (supremum) respectively, and $e(x/a)$ is the evaluation identical to e except that the variable x is evaluated by $a \in D$. To ensure the existence of suprema and infima, we define a truth value set as a (topologically) *closed* set V such that $\{0, 1\} \subseteq V \subseteq [0, 1]$. In full analogy with the classical case, a formula φ is a *logical truth* of a set V if it is valid in each structure $\mathcal{J}$ based on V, i.e. if $\mathcal{J}(\varphi)[e] = 1$ for each structure $\mathcal{J}$ based on V and each evaluation e of variables.

Example 1 Let $V = \{\frac{1}{2}, 1\} \cup \{\frac{1}{2} - \frac{1}{k}; k \geq 2\}$ and consider a language $\{P\}$ with a single unary predicate P. Let the domain D be the set $\{d_2, d_3, d_4, \ldots\}$ and let the truth assignment be defined by $\mathcal{J}(P(x))[e(x/d_k)] = \frac{1}{2} - \frac{1}{k}$. Note that the numbering of elements of D is chosen so that we have the same k on both sides of the latter equality. Then

$$\mathcal{J}(\exists yP(y))[e] = \sup_{k\geq 2}\mathcal{J}(P(y))[e(y/d_k)] = \frac{1}{2}$$

regardless of e, and $\mathcal{J}(\exists yP(y) \to P(x))[e(x/d_k)] = \frac{1}{2} - \frac{1}{k}$ by the definition of Gödel implication function. So $\mathcal{J}$ is a structure based on V in which the sentence $\exists x(\exists yP(y) \to P(x))$ is not valid because its truth value is $\frac{1}{2}$ under some (and also any) truth evaluation of variables. Thus that sentence is not a logical truth either of our V or of the full real interval $[0, 1]$.

One can even think a little further and verify that the existence of a truth value $a < 1$ in V which is a limit of lower values is essential for Example 1 to work. The sentence $\exists x(\exists yP(y) \to P(x))$ is a logical truth of any truth value set containing no $a < 1$ which is a limit of lower values, and in particular it is a logical truth of any finite truth value set. So Example 1 also shows that finite model property is not true for predicate Gödel-Dummett logic.

The usual lemma saying that if e_1 and e_2 are two evaluations of variables that agree on all free variables of a formula φ then $\mathcal{J}(\varphi)[e_1] = \mathcal{J}(\varphi)[e_2]$ is true also for multi-valued structures. So if φ is a sentence then we can write only $\mathcal{J}(\varphi)$ without specifying the evaluation e. Also, we will write for example $\mathcal{J}(P(d))$ instead of the more correct $\mathcal{J}(P(x))[e(x/d)]$.

By a *logic* we mean any deductively closed set of formulas, i.e. any set of predicate formulas that is closed under the modus ponens and generalization rules. Let G_V, the *Gödel-Dummett logic based* on a truth value *set V*, or a logic *determined by V*, be the logic of all logical truths of V. The basic Gödel-Dummett logic BG is defined as the logic based on the real interval [0, 1], in symbols, BG = $G_{[0,1]}$. The logic G_m for $m \geq 2$ is, as in the propositional case, the logic based on (any) m-element truth value set. In predicate logic it is not true that all infinite truth value sets determine the same logic; this can also be deduced from Example 1. If the properties (i)–(iv) from the second paragraph of Introduction are reformulated for predicate logic, (i)–(iii) remain true, but (iv) is false.

The logic BG is axiomatizable, see e.g. Takano (1987). Its axiomatization is obtained by taking the propositional calculus for BG mentioned above and by adding one quantifier schema

$$S_1: \qquad \forall x(\psi \vee \varphi(x)) \to \psi \vee \forall x\varphi(x),$$

where x is not free in ψ (recall the convention for omitting parentheses above). Each of the logics G_m is axiomatizable as well, see Preining (2003). Baaz et al. (2003) define two more interesting logics $G_\downarrow$ and $G_\uparrow$ as logics determined by the sets $V_\downarrow = \{0\} \cup \{\frac{1}{k}; k \geq 1\}$ and $V_\uparrow = \{1\} \cup \{1 - \frac{1}{k}; k \geq 1\}$ respectively. The formula $\exists x(\exists y P(y) \to P(x))$ is a logical truth of both logics $G_\downarrow$ and $G_\uparrow$. Baaz et al. (2003) also show that neither $G_\downarrow$ and $G_\uparrow$ nor any logic based on a countable infinite truth value set is axiomatizable. Petr Hájek in Hájek (2005) recently obtained more accurate results about the position of the logics $G_\downarrow$ and $G_\uparrow$ in arithmetical hierarchy.

Recall that, in classical logic, prenex operations are formulated as eight equivalences, i.e. sixteen implications, and the schema S1 is one of only three prenex implications that are not intuitionistically valid. The remaining two intuitionistically non-valid prenex implications are

$$S_2: \qquad (\psi \to \exists x\varphi(x)) \to \exists x(\psi \to \varphi(x)),$$
$$S_3: \qquad (\forall x\varphi(x) \to \psi) \to \exists x(\varphi(x) \to \psi),$$

where again x is not free in ψ. Since S_1 is so important in the axiomatization of the logic GB, it seems interesting to think also about S_2 and S_3 as potential axiom schemas. So we *define* S2G, S3G, and PG to be the logics obtained by adding S_2, or S_3, or both S_2 and S_3 respectively, as additional axiom schema(s) to the basic logic BG. Thus PG is the weakest extension of BG in which all the classical prenex operations are available. We will discuss some properties of the logics S2G, S3G, and PG, and we will relate them to the logics $G_\downarrow$, $G_\uparrow$, G_m known from literature.

The idea to study the extensions of the logic BG given by axioms of prenexability may look somewhat unusual because these logics are not determined by truth value sets. Our approach is that a schematical extension of a Gödel-Dum-

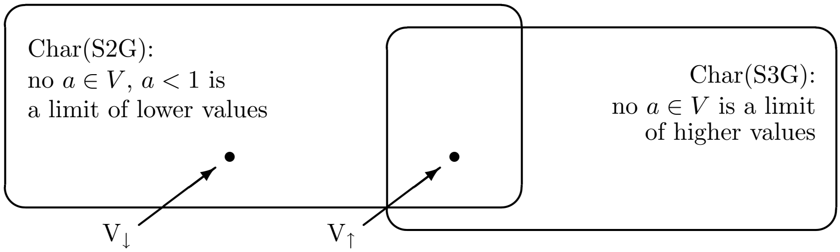


Figure 1: Characteristic classes of S2G, S3G, and PG

mett logic can still be called Gödel-Dummett logic. This is, I suppose, fully in the spirit of Hájek (1998).

Let Char(T), the *characteristic class* of a logic T, be defined as the class of all truth value sets V such that all logical truths of T are valid in all multi-valued structures based on V.

Lemma 2 *(a) If $T_1 \subseteq T_2$, i.e. if each logical truth of a logic T_1 is simultaneously a logical truth of T_2, then* Char(T_2) $\subseteq$ Char(T_1).
(b) If V is a truth value set and T a logic, then $V \in$ Char(T) *if and only if $T \subseteq G_V$.*

Proof If φ is a logical truth of T then φ is valid in any structure $\mathcal{J}$ based on any set in Char(T). If, in addition, $V \in$ Char(T) then φ is valid in any structure based on V. So $\varphi \in G_V$. On the other hand, if $V \notin$ Char(T) then there exists a structure $\mathcal{J}$ based on V and a sentence $\varphi \in T$ not valid in $\mathcal{J}$. Since $\varphi \notin G_V$, we have $T \nsubseteq G_V$. The proof of (a) is similar. ■

Theorem 3 *Over* BG*, the logic* S2G *is equivalently axiomatized by any of the schemas*
$C_\downarrow$: $\exists x(\exists y\varphi(y) \to \varphi(x))$,
E: $\forall x(\forall y(\varphi(y) \to \varphi(x)) \to \varphi(x)) \to \exists x\varphi(x)$.
Its characteristic class is the class of all truth value sets where no value except possibly 1 *is a limit of lower values.*

Proof We show that $C_\downarrow$ and E are (already intuitionistically) equivalent. We omit the proof that S_2 is equivalent to $C_\downarrow$ because it is known or implicit in literature, i.e. in Baaz et al. (2003). We proceed informally, the reader should have no difficulty with formalizing the argument in the appropriate calculus.

$C_\downarrow \Rightarrow$ E: Assume that $\forall x(\forall y(\varphi(y) \to \varphi(x)) \to \varphi(x))$ and let x_0 be such that $\exists y\varphi(y) \to \varphi(x_0)$. We have $\forall y(\varphi(y) \to \varphi(x_0)) \to \varphi(x_0)$. Since $\exists y\varphi(y) \to \varphi(x_0)$ is intuitionistically equivalent to $\forall y(\varphi(y) \to \varphi(x_0))$, we have $\varphi(x_0)$. So indeed, $\exists x\varphi(x)$.

E $\Rightarrow$ $C_{\downarrow}$: To show that $\exists x(\exists y\varphi(y) \to \varphi(x))$, the schema E says that it is sufficient to verify that

$$\forall x(\forall z((\exists y\varphi(y) \to \varphi(z)) \to (\exists y\varphi(y) \to \varphi(x))) \to (\exists y\varphi(y) \to \varphi(x))).$$

So let x be given. Since $A \to (B \to C)$ is equivalent to $A \,\&\, B \to C$, and $(A \to B) \,\&\, A$ is equivalent to $A \,\&\, B$, to verify that

$$\forall z((\exists y\varphi(y) \to \varphi(z)) \to (\exists y\varphi(y) \to \varphi(x))) \to (\exists y\varphi(y) \to \varphi(x))$$

it is sufficient to verify that

$$\forall z(\exists y\varphi(y) \,\&\, \varphi(z) \to \varphi(x)) \,\&\, \exists y\varphi(y) \to \varphi(x). \qquad (*)$$

Taking y_0 such that $\varphi(y_0)$, which is possible by the right conjunct, and then applying the left conjunct to $z := y_0$ quickly shows that $(*)$ is true.

Assume now that V is a truth value set such that no its element, except possibly 1, is a limit of lower values. We have to verify that $\exists x(\exists y\varphi(y) \to \varphi(x))$ is valid in any structure $\mathcal{J}$ based on V. So let $\mathcal{J}$ with domain D be given and take $a_0 = \mathcal{J}(\exists y\varphi(y)) = \sup_{d\in D} \mathcal{J}(\varphi(d))$. If $a_0 = 1$ then $\mathcal{J}(\exists x(\exists y\varphi(y) \to \varphi(x))) = \sup_{d\in D} \mathcal{J}(\exists y\varphi(y) \to \varphi(d)) \geq \sup_{d\in D} \mathcal{J}(\varphi(d)) = 1$. If a least upper bound of a set is not a limit of lower values then it must be an element of that set. So, in the remaining case where $a_0 < 0$, there exists an element $d_0 \in D$ such that $a_0 = \sup_{d\in D} \mathcal{J}(\varphi(d)) = \mathcal{J}(\varphi(d_0))$. Then $\mathcal{J}(\exists x(\exists y\varphi(y) \to \varphi(d_0)) = 1$. Note that in both cases the definition of the Gödel implication function $\Rightarrow$ played a role.

It remains to verify that if the truth value set V contains a value $a < 1$ which is a limit of lower values then there exists a structure $\mathcal{J}$ based on V such that some instance of the schema C is violated. This is however already clear from Example 1. ■

Since the following Theorem 4 does not involve a new schema (like the schema E above), we omit its proof. It is similar to that of Theorem 3.

Theorem 4 S3G *is equivalently axiomatized by* $\exists x(\varphi(x) \to \forall y\varphi(y))$. *Its characteristic class is the class of all truth value sets where no value is a limit of higher values.*

Characteristic classes of logics S2G, S3G, and PG, and the membership of the prominent truth value sets $V_{\downarrow}$ and $V_{\uparrow}$, are depicted in Fig. 1; it is evident that $\mathrm{Char(PG)} = \mathrm{Char(S2G)} \cap \mathrm{Char(S3G)}$. It is important to observe that Char(PG) is rather small: if $V \in \mathrm{Char(PG)}$, i.e. if no element of V, except possibly the element 1, is a limit of other values, then V is finite or isomorphic to $V_{\uparrow}$.

$$\begin{tikzcd}
 & \mathrm{S3G} \arrow[dr] & & & & & & \\
\mathrm{BG} \arrow[ur] \arrow[dr] & & \mathrm{PG} \arrow[dr] & & & & & \\
 & \mathrm{S2G} \arrow[ur] \arrow[dr] & & \mathrm{G}_{\uparrow} \arrow[r] & \cdots \arrow[r] & \mathrm{G}_4 \arrow[r] & \mathrm{G}_3 \arrow[r] & \mathrm{G}_2 \\
 & & \mathrm{G}_{\downarrow} \arrow[ur] & & & & &
\end{tikzcd}$$

Figure 2: Relationships between Gödel-Dummett logics

It is easy to verify that the schema $\forall x(\forall y(\varphi(y) \to \varphi(x)) \to \varphi(x)) \equiv \exists x \varphi(x)$, resulting from replacing the outermost implication in the schema E by equivalence, is also provable in S2G. So we have the following Theorem.

Theorem 5 *In* S2G *and thus in all its extensions, the existential quantifier is expressible in terms of the remaining logical symbols.*

Theorem 6 *The relationships between the logics we consider are as shown in Fig. 2.*

Proof S2G $\subseteq$ PG and S3G $\subseteq$ PG is immediate. S2G $\subseteq \mathrm{G}_{\downarrow}$ follows from Lemma 2(b), as well as PG $\subseteq \mathrm{G}_{\downarrow}$. The inclusions $\mathrm{G}_{\downarrow} \subseteq \mathrm{G}_m$ and $\mathrm{G}_{\uparrow} \subseteq \mathrm{G}_m$, for each m, follow from property (ii) in the Introduction. Baaz et al. (2003) show that $\mathrm{G}_{\uparrow} = \bigcap_{m \geq 2} \mathrm{G}_m$. From this we have $\mathrm{G}_{\downarrow} \subseteq \mathrm{G}_{\uparrow}$.

As to non-inclusions, the fact that S3G $\not\subseteq \mathrm{G}_{\downarrow}$ follows from $\mathrm{V}_{\downarrow} \notin$ Char(S3G) and Lemma 2(b). Also, S2G $\not\subseteq$ S3G follows from Char(S3G) $\not\subseteq$ Char(S2G) and Lemma 2(a). For the more complicated proof of $\mathrm{G}_{\downarrow} \not\subseteq$ PG see Kozlíková and Švejdar (2006); the proof is also outlined in Section 3 below. ■

So, by Theorem 5, the quantifier $\exists$ is expressible in terms of $\forall$ and logical connectives in the logics S2G, PG, $\mathrm{G}_{\downarrow}$, $\mathrm{G}_{\uparrow}$, and all G_m. Petr Cintula verified that the schema E, with equivalence as the outermost symbol, is provable also in logics that we do not consider here, namely in all logics extending MTL+S_2, where the logic MTL is defined in Esteva and Godo (2001). So also in all these logics the existential quantifier is expressible in terms of the remaining logical symbols. Petr Cintula also remarked that the fact that the existential quantifier is expressible using *only* the symbols $\forall$ and $\to$ may be new even for the logic G_2, the classical two valued logic.

Further results in Kozlíková and Švejdar (2006) say that the quantifier $\exists$ is not expressible in terms of $\forall$ and logical connectives in S3G, and the quantifier $\forall$ is not expressible in terms of $\exists$ and logical connectives even in G_3. Also, for both logics S2G and S3G there exist formulas that are not equivalent to prenex formulas. To obtain these results, Kripke semantics is sometimes used as well. It is important to realize that one can work with a semantics – multi-valued or Kripke – even in the absence of completeness theorem: for some results, the soundness theorem is sufficient.

While PG is the weakest extension of the basic logic BG in which all the classical prenex operations are valid, it still seems to be an interesting problem whether PG is the weakest extension of BG in which any formula is equivalent to a prenex formula.

3. Remarks on semantics and completeness

The non-inclusion $G_\downarrow \not\subseteq$ PG asserts the existence of a sentence $\varphi \in G_\downarrow$ such that $\varphi \notin$ PG. However, if V is a set in Char(PG), i.e. if V is finite or isomorphic to $V_\uparrow$ then, by $G_\downarrow \subseteq G_\uparrow$, the sentence φ is valid in any structure based on V. So we conclude that $\varphi \notin$ PG cannot be shown by taking a truth value set from the logic's characteristic class and defining a structure $\mathcal{J}$ based on V such that $\mathcal{J}(\varphi) < 1$. The logic PG is *incomplete* with respect to its characteristic class.

The problem whether PG (or S2G, or S3G) is complete with respect to some semantics is left open in Kozlíková and Švejdar (2006). Hájek and Cintula (2006) offer a solution: the logic PG is complete with respect to witnessed structures. Their result can probably be generalized also for S2G and S3G. A structure $\mathcal{J}$ with a domain D is *witnessed* if, whenever $\varphi(x,y_1,..,y_n)$ is a formula and the variables $y_1,..,y_n$ are evaluated by $d_1,..,d_n \in D$, the set $\{\mathcal{J}(\varphi(d,d_1,..,d_n))\ ;\ d \in D\}$ of truth values has both maximal and minimal element.

Without using the notion of witnessed structure, a structure $\mathcal{J}$ satisfying the definition is constructed in Kozlíková and Švejdar (2006) to show that $G_\downarrow \not\subseteq$ PG. The structure $\mathcal{J}$ looks as follows. The truth value set V contains a value $a_0 < 1$ which is a limit of lower values. There are only finitely many values greater than a_0 and all values in V except a_0 are isolated. Let Q be a function from V to V defined by $Q(a) = a$ for $a \leq a_0$ and $Q(a) = a_0$ for $a \geq a_0$. Importantly, the function $[a,b] \mapsto Q(a \Rightarrow b)$, from V^2 to V, is continuous. The structure $\mathcal{J}$ is chosen so that its domain D is equipped with a compact topology and so that for each atomic formula $\varphi(x_1,..,x_n)$ the function $[d_1,..,d_n] \mapsto Q(\mathcal{J}(\varphi(d_1,..,d_n)))$ is continuous as a function from D^n to V. Then using some topological knowledge and equations like $Q(\min\{a,b\}) = \min\{Q(a),Q(b)\}$ *an*d $Q(a \Rightarrow b) = Q(Q(a) \Rightarrow Q(b))$ one can show that the function $[d_1,..,d_n] \mapsto Q(\mathcal{J}(\varphi(d_1,..,d_n)))$ is continuous for *every* formula φ. So every set of the form $\{Q(\mathcal{J}(\varphi(d,d_1,..,d_n)))\ ;\ d \in D\}$ is topologically closed, and as such it must have both maximal and minimal element.

The set $\{J(\varphi(d,d_1,..,d_n))\,;\, d \in D\}$ may be not closed, but one can conclude that it must have both maximal and minimal element, too.

The construction described in the previous paragraph suggests that, in particular case, it may not be so easy to verify that a given structure is witnessed.

Vítězslav Švejdar
Department of Logic, Charles University
Palachovo nám. 2, 116 38 Praha 1, Czech Republic
vitezslav.svejdar@cuni.cz, http://www.cuni.cz/~svejdar/

References

Baaz, M., & Preining, N., & Zach, R. (2003). Characterization of the axiomatizable prenex fragments of first-order Gödel logics. In *33rd International Symposium on Multiple-valued Logic, May 16–19, 2003* (pp. 175–180). Tokyo: IEEE Computer Society Press. (A longer version will appear in *Annals Pure Appl. Logic.*)

Dummett, M. (1959). A propositional calculus with denumerable matrix. *Journal of Symbolic Logic*, *25*, 97–106.

Esteva, F., & Godo, L. (2001). Monoidal t-norm based logic: Towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, *123* (3), 271–288.

Gödel, K. (1932). Zum intuitionistischen Aussagenkalkül. *Anzeiger Akademie der Wissenschaften Wien, Math.-naturwissensch. Klasse*, *69*, 65–66. (See also Ergebnisse eines matematischen Kolloquiums 4 (1933), 40.)

Hájek, P., & Cintula, P. (2006). On theories and models in fuzzy predicate logics. (To appear in *Journal of Symbolic Logic.*)

Hájek, P. (1998). *Metamathematics of fuzzy logic*. Dordrecht: Kluwer.

Hájek, P. (2005). A non-arithmetical Gödel logic. *Logic Journal of the IGPL*, *13* (4), 435–441. (A special issue devoted to selected papers presented at the Challenge of Semantics workshop, Vienna, July 2004.)

Kozlíková, B., & Švejdar, V. (2006). On interplay of quantifiers in Gödel-Dummett fuzzy logics. *Archive for Math. Logic*, *45* (5), 569–580.

Kozlíková, B. (2004). *Sémantické metody v intuicionistické predikátové logice (Semantical Methods in Intuitionistic Predicate Logic)*. Master's thesis, Faculty of Arts and Philosophy of Charles University, Department of Logic.

Preining, N. (2003). *Complete recursive axiomatizability of Gödel logics*. Unpublished doctoral dissertation, Vienna University of Technology, Austria.

Takano, M. (1987). Another proof of strong completeness of the intuitionistic fuzzy logic. *Tsukuba Journal of Math.*, *11*, 101–105.

Proof and Explanation in Mathematics[1]

Susan Vineberg

Introduction

Mathematics, like empirical science, aims not only at establishing facts, but also at explaining them. Although scientific explanation has received considerable attention, very little work has been done on mathematical explanation. To the extent that mathematical explanation has been considered, much of the attention has focused on distinguishing explanatory from non-explanatory proofs. Drawing on some of this work, it is argued that the main accounts of scientific explanation do not fare well at this task. It is also shown that there are mathematical explanations that arise from proof in which it is not the theorem proven that is explained by the proof. This indicates that distinguishing explanatory from non-explanatory proofs fails to exhaust the subject of explanation in mathematics. Furthermore, the principal accounts of explanation in science do not cover this additional kind of mathematical explanation. An alternative view is suggested that treats mathematical explanation as a special case of a very general account of explanation, as the displaying of dependencies, which Thalos has developed to treat explanation in science in terms of physical dependence. Finally, it is suggested that this view helps us to understand what is missing in probabilistic computer proofs.

Distinguishing explanatory and non-explanatory proofs

The discussion of explanation in mathematics has thus far largely been confined to distinguishing explanatory proofs, which provide insight, from non-explanatory proofs, which establish theorems without explaining why they are true. Indeed, there are entire classes of proofs, such as those by induction, which seem to offer little in the way of understanding. As an example, Steiner considers two proofs of the following formula, which gives the sum of the first n positive natural numbers:

[1] I especially wish to thank to Madeline Muntersbjorn, Sean Stidd, and Eric Hiddleston for discussion and support. I also wish to thank the audience members at the meeting of the Society for Exact Philosophy 2006 and at Logica 2006 for their useful comments.

(1) $S(n) = 1 + 2 + 3 + ... + n = n(n+1)/2.$ [2]

The theorem is easily proven by induction, with the following induction step:

$$S(n+1) = S(n) + n + 1 = n(n+1)/2 + 2(n+1)/2 = (n+1)(n+2)/2$$

But this proof does not appear explanatory at all. In particular, it fails to show us why the equality holds, but only shows us that it holds. Steiner contrasts this inductive proof of the theorem with the following geometric proof, which he takes as explanatory:

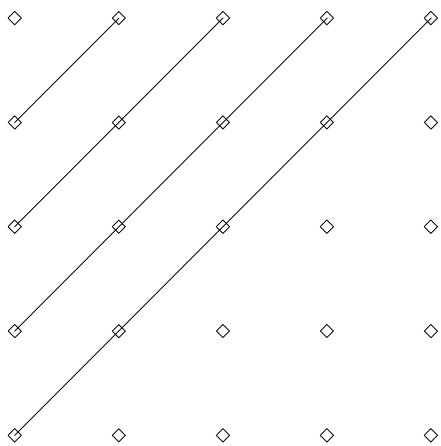


Figure 1

The square of n^2 dots is constructed of two triangles. When we put these together the diagonal is counted twice, so

$$S(n) + S(n) = n^2 + n.$$

This second proof makes it entirely clear that the formula on the left side of (1) is correct. Of course this is not the same as providing an understanding of exactly why the theorem is true, though it seems that the second proof is more explanatory than the first. These examples suggest that it is in fact misleading to divide proofs into those that are explanatory and those that are not. Although some proofs may offer so little by way of explanation that we may be inclined to say that they are not explanatory at all, and it may be tempting to put the first proof into this category, in general proofs should be considered to be more

[2] Steiner, 1978.

or less explanatory, rather than explanatory or not. A satisfactory account of explanation and proof should be sensitive to this point. However, it is important to observe that it is not merely that the second proof seems more explanatory than the first, but that the second proof offers a kind of insight into the theorem that is lacking in the proof by induction. Given two proofs of the same theorem, it may be that overall one proof is more explanatory than the other, but in some cases they may each offer a distinct form of understanding.

Theories of explanation

Whereas relatively little has been written about explanation in mathematics, a variety of theories of scientific explanation have been proposed and defended. Although the general differences between scientific and mathematical practice suggest that there may be significant differences in the nature of explanation between the two, it is plausible that some insight can be gained by applying some of the accounts of scientific explanation to mathematics. Most of the contemporary theories of scientific explanation arose in response to the Deductive Nomological account of explanation, according to which a fact is said to be explained by deducing it from general laws.[3] Despite the many problems that make it inadequate as a theory of scientific explanation, it does have some features that seem useful for describing explanatory proofs. After all, such proofs often take place by deducing particular claims from more general axioms and theorems. Moreover, it has the advantage that we might account for degrees of explantoriness in terms of the generality of the principles from which particular propositions are deduced.

As Steiner discusses, there are cases in which the most explanatory proof is indeed the most general. He gives as an example the proof of the Pythagorean theorem that proceeds by observing the similarlity of ABC, DAB and DAC.

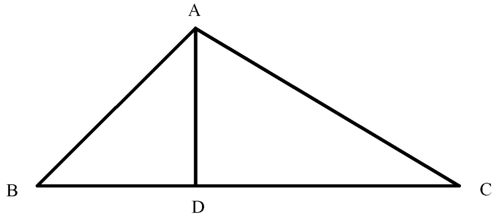


Figure 2

[3] Hempel & Oppenheim, 1948. Of course, Hempel and Oppenheim targeted scientific explanation, and accordingly placed the requirement on DN explanations that they must have empirical content. Unless mathematics is to be regarded as itself empirical, then this condition would have to be dropped in order to apply the account to mathematical explanation.

The proof relies on a special case of the fact that the areas of any two similar plane figures are to each other as the squares of their corresponding sides. This makes the proof both general and explanatory according to Steiner[4], and fits with the view he takes Feferman to endorse:

> Abstraction and generalization are constantly pursued as a means to reach really satisfactory explanations which account for scattered individual results. In particular, extensive developments in algebra and analysis seem necessary to give us real insight into the behavior of the natural numbers. [Systems of Predicative Analysis (Feferman, 1969)][5]

The idea that explanation involves generality not only corresponds with the DN account, but also with the view that explanation consists in unifying different phenomena that were previously taken to be unrelated under a common set of principles, which is exemplified by Newton's explanation of the tides in which he brought them under the scope of his general law of gravitation. Along these lines, Friedman (1974) put forth the idea that explanation involves reducing the number of laws needed to explain various phenomena in diverse domains, thus unifying them, and Kitcher (1989) later offered a related account whereby explanation consists in reducing the number of argument patterns from which the facts to be explained can be derived. Most notably Kitcher suggested that one of advantages of his unification theory is that it applies to both scientific and mathematical explanation (Kitcher, 1975 and 1989).

The unification view of scientific explanation can be seen in a number of respects as a successor of Hempel's theory.[6] Indeed Kitcher held that his version remedied various problems with that account including its inability to handle the asymmetry of explanation. It clearly solves another problem with the DN account of scientific explanation, which if anything is even worse in the mathematical case, namely that it counts any derivation of a fact from general principles as explanatory. Such derivations do increase understanding in mathematics, making the DN model fit some cases of mathematical explanation. However, the fact that the DN model as applied to mathematics would count as explanatory any derivation from general principles makes it unsuitable as an account of explanatory proof. To begin with, it threatens to make virtually every proof explanatory. Perhaps this could be avoided by putting some restriction on what counts as a general axiom, and thus as satisfying the model. But this would have

---

[4] The triangles *DAB* and DAC are similar to *ABC* and thus to each other. The areas of similar plane figures are to each other as the squares of the corresponding sides. Therefore, $ADC/ABC = AC^2/BC^2$ and $ADB/ABC = AB^2/BC^2$. But $ADC + ADB = ABC$, which entails that $AB^2 + AC^2 = BC^2$.

[5] Quoted in Steiner (1978).

[6] See (Salmon, 1989) for discussion.

the unwelcome consequence of allowing us to turn an unexplanatory proof into an explanatory one simply by replacing one of the assumptions by a more general one, and this need not correspond to a genuine increase explanatory power. In trying to capture how generalization fits into mathematical explanation, we should simply jettison the DN model, whereas some version of the unification theory might be retained as capturing this aspect of explanation.

There are several ways in which what can be described as unifying proofs can contribute to explanation in mathematics. The first, in which familiar mathematical truths are deduced from more fundamental principles, most clearly fits the unification theory. Russell and Whitehead's work in Principia Mathematica (Russell & Whitehead, 1910), in which they derive arithmetical truths from what they take to be simpler (logical) axioms, thus unifying arithmetic and logic, is a prime example. It is notable that such "explanations" do not consist in appealing to principles that are clearer or more easily grasped, quite the opposite, but we must distinguish between explanation as making easier to grasp, and explanation as locating fundamental reasons. It seems quite right to say that in reducing arithmetical truths, such as 2 + 2 = 4, to logical truths, we gain understanding, though in this case perhaps not the sort of understanding that is terribly helpful in developing mathematics. Another way in which unifying proofs increase understanding can be seen in cases where facts or principles from one domain are used in proving results in another area. Such an example of unification and explanation involves the linking of topology, specifically the theory of elliptic curves, and number theory in the proof of Fermat's last theorem. The use of algebraic methods in topology and the application of category theory not only provide examples of explanatory unification, but also examples in which such explanatory activity contributes tremendously to the development of mathematics. In some cases it is not so much that facts or techniques are brought in from other domains as that new techniques are developed and employed in a variety of proofs, which results in a broad unified mathematical domain. Riemann's approach to complex analysis provides an important example of this, and as Tappenden (2005) discusses, it offers a form of understanding that falls under the unification model proposed by Kitcher.

Despite these considerations, as in empirical science, it appears doubtful that explanation in mathematics consists in the unification of disparate facts and that the proposed models of explanation as unification can capture explanatory activity within mathematics in its entirety. Consider again the example of the sum of the first n numbers. The second proof does appear to be more explanatory than the first, but it neither involves the unification of principles nor even any sort of subsumption under general laws. The proof has a geometrical character, but its explanatoriness does not stem from appeal to unifying principles of geometry in any clear way. One response here would be to invoke the distinction between explanation as making clear and explanation as locating fundamental reasons. It might be suggested that the second proof makes it clear that $S(n) = n(n+1)/2$, but does not provide us with an underlying reason. While there may be a sense in which the proof fails to make such reasons explicit, it nonetheless accounts for the relationship between $S(n)$ and the square of n. It accomplishes this by giving a geometric representation to the sum of the first n natural numbers, but it does not appeal, at least explicitly, to fundamental geometric laws. As I see it, this proof is explanatory, but does not appeal to any general principles, and thus fits neither the unification nor of course the DN model.

Perhaps the geometrical proof is a borderline case of an explanatory proof, and hence not a challenge for the unification account. However, Steiner, who argues against the view that generality is required for explanation, does want to count this as an explanatory proof. On his view,

> An explanatory proof makes reference to a characterizing property of an entity or structure mentioned in the theorem, such that from the proof it is evident that the result depends on the property. (Steiner, 1978, p. 143)

Curiously, this account does not appear to fit with Steiner's own claim that the geometric proof is explanatory. While Steiner does not define "characterizing property", leaving us with only a vague description of what he takes as an explanatory proof, it is hard to see how the notion could reasonably cover the geometric proof. The relevant structure would seem to be that of the natural numbers, but the proof does not mention the characterizing properties of that structure. Understanding the proof as proceeding from the characterizing properties of natural numbers just takes us away from those features that make it explanatory; it is in fact the non-explanatory proof by induction that is tied more closely to the characterizing properties of the natural numbers. Instead, we might take the relevant object to be the sum of the first n positive integers. While the proof may be said to turn on a feature of that sum, that feature is not a characterizing property.

Putting aside the lack of clarity in his notion of a characterizing property, a limitation of Steiner's account is that it is too narrow to cover the various forms of explanation that occur in mathematical proof. In fairness to Steiner, he does not appear to be after all of these, but rather those proofs that would generally be regarded by mathematicians as illuminating why the theorem proven is true. As suggested above, there is a sense in which the proofs in Principia are explanatory, for they show how many truths of arithmetic can be reduced to what are, at least arguably, principles of logic. Nevertheless, these proofs hardly make the theorems more transparent, and so are not explanatory in the sense that Steiner wants to capture. But, as his example of the geometric proof of the formula for the sum of the first n positive natural numbers suggests, his account is unsatisfactory even if it is only supposed to apply to "illuminating" proofs. In any case, there are many other forms of mathematical explanation that occur through proof, including the foundational proofs of Principia, proofs that relate one area of mathematics to another, proofs that contribute to overall unity and ones that account for earlier results.[7] Whether there can be any single theory of mathematical explanation that illuminates all of these remains to be seen.

In addition to "illuminating proofs", there are cases in which proofs provide understanding of previous results that are not well accounted for on the unification model. Consider the case of the Banach-Tarski Theorem (Banach, 1924), which says that a ball in R^3 can be decomposed into a finite number of pieces to produce two balls of the same volume as the first. Banach and Tarski also showed in their (1924) proof that the result generalizes to every dimension greater than three. They emphasized their use of the Axiom of Choice in obtaining the paradoxical result, which is often called the Banach-Tarski Paradox. It seemed at the time that the proof of the theorem depends upon the axiom of choice. In 1963 Paul Cohen proved by his method of forcing that the Axiom of Choice is independent of the basic axioms of Zermelo Frankel set theory (Cohen, 1966). Later Robert Solovay (1970) extended Cohen's results to show that if a slight strengthening of the Zermeo-Frankel axioms is consistent, then these axioms cannot prove the Banach-Tarski Theorem. Solovay's result provides insight into the earlier proof by explaining why it indeed required the axiom of choice, and it is this explanatory fact that gives interest to Solovay's theorem. Notice that the explanatory power of Solovay's result does not come from the features of its proof, nor from any unifying feature of the work. This case also shows clearly that explanations in mathematics, while arising from mathematical activity and hence from proof, need not coincide with explanatory proofs in the usual sense. Moreover, the example is suggestive of an account of explanation that can make sense of many cases of mathematical explanation, including those considered here.

Advocates of the unification view take Newton's explanation of the tides to consist in bringing the phenomenon under the umbrella of his theory of gravitation, and see the explanatory power of the theory as stemming from its ability to unite astronomical and terrestrial phenomena. A different view takes the explanation to consist in showing how the tides are produced by the gravitational force exerted by the moon, rather than in showing how the tides are related to other phenomena. This interpretation is generally associated with the causal theory of explanation, which is the one contemporary view of scientific explana-

---

[7] There are other examples in which we explain some step in a proof by taking note of some underlying fact, which does not involve any sort of characterizing property of an entity or structure mentioned in the theorem. For a detailed case study see (Hafner & Mancosu, 2005).

tion that does not seem applicable to mathematics, since causal relationships do not obtain between mathematical objects. However, causal explanation can be understood as a special case of a more general account that takes explanation to consist in the exhibiting of dependence relations. Thalos (2002) has put forth such a view of explanation within physics as exhibiting dependence relations between quantities. She argues that causal dependence is merely one form of dependence that may figure in explanation, and develops the general idea of explanation as the exhibiting of dependence relations as an extension of an account of logical dependence first proposed by Grelling (1988).

The idea that explanation involves exhibiting dependency relations is helpful in illuminating various cases of mathematical explanation. There are examples throughout mathematics, but they arise frequently in foundational studies. There is much to be said about such cases that would be useful in illustrating the view that revealing dependence relations has a central role in mathematical explanation, but here a brief indication of some examples will have to suffice. To begin, it makes sense to say that Solovay's result was explanatory because it showed that the Banach-Tarski theorem depends on the axiom of choice. More generally this idea allows us to understand foundational studies as aiming not only at providing an edifice from which mathematics can be reconstructed, which has been viewed as a dubious enterprise on both philosophical and mathematical grounds, but also at bringing out fundamental dependencies, which plays a legitimate and central role in mathematics. While the logicist projects of Frege and Russell were not entirely successful, it is reasonable to say that they showed that some arithmetical truths depend on logical truths, and that this work was explanatory. More recent foundational work within the theory of definability further illustrates this role. Hilbert's axiomatization of geometry also offers evidence for the idea that explanation in mathematics involves exhibiting dependency relationships. Hilbert's axiomatization of geometry was instrumental to our understanding geometry and in particular to the algebraic structures corresponding to geometries. His system contained a great number of axioms, which allowed him to show precisely which assumptions the various theorems depended on, which led to considerable advance.

The revealing of connections between mathematical phenomena, as suggested by Kitcher, and the articulating of underlying dependencies are fundamental goals in mathematics. [8] It is these key tasks of foundational studies that unite it with mathematics as a whole. Interestingly, this allows us to view category theory and the more traditional foundational studies of logic and set theory as having overlapping, yet also somewhat different, foundational and explanatory

[8] In suggesting that mathematics seeks to display underlying dependencies, I do not presuppose a realist view of mathematics, according to which mathematicians are simply discovering pre-existing truths. One might hold that displaying dependencies is in fact a constructive activity.

roles. Both study mathematical structures as such, and thus can offer mathematical explanations that involve exhibiting underlying mathematical structure. Category theory is more specifically concerned with revealing structural connections between mathematical phenomena, whereas logic and set theory address logical and propositional, or axiomatic, dependencies, making them not so much alternative foundations for mathematics, but rather approaches that address different aspects of a central feature of mathematical practice. It is significant that both offer explanations through the development of a unifying edifice, which suggests that such explanations can be assimilated to the unification view. Indeed, the idea that explanations point to various relations of dependence is embraced by Kitcher, but he maintains that such dependence is grounded in the inferential ordering of our beliefs.[9] This would seem to allow him to piggyback on the gains offered by those who take explanation in science to consist in exhibiting relations of causal dependence, such as being able to account for the asymmetry of explanation, without taking such dependence as basic. The idea that dependence of various sorts is taken as arising from the organization of our belief system is very much in keeping with Hempel's empiricist approach, but more importantly, given the formal development of mathematics, it makes the view especially promising as an account of mathematical explanation. Such a view has the advantage that mathematical dependence is ultimately given by the organizational features of the theoretical systems that we adopt, thus avoiding an appeal to any sort of fundamental (and perhaps mysterious) concept of mathematical dependence.

While there are attractions to reducing the notion of dependence at work in explanations to structural features of our inferential practice, it remains doubtful that the unification theory can accommodate various examples in a way that does justice to the way in which they figure in increasing our mathematical understanding. While Solovay's work explains why the axiom of choice is needed to prove the Tarski-Banach theorem, it does so without producing greater unification. This is not to say that Solovay's result and methods do not exhibit generality, but that the explanation it gives rise to is independent of such generality. Steiner's second proof of the formula for S(n) shows us why the quantity depends upon n^2, whereas the inductive proof does not. It accomplishes this by relating the quantity to a geometrical representation, but this association alone does not involve reducing our set of argument patterns. To the extent that this can be accomplished, it would require so much more than is exhibited in the proof that it must be said that we simply do not grasp the unification through this proof. Yet our understanding is increased in these cases through the dependence relationships displayed in the proofs. A defender of the view that explanation just is unification must insist that any sort of dependence that is taken

[9] Kitcher, 1989, p. 436.

to be explanatory reduces to facts about the ordering of our actual inference patterns, which entails that there is form of unification that characterizes our reasoning even in cases where this is not at all apparent. Even if such a claim is defensible, our understanding is increased through the simple display of dependence, whereas it is unclear that we should regard ourselves as having increased understanding where the basis for it (i.e. increased unification) remains beyond our grasp.[10]

Instead of trying to fit mathematical explanation into the unification theory, the suggestion here is to focus on the various forms of dependence that are cited in mathematical explanations. Such dependence relationships can often be described broadly as structural. The developments in the late 20th century that led to the final proof of Fermat's last theorem illustrate an unfolding of such dependence relationships. Wiles finally proved the theorem by establishing a special case of the Taniyama-Shimura Conjecture, concerning modular forms, which implies it. In this case, the dependency was suggested through extended work on modular forms and elliptic curves that developed a relationship between the structures, culminating in Wiles' proof which definively established the connection for the important special case (of Taniyama-Shimura) that yields Fermat's last theorem as a consequence.

The contemporary view of mathematics as concerned with abstract structure,[11] and indeed that this constitutes its subject matter, fits naturally with the idea that mathematical explanation involves showing that a mathematical object or collection of objects has a particular structure or that one structure can be embedded in another.[12] Explanations are thus associated with functions from one structure to another or from one set of quantities to another. This follows the work of Grelling who took dependence relations to be straightforwardly characterized by functions, of which he gave definitions for several types. Of course, in mathematics functional dependence of one sort or another is always at hand; understanding is achieved by pointing to the right form of dependence. In some cases, explanatory dependence is not to be located by considering mathematical objects or structures as such, but rather emerges at the level of description. For example, theorems are sometimes to be explained in terms of the quantifier structure in their statement. Of course these cases too involve a kind of functional dependence.

Paying attention to the various forms of functional dependence allows us to make sense out of the idea that some proofs seem more explanatory than others. Consider once again the two proofs of formula for S(n). While the second proof

---

[10] Woodward discusses this problem for the unification account that it doesn't seem to account for our increased understanding in (Woodward, 2003).

[11] Shapiro, 1997 and Resnik, 1997.

[12] Here we see a division of explanations similar to Hempel's between the explanation of particular facts and that of general regularities.

strikes us as more explanatory, it is not that the first, which by employing induction establishes the result from the fundamental axioms of arithmetic, should be taken as entirely without explanatory value. As noted previously, foundational work can indeed be explanatory, but the level of dependence is typically different from that exhibited in the proofs that we usually find illuminating. This suggests that there is a pragmatic dimension to what we consider explanatory, and which singles out the relevant form of dependence.[13] Perhaps most often it is our concerns in developing mathematics that determine the kind of understanding, and in particular the form of dependence, that we take to be explanatory.

Computer proofs

Let us turn briefly to the to the controversy over probabilistic computer proofs. Some think that that these do not count as proofs.[14] Others say that they do count, because they provide evidence for their conclusions, indeed they give us what is in some ways stronger evidence than we may be able to produce through more traditional proofs.[15] The idea here is that we can never be absolutely certain of a result, or at least one whose proof has any complexity, for it is always possible that we have made some kind of error. Computers, on the other hand, provide us with an increased capacity to rule out errors, both in the case of probabilistic and non-probabilistic proofs.

If the aim of proof is simply to provide evidence, then I think it must be granted that a probabilistic proof of a theorem can provide us with as much reason to accept it as true as some more traditional proofs do. However, such a view of mathematical proof is surely too limited. For one thing, it cannot account for the differences in value that various proofs are said to have. One might try to account for these differences in terms of their complexity, the techniques developed, the difficulty in obtaining the proof, or perhaps on some kind of aesthetic grounds. No doubt at least some of these criteria do contribute to the value that we place on various proofs, but it is highly doubtful that these will suffice.

I suggest that along with evidence a primary role of proof is to explain. Thus, while probabilistic computer proofs may be perfectly good on evidential grounds, in that they can provide overwhelming reason to think that certain theorems are true, they fail to be explanatory. In particular they fail to exhibit the dependence relationships that are the cornerstone of mathematical explanation. They may indicate that such relationships exist, but they do not display

[13] Sandborg applies the pragmatic theory of explanation to mathematics, but finds that it is not entirely satisfactory (Sandborg, 1998). However, his point does not undercut the idea that there is a pragmatic dimension to mathematical explanation.

[14] See Jaffe & Quinn, 1993.

[15] The thesis that these should count as proofs is defended in (Fallis, 1997).

them. This is not to say that some computer-generated proofs cannot be quite explanatory. In cases where a proof is produced through a series of deductive rules from a set of axioms, the proof may indicate some dependence relations. It is rather those probabilistic computer proofs in which the computer output does not correspond to the lines in a traditional proof where there is evidence without understanding.

Department of Philosophy
Wayne State University
Detroit MI
susan.vineberg@wayne.edu

References

Banach, S., & Tarski, A. (1924). Sur la décomposition des ensembles de points en parties respectivement congruentes. *Fund. Math, 6,* 244–277.

Cohen, P. (1966). *Set Theory and the Continuum Hypothesis*. New York: Benjamin.

Fallis, D. (1997). The Epistemic Status of Probabilstic Proof. *The Journal of Philosophy, 94* (4), 165–186.

Feferman, S. (1969). Systems of Predicative Analysis. In J. Hintikka (Ed.). *The Philosophy of Mathematics*. Oxford Univeristy Press.

Friedman, M. (1974). Explanation and Scientific Understanding. *Journal of Philosophy, 71*, 5–19.

Grelling, K. (1988). A Logical Theory of Dependence. In B. Smith (Ed.). *Foundations of Gestalt Theory* (p. 217–226). Munich: Philosophia Verlag.

Hafner, J., & Mancosu, P. (2005). The Varieties of Mathematical Explanation. In P. Mancosu, J. Jørgensen & S.A. Pedersen (Eds.). *Visualization, Explanation and Reasoning Styles in Mathematics,* (p. 215–250). Springer.

Hempel, C. G., & Oppenheim. P. (1948). Studies in the Logic of Explanation. *Philosophy of Science, 15,* 567–579.

Jaffe, A., & Quinn, F. (1993). Theoretical Mathematics: Toward a Cultural Synthesis of Mathematics and Theoretical Physics. *Bulletin of the American Mathematical Association, 29,* 1–13.

Kitcher, P. (1975). Bolzano's Ideal of Algebraic Analysis. *Studies in History and Philosophy of Science, 6,* 229–269.

Kitcher, P. (1989). Explanatory Structure and Causal Unification. In P. Kitcher, & W. C. Salmon (Eds.). *Scientific Explanation* (p. 410–505). Minneapolis: University of Minnesota Press.

Resnik, M. (1997). *Mathematics as a Science of Patterns*. New York: Oxford University Press.

Russell, B., & Whitehead, A. N. (1910). *Principia Mathematica.* Cambridge: Cambridge University Press.

Salmon, W. C. (1989). Four Decades of Scientific Explanation. In P. Kitcher, & W. C. Salmon (Eds.). *Scientific Explanation,* (p. 3–219). Minneapolis: University of Minnesota Press.

Sandborg, D. (1998). Mathematical Explanation and the Theory of Why-Questions. *British Journal for the Philosophy of Science, 49,* 603–624.

Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology*. New York: Oxford University Press.

Solovay, R. (1970). A model of set theory in which every set of reals is Lebegue measurable. *Annals of Mathematics, 92,* 1–56.

Steiner, M. (1978). Mathematical Explanation. *Philosophical Studies, 34,* 135–151.

Tappenden, Jamie. (2005). Proof Style and Understanding I: Visualization, Unification and Axiom Choice. In P. Mancosu, J. Jørgensen and S.A. Pedersen (Eds.). *Visualization, Explanation and Reasoning Styles in Mathematics.* Springer.

Thalos, M. (2002). Explanation is Genus: An Essay on the Varieties of Scientific Explanation. *Synthese, 130,* 317–354.

Woodward, J. (2003). Scientific Explanation. Stanford Encyclopedia of Philosophy.

Meinongian Theories without Ad Hoc Restriction – Taking Two-Modes-of-Predication Approach as an Example

Wen-fang Wang

The ideas of fixed points (Kripke, 1975 and Martin & Woodruff, 1975) and revision sequences (Gupta & Belnap, 1993 and Gupta, 2001) have been exploited to provide solutions to the liar paradox and have achieved some success. This happy situation naturally encourages one to look for other philosophical areas of their applications where paradoxical results seem to follow from intuitively acceptable principles. In this paper, I propose to extend the use of these ideas to give two new treatments of Meinongian objects.

1. The naïve Meinongian theory

I begin with a second-order language L. The primitive symbols of L include "$\sim$", "$\supset$", "$=$", "$\exists$", "(", ")", individual constants a, b, c, a_1, b_1, c_1, ..., individual variables x, y, z, x_1, y_1, z_1, ..., predicate constants F^n, G^n, H^n, $F^n{}_1$, $G^n{}_1$, $H^n{}_1$, ..., and n-place predicate variables X^n, Y^n, Z^n, $X^n{}_1$, $Y^n{}_1$, $Z^n{}_1$, ...for each n. I will sometimes omit superscripts for predicates when contexts make it clear what they are. Other connectives and universal quantifiers are defined in the usual way. The grammar of L is the standard one except that it allows identity sign "=" to be flanked by, and only by, either individual terms or predicate terms of the same superscripts on both sides. I use meta-variables "d", "d_1", "d_2", ... for individual constants, "v", "v_1", "v_2", ... for individual variables, "α", "α_1", "α_2", ... for individual terms in general, "P", "P_1", "P_2", ..., with or without superscripts, for predicate constants, "V", "V_1", "V_2", ..., with or without superscripts, for predicate variables, and "β", "β_1", "β_2",..., with or without superscripts, for predicate terms in general. When our discussion proceeds, we will have opportunities to add new symbols and formation rules to L.

As usual, a model M for L is an order pair <D, I>, where D is a non-empty set and I is an interpretation function assigning members of D to individual constants and subsets of D^n to n-place predicate constants. An assignment function g is a function assigning members of D to individual variables and subsets of D^n to n-place predicate variables. The valuation function $v_{M,g}$ for formulas is defined in the standard way with the addition clause that

$$v_{M,g}(\beta_1=\beta_2)=1 \text{ iff } v_{M,g}(\beta_1)= v_{M,g}(\beta_2).$$

For a Meinongian who believes that every thought is a thought about some object, the following two principles look attractive and, perhaps, intuitively compelling too:

(OB_S) For every set S of properties, there is an object having exactly those properties in S.

(CP_S) For every set S of objects, there is a property had by exactly those objects in S.

(OB_S) and (CP_S) are apparently contradictory, however, for (OB_S) implies that the cardinality of the set of objects is greater than that of the set of properties, while (CP_S) implies the contrary. Without trying to remedy this problem at the moment, In L we approximate (OB_S) and (CP_S) by:

(OB_N) $(\exists x)(X)(Xx \equiv \phi(X))$, where "$\phi(X)$" is a wff in which "$x$" does not occur free;

and

(CP_N) $(\exists X)(x)(Xx \equiv \phi(x))$, where "$\phi(x)$" is a wff in which "$X$" does not occur free.

Though weaker, (OB_N) and (CP_N) are more congenial than (OB_S) and (CP_S) for a Meinongian who wants to avoid the apparent contradiction involved in the latter and who assumes that only sets of properties that are describable or expressible in our language or graspable in our thought correspond to objects. Note that (CP_N) is no more than a special case of the general comprehension principle in standard second-order logic:

(CP) $(\exists X^n)(x_{1...}x_n)(X^n x_{1...}x_n \equiv \phi(x_{1...}x_n))$, where "$\phi(x_{1...}x_n)$" is a wff in which "$X^n$" does not occur free.

We call the standard (incomplete) axiomatization of second-order logic (as given in Shapiro 2001) together with (OB_N) and identity axioms for relations "the naïve Meinongian theory". It can be shown that every axiom of the naïve Meinongian theory except (OB_N) is valid in the semantics given above.

2. The main problem

The main problem of the naïve Meinongian theory is that (OB_N) and (CP), together with other logical axioms, form an inconsistent whole.

The inconsistency arises in two ways. First, it may arise because some objects have incomplete or contradictory natures. Thus, the object o_1 having exactly the properties of being a golden mountain exemplifies neither the property of being red nor the property of being not red, while the object o_2 having both the property of being red and the property of being not red exemplifies both by its nature. Both o_1 and o_2 lead directly to a contradiction. We call this kind of inconsistency "Sosein Paradox" for Meinongianism.

Second, the inconsistency may arise because (OB_N) and (CP_N) make incompatible claims about the size of the set of objects and that of properties. The inconsistency is obvious when we look at the set-theoretical version (OB_S) and (CP_S), but it also emerges in the naïve Meinongian theory. Thus, either (CP_N) plus a specialized case of (OB_N) , viz., "$(\exists x)(X)(Xx \equiv (\exists y)((z)(Xz \equiv z = y) \;\&\; {\sim}Xy))$", or ($OB_N$) plus a specialized case of (CP_N), viz., "$(\exists X)(x)(Xx \equiv (\exists Y)(Yx \;\&\; {\sim}Yx))$", gives rise to a contradiction. The proofs of these contradictions are proceeded by the familiar diagonal method[1]. We call this kind of inconsistency "Cardinality Paradox" for Meinongianism.

So the naïve Meinongian theory has no model. Consequently, no model M and value assignment g can be such that, for every formula "$\phi(X)$" not containing free occurrence of "x", every formula "$\phi(x_{1\ldots}x_n)$" not containing free occurrence of "X", and every $x_{1\ldots}x_n$-alternatives g^* of g and X-alternatives g' of g:

$$v_{M,g'}(Xx)=v_{M,g'}(\phi(X)),$$

and

$$v_{M,g^*}(Xx_{1\ldots}x_n)=v_{M,g^*}(\phi(x_{1\ldots}x_n)).$$

3. Two modes of predication

There are two existing Meinongian approaches devoted to avoid the inconsistency affecting the naïve theory: the first distinguishes two modes of predication – exemplifying and encoding (Zalta, 1983; Rapaport, 1979; Castaneda, 1989), while the second distinguishes two kinds of properties – nuclear and extranuclear properties (Parsons, 1980). Both treatments work in a classical setting and both put some restriction on (OB_N) and/or (CP). I will only deal with the two-modes-of-predication approach in this paper, and shall start with a theory

[1] See Zalta (1983) Appendix A for similar proofs.

which is similar to, but simpler than, Zalta's (1983). (I will have a few words to say about the other approach, however. See the concluding part of this paper.)

Grammatically, the two-mode-of-predication approach adds to our grammar, beyond the familiar mode of exemplification "$\beta\alpha$", a new mode of predication "$\alpha\beta$", read as "α encodes β", where "β" is a one-place predicate term and "α" is an individual term. The idea behind the new mode of predication is that, when an object encodes a property, it merely contains, "is determined by", or "is ascribed to" that property, which does not mean that it also exemplifies the property. Distinguishing two modes of predication only complicates the semantics a little bit: every one-place predicate term in L will have both an exemplification extension and an encoding extension by I and *g*, and two predicate terms are identical, given I and *g*, iff they have both extensions exactly the same. Otherwise, the semantics is not revised. It can still be shown that every axiom of the naïve Meinongian theory except (OB_N) is valid in the new semantics. Note that (CP) is a principle about the exemplification extension, rather than about the encoding extension, of a n-place predicate; it does not apply when the left part of the equivalence is an encoding formula.

The two-mode-of-predication approach insists, however, that objects of thoughts are merely ascribed to properties in thoughts, which by no means entail that they will also exemplify those properties ascribed to them. Thus, a more reasonable object-comprehension principle for such an approach is not (OB_N), but rather:

(OB_Z) $\quad (\exists x)(X)(xX \equiv \phi(X))$,

where $\phi(X)$ is a wff in which "x" does not occur.

We call the result of replacing (OB_N) by (OB_Z) in the naïve Meinongian theory "the Meinongian theory Z*", or simply "theory Z*".

It is easy to see that Sosein Paradox do not arise in theory Z*, and this is partly due to the failure of mutual entailment between the encoding relation and the exemplifying relation and partly due to the fact that (CP) is only a principle about the exemplification relation. However, while Z* enables a Meinongian to shun away from Sosein Paradox, Cardinality Paradox persists. Thus, either (CP) together with a specialized case of (OB_Z), viz., "$(\exists x)(X)(xX \equiv (\exists y)((z)(Xz \equiv z = y)\ \&\ {\sim}yX))$", or ($OB_Z$) together with a specialized case of "(CP), viz., "$(\exists X)(x)(Xx \equiv (\exists Y)(Yx\ \&\ {\sim}xY))$", give rise to contradictions. Once again, the proofs are proceeded by the familiar diagonal method[2], and therefore Z* has no model.

To fix the problem, Zalta suggests that we restrict (CP) to those "$\phi(x_{1\ldots}x_n)$"'s

[2] See Zalta (1983) Appendix A for relevant proofs.

that "are propositional", i.e., have no encoding subformula, and define identity statements in terms of encoding ones:

$$\text{(D) } \alpha_1{=}\alpha_2 =_{df} \alpha_1{=}_E\alpha_2 \vee (\sim E\alpha_1 \ \& \sim E\alpha_2 \ \& \ (X)(\alpha_1 X \equiv \alpha_2 X)),$$

where "E" is a primitive predicate for "exist", and "$=_E$" is primitive relation supposed to hold only between an existent object and itself. We call the revised (CP) "(CP_Z)" and the theory resulting from such a restriction on (CP) "theory Z".

Theory Z may very well be paradox-free, but the price of taking it is not cheap. First, though saving the theory from being inconsistent, the restriction in (CP_Z) is *ad hoc*. For, if it is natural to say that o exemplifies the property of being an exemplifier of F when o satisfies "F*x*", then it is equally natural to say that o exemplifies the property of being an encoder of F when o satisfies "*x*F". Second, the definition of the identity relation is not metaphysically plausible, for even a formula like "$(X)(xX \equiv yX)$" would requires us to recognize that the properties that *x* encodes are *exactly the same as* the properties that *y* encodes. To cite from Hawthorne: "That a predicate expressing identity *could be* explicitly introduced by one of the mechanisms stated does not imply that the concept of identity is dispensable or parasitic: the point remains that mastery of the apparatus of quantification would appear to require an implicit grasp of identity and difference." (Hawthorne 2003, p. 105) Third, Zalta's definition for the identity relation leaves it possible that two "identical" objects may nevertheless exemplify different properties, for the definition requires only that two non-existent objects be identical when they *encode* the same properties. It therefore takes Zalta to formulate a "proper axiom" to get rid of the unfavorable result that identical objects may be discernible. Fourth, so far, nothing prevents us from having a *primitive* relation *Y* such that $(x)(y)(Yxy \equiv x{=}y)$. Yet, were there such a relation *Y*, "$(\exists x)(X)(xX \equiv (\exists y)((z)(Xz \equiv Yzy) \ \& \sim yX))$", ($CP_Z$), and the meaning postulate "$(x)(y)(Yxy \equiv x{=}y)$" would give rise to a contradiction again[3].

4. Fixed-point approach

We believe that the restriction on (CP) and the definition of identity made in theory Z are unnecessary if either fixed-point techniques or ideas about revision sequences are adequately employed. We describe the use of fixed-point techniques in this section.

To remove the *ad hoc* restriction on (CP), we first note that the law of the

[3] The proofs are similar to those given in Zalta (1983) Appendix A. I will omit them for the limit of space.

excluded middle (LEM) is used in deriving the Cardinality Paradox. This observation motivates the use of many-valued logics in which (LEM) fails. Further, it is natural for a Meinongian theory to opt for a four-valued logic, in view of the incomplete or impossible natures of some objects. Thus, the use of a four-valued scheme is well-motivated, and fixed-point techniques can be used to generate a model in which *a form* of (OB_Z) and (CP) holds unrestrictedly.

To fulfill the plan, we first need to choose among all possible four-valued valuation schemes. We decide to choose the relational semantics for FDE as invented by Dunn (1986), but extended it to second-order logic. (Other choices are also possible, so long as schemes being chosen are monotonic.) A model for FDE is a pair $\langle D, I\rangle$, where D is a non-empty set of objects and I is a function assigning each individual constant d a member of D, each n≥2-place predicate constant P^n a pair of sets $\langle ex(P^n), aex(P^n)\rangle$, each set being a subset of D^n, and each one-place predicate constant P^1 a pair of pairs of sets $\langle\langle ex(P^1), aex(P^1)\rangle, \langle en(P^1), aen(P^1)\rangle\rangle$, each set being a subset of D. Intuitively, $ex(P^n)$ and $aex(P^n)$ are, separately, the exemplification extension and anti-exemplification extension of P^n, and $en(P^1)$ and $aen(P^1)$ are, separately, the encoding extension and anti-encoding extension of P^1. A value assignment g is a function similar to I except that it assigns objects or ordered pairs to variables. I use "$\mathbf{P}$", "$\mathbf{Q}$", "$\mathbf{P_1}$", "$\mathbf{Q_1}$", ... as variables for pairs of pairs of sets, "Γ", "Δ", "Γ_1", "Δ_1", ... as variables for sets of (n-tuples of) objects, and "$\boldsymbol{\Gamma}$", "$\boldsymbol{\Delta}$", "$\boldsymbol{\Gamma_1}$", "$\boldsymbol{\Delta_1}$", ... as variables for sets of properties, i.e., sets of pairs of pairs of sets. When $\langle S, T\rangle$ is an ordered pair, we let $l(\langle S, T\rangle)$ be S and $r(\langle S, T\rangle)$ be T. We now define the valuation function $v_{M,g,}$ as follows:

(i) $v_{M,g}(d)=I(d)$
(ii) $v_{M,g}(v)=g(v)$
(iii) $v_{M,g}(P^n)=I(P^n)$
(iv) $v_{M,g}(V^n)=g(V^n)$
(v) $v_{M,g}(\alpha\beta^1)=1$ iff $v_{M,g}(\alpha)\in l(r(v_{M,g}(\beta^1)))$
(vi) $v_{M,g}(\alpha\beta^1)=0$ iff $v_{M,g}(\alpha)\in r(r(v_{M,g}(\beta^1)))$
(vii) $v_{M,g}(\beta^1\alpha)=1$ iff $v_{M,g}(\alpha)\in l(l(v_{M,g}(\beta^1)))$
(viii) $v_{M,g}(\beta^1\alpha)=0$ iff $v_{M,g}(\alpha)\in r(l(v_{M,g}(\beta^1)))$
(ix) $v_{M,g}(\beta^n\alpha_1...\alpha_n)=1$ iff $\langle v_{M,g}(\alpha_1)...v_{M,g}(\alpha_n)\rangle\in l(v_{M,g}(\beta^n))$, for n≥2
(x) $v_{M,g}(\beta^n\alpha_1...\alpha_n)=0$ iff $\langle v_{M,g}(\alpha_1)...v_{M,g}(\alpha_n)\rangle\in r(v_{M,g}(\beta^n))$, for n≥2
(xi) $v_{M,g}(\alpha_1=\alpha_2)=1$ iff $v_{M,g}(\alpha_1)=v_{M,g}(\alpha_2)$
(xii) $v_{M,g}(\alpha_1=\alpha_2)=0$ iff $v_{M,g}(\alpha_1)\neq v_{M,g}(\alpha_2)$
(xiii) $v_{M,g}(\beta_1=\beta_2)=1$ iff $v_{M,g}(\beta_1)=v_{M,g}(\beta_2)$
(xiv) $v_{M,g}(\beta_1=\beta_2)=0$ iff $v_{M,g}(\beta_1)\neq v_{M,g}(\beta_2)$
(xv) $v_{M,g}(\sim\phi)=1$ iff $v_{M,g}(\phi)=0$
(xvi) $v_{M,g}(\sim\phi)=0$ iff $v_{M,g}(\phi)=1$
(xvii) $v_{M,g}(\phi\supset\psi)=1$ iff $v_{M,g}(\phi)=0$ or $v_{M,g}(\psi)=1$

(xviii) $v_{M,g}(\phi \supset \psi)=0$ iff $v_{M,g}(\phi)=1$ and $v_{M,g}(\psi)=0$
(xix) $v_{M,g}((\exists v)\phi)=1$ iff there is an v-alternative g' of g such that $v_{M,g'}(\phi)=1$
(xx) $v_{M,g}((\exists v)\phi)=0$ iff all v-alternative g'of g are such that $v_{M,g'}(\phi)=0$
(xxi) $v_{M,g}((\exists V^n)\phi)=1$ iff there is a V-alternative g' of g such that $v_{M,g'}(\phi)=1$
(xxii) $v_{M,g,g'}((\exists V^n)\phi)=0$ iff all V-alternative g'of g are such that $v_{M,g'}(\phi)=0$

To get the model and the value assignment function that we are searching for, we first associate each formula of the form "$\phi(x_{1...}x_n)$" with a unique n-place predicate variable V^n by a function p, distinct formulas with distinct predicate variables, and associate each formula of the form "$\phi(X)$" with a unique individual variable v by a function i, distinct formulas with distinct individual variables. Given an infinite[4] model **M** and a value assignment g that assigns different objects to different individual variables, we define a function s from the range of p to pairs (of pairs) of sets and a function t from the range of i to pairs of properties as follows: (To get a hint of what functions p, i, s, t, u, and κ are doing, think of p and i as giving a temporary "name" for each relation and individual defined by (OB_Z) and (CP). Given **M** and g, we want to achieve a new model **M**' and a new value assignment g' by κ such that **M** and **M**' agree in almost everything except that **M**' interprets the encoding extensions and anti-encoding extensions of one-place predicates in accordance with (OB_Z), while g' and g differ radically. We first want g' to assign new exemplification extensions and anti-exemplification extensions to those new "names" of relations according to (CP), which is done by s, and then assigned new encoding extensions and anti-encoding extensions of "names" of one-place relations according to (OB_Z), which is partly done by t and partly done by u.)

$s(V^1)=<<\Gamma_1, \Delta_1>, <\Gamma_2, \Delta_2>>$, where $<\Gamma_2, \Delta_2>=r(g(V^1))$, $o_1\in\Gamma_1$ iff $v_{M,g'}(p^{-1}(V^1))=1$ for some $g'=g[x_1/o_1]$, and $o_1\in\Delta_1$ iff $v_{M,g'}(p^{-1}(V^1))=0$ for some $g'=g[x_1/o_1]$.
$s(V^{n\geq 2})=<\Gamma, \Delta>$, where $<o_1...o_n>\in\Gamma$ iff $v_{M,g'}(p^{-1}(V^n))=1$ for some $g'=g[x_1/o_1,...,x_n/o_n]$, and $<o_1...o_n>\in\Delta$ iff $v_{M,g'}(p^{-1}(V^n))=0$ for some $g'=g[x_1/o_1,...,x_n/o_n]$.
$t(v)=<\mathbf{\Gamma}, \mathbf{\Delta}>$, where $\mathbf{P}\in\mathbf{\Gamma}$ iff $v_{M,g'}(i^{-1}(v))=1$ for some $g'=g[X/\mathbf{P}]$, and $\mathbf{P}\in\mathbf{\Delta}$ iff $v_{M,g'}(i^{-1}(v))=0$ for some $g'=g[X/\mathbf{P}]$.

Given s and t, we define a function u from the union of one-place predicates and the range of p to their extensions as:

[4] Finite model will not do if we want objects encoding different properties to be distinct. For if there are only finitely many, say n, objects, then (CP) will assure us that there are at least 2^n different properties while (OB_Z) will generate still greater number of objects. However, we allow objects encoding the same properties to be distinct, which seems to be harmless for our purpose.

(i) $u(V^{n\geq 2})=s(V^{n\geq 2})$.
(ii) $l(u(V^1))=l(s(V^1))$.
(iii) $r(u(V^1))=\langle\Gamma, \Delta\rangle$, where $g(v)\in\Gamma$ iff $s(V^1)\in l(t(v))$ and $g(v)\in\Delta$ iff $s(V^1)\in r(t(v))$.
(iv) $l(u(\mathrm{P}^1))=l(\mathrm{I}(\mathrm{P}^1))$.
(v) $r(u(\mathrm{P}^1))=\langle\Gamma, \Delta\rangle$, where $g(v)\in\Gamma$ iff $\mathrm{I}(\mathrm{P}^1)\in l(t(v))$ and $g(v)\in\Delta$ iff $g(\mathrm{P}^1)\in r(t(v))$.

Now, we define a "jump" κ as $\kappa(\langle \mathrm{M}, g\rangle)=\langle \mathrm{M}', g'\rangle$, where $\mathrm{M}'=\langle \mathrm{D}, \mathrm{I}'\rangle$ is like $\mathrm{M}=\langle \mathrm{D}, \mathrm{I}\rangle$ except that $\mathrm{I}'(\mathrm{P}^1)=u(\mathrm{P}^1)$, and g' is like g except that $g'(V^n)=u(V^n)$.

Now, if we define:

(i) $\langle\Gamma_1, \Delta_1\rangle\leq\langle\Gamma_2, \Delta_2\rangle$ iff $\Gamma_1\subseteq\Gamma_2$ and $\Delta_1\subseteq\Delta_2$.
(ii) $\langle\langle\Gamma_1, \Delta_1\rangle, \langle\Gamma_2, \Delta_2\rangle\rangle\leq\langle\langle\Gamma_1', \Delta_1'\rangle, \langle\Gamma_2', \Delta_2'\rangle\rangle$ iff $\langle\Gamma_i, \Delta_i\rangle\leq\langle\Gamma_i', \Delta_i'\rangle$.
(iii) $\mathrm{I}\leq\mathrm{I}'$ iff $\mathrm{I}(\mathrm{d})=\mathrm{I}'(\mathrm{d})$ and $\mathrm{I}(\mathrm{P}^n)\leq\mathrm{I}'(\mathrm{P}^n)$.
(iv) $g\leq g'$ iff $g(v)=g'(v)$ and $g(V^n)\leq g'(V^n)$.
(v) $\mathrm{M}\leq\mathrm{M}'$ iff $\mathrm{M}=\langle \mathrm{D}, \mathrm{I}\rangle$, $\mathrm{M}'=\langle \mathrm{D}, \mathrm{I}'\rangle$ and $\mathrm{I}\leq\mathrm{I}'$.
(vi) $\langle \mathrm{M}, g\rangle\leq\langle \mathrm{M}', g'\rangle$ iff $\mathrm{M}\leq\mathrm{M}'$ and $g\leq g'$

It then can be shown that κ is monotonic in the sense that, for any M, M', g, and g', $\kappa(\langle \mathrm{M}, g\rangle)\leq\kappa(\langle \mathrm{M}', g'\rangle)$ if $\langle \mathrm{M}, g\rangle\leq\langle \mathrm{M}', g'\rangle$[5].

Finally, we define "the result of n-applications of κ to $\langle \mathrm{M}, g\rangle$" (short as $\kappa^n(\langle \mathrm{M}, g\rangle)$) by the following transfinite inductive definition:

(i) $\kappa^0(\langle \mathrm{M}, g\rangle)=\langle \mathrm{M}^0, g^0\rangle=\langle \mathrm{M}, g\rangle$.
(ii) $\kappa^{i+1}(\langle \mathrm{M}, g\rangle)=\langle \mathrm{M}^{i+1}, g^{i+1}\rangle=\kappa(\kappa^i(\langle \mathrm{M}, g\rangle))$.
(iii) For limit ordinals π, $\kappa^\pi(\langle \mathrm{M}, g\rangle)=\langle \mathrm{M}^\pi, g^\pi\rangle$, where, for every d and P^n, $r(\mathrm{M}^\pi)(\mathrm{d})=\mathrm{I}(\mathrm{d})$ and $r(\mathrm{M}^\pi)(\mathrm{P}^n)=\cup_{i<\pi} r(\mathrm{M}^i)(\mathrm{P}^n)$, and, for every v and V^n, $g^\pi(v)=g(v)$ and $g^\pi(V^n)=\cup_{i<\pi} g^i(V^n)\rangle$.

Now if we start with an infinite model $\mathrm{M}=\langle \mathrm{D}, \mathrm{I}\rangle$ and an assignment g such that, for every P^1, we have $r(\mathrm{I}(\mathrm{P}^1))=\langle\varnothing, \varnothing\rangle$, and for every V^1 and V^n ($n\geq 2$), we have $g(V^1)=\langle\langle\varnothing, \varnothing\rangle, \langle\varnothing, \varnothing\rangle\rangle$ and $g(V^n)=\langle\varnothing, \varnothing\rangle$, then there is an ordinal number p such that $\kappa^{p+1}(\langle \mathrm{M}, g\rangle)=\kappa(\kappa^p(\langle \mathrm{M}, g\rangle))=\kappa^p(\langle \mathrm{M}, g\rangle)$. So, $\kappa^p(\langle \mathrm{M}, g\rangle)$ is a fixed point of κ, and a minimal or smallest fixed point of κ. The model $l(\kappa^p(\langle \mathrm{M}, g\rangle))$ and the value assignment $r(\kappa^p(\langle \mathrm{M}, g\rangle))$ are such that, for every formula "$\phi(X)$" not containing free occurrence of "x", every formula "$\phi(x_{1\ldots}x_n)$" not containing free occurrence of "X", and every $x_{1\ldots}x_n$-alternatives g^* of $r(\kappa^p(\langle \mathrm{M}, g\rangle))$ and X-alternatives g' of $r(\kappa^p(\langle \mathrm{M}, g\rangle))$:

[5] This follows immediately form the monotonicity of $v_{\mathrm{M},g}$: if $\langle \mathrm{M}, g\rangle \leq \langle \mathrm{M}', g'\rangle$, then $v_{\mathrm{M}',g'}$ is at least as informative as $v_{\mathrm{M},g}$. Hence formulas satisfied (unsatisfied) by g under M must also be satisfied (unsatisfied) by g' under M'. Consequently, $\kappa(\langle \mathrm{M}, g\rangle)\leq\kappa(\langle \mathrm{M}', g'\rangle)$.

$$v_{M,g'}(xX)=v_{M,g'}(\phi(X)),$$

and

$$v_{M,g^*}(Xx_{1...}x_n)=v_{M,g^*}(\phi(x_{1...}x_n)).$$

The minimal fixed point given above, however, does not verify (CP) and (OB_Z), though it verifies a restricted version of them: when "≡" does not occur in "$\phi(X)$" and "$\phi(x_{1...}x_n)$", and this is because, when both $v_{M,g'}(xX)$ and $v_{M,g'}(\phi(X))$ or both $v_{M,g^*}(Xx_{1...}x_n)$ and $v_{M,g^*}(\phi(x_{1...}x_n))$ are "undefined", i.e., neither true nor false, putting an equivalence sign between them will yield an undefined formula too. Nevertheless, the minimal fixed point is not the only fixed point of κ; actually κ has infinitely many fixed points. Some of its fixed points are such that both the union of the exemplification extension and the anti-exemplification extension and the union of the encoding extension and the anti-encoding extension of *every* predicate exhaust D. If we interpreted the language by such fixed points, the resulted interpretations will verify both (CP) and (OB_Z). Furthermore, they will be such that paradoxical sentences, such as "$(\exists x)(X)(xX \equiv (\exists y)((z)(Xz \equiv z = y)\ \&\ {\sim}yX))$" and "$(\exists X)(x)(Xx \equiv (\exists Y)(Yx\ \&\ {\sim}xY))$", are both true and false. This is not to say, however, that these interpretations are completely satisfactory; they are not for at least two reasons. First, the four-value valuation scheme we used for proving the existence of fixed points is expressively incomplete: Lukasiewicz's biconditionals and exclusion negation are not expressible in it, for example. Second, those interpretations that verify (CP) and (OB_Z) are also those that make every sentence either true or false, which ruins our motivation for choosing a four-value valuation scheme. Whether these "defects" are serious ones deserves further philosophical investigation.

5. Revision sequence approach

Another way to remove the *ad hoc* restriction made by Zalta on (CP) is to, first of all, regard both (CP_Z) and (OB_Z) as "circular definitions", and then employ revision sequences for every hypothesis of the extension of *each* predicates to provide a semantics for the language. In order to justify this treatment, we first add to our language an infinite list of individual constants of the form "$<\lambda Y^1\ \phi(Y^1)>$", where "$\phi(Y^1)$" is a formula that may contain free occurrences of "Y^1", and an infinite list of predicate terms of the form "$[\lambda y_1...y_n\ \phi(y_1...y_n)]$", where "$\phi(y_1...y_n)$" is a formula that may contain free occurrences of "y_1" ... "y_n". (CP) and (OB_Z) now becomes the closures of:

($CP^{\#}$) $$[\lambda y_1...y_n\ \phi(y_1...y_n)](x_{1...}x_n) \equiv \phi(x_{1...}x_n),$$

and

$$(\mathrm{OB_Z}^{\#}) \qquad X{<}\lambda Y^1\ \phi(Y^1){>} \equiv \phi(X),$$

where "$\phi(Y^1)$" is a wff in which "$<\lambda Y^1\ \phi(Y^1)>$" does not occur.

Now, it should be obvious that ($\mathrm{CP}^{\#}$) and ($\mathrm{OB_Z}^{\#}$) are circular if we take "≡" in them to be a sign of definition. Actually, ($\mathrm{CP}^{\#}$) and ($\mathrm{OB_Z}^{\#}$) are doubly circular in the sense that each principle defines entities of a sort in terms of entities of the other sort defined by the other principle, and the condition in each principle may involve a quantifier ranging over entities of the very sort defined by the principle itself. Thus, Gupta and Belnap's semantics and logic for circular definitions can be readily applied to ($\mathrm{CP}^{\#}$) and ($\mathrm{OB_Z}^{\#}$).

To be more specific, we let L' be our original language L expanded with terms of the form "$<\lambda Y^1\ \phi(Y^1)>$" and "$[\lambda y_1...y_n\ \phi(y_1...y_n)]$". Let M=<D, I> be an infinite *base* model for L, where I is a function from individual constants to members of D, from n-place predicates to sets of n-tuples of D, and from one-place predicates to ordered pairs $<\Gamma, \varnothing>$, where $\Gamma \subseteq D$ and $\varnothing$ is the empty set. The valuation function $v_{M,g}$ is the standard classical one with the additional clauses that:

$$v_{M,g}(P^1\alpha)=1 \text{ iff } v_{M,g}(\alpha)\in l(v_{M,g}(P^1))$$
$$v_{M,g}(\alpha P^1)=1 \text{ iff } v_{M,g}(\alpha)\in r(v_{M,g}(P^1))$$

We call any function $f = f_0 \cup f_1 \cup f_n$ a "hypothesis", where

(i) $f_0(<\lambda Y^1\ \phi(Y^1)>)\in D$ such that distinct constants of the form "$<\lambda Y^1\ \phi(Y^1)>$" are assigned distinct objects in D;
(ii) $f_1(P^1)=f_1([\lambda y\ P^1(y)])=<l(I(P^1)), \Delta>$, where $\Delta\subseteq D$;
(iii) $f_1([\lambda y\ \phi(y)])\in \mathcal{P}(D)\times\mathcal{P}(D)$, where "$[\lambda y\ \phi(y)]$" is a predicate different from those in (ii);

and

(iv) $f_n(\beta^{n\geq 2})\in\mathcal{P}(D^n)$, where β^n is of the form "$[\lambda y_1...y_n\ \phi(y_1...y_n)]$".

We let M+f =<D, I'> be a model exactly like M except that it interprets each individual constant "$<\lambda Y^1\ \phi(Y^1)>$" as $f(<\lambda Y^1\ \phi(Y^1)>)$ and each predicate β^n as $f(\beta^n)$. Then the revision rule $R_M(f)=R^b_M(R^a_M(f))$ (we omit the subscript subsequently) in M can be conceived as a function from the set of hypotheses into itself such that:

(i) $R^a(f)(<\lambda Y^1\ \phi(Y^1)>)=(f)(<\lambda Y^1\ \phi(Y^1)>)$;
(ii) $R^a(f)(P^1)=R^a(f)([\lambda y\ P^1(y)])=f(P^1)=f([\lambda y\ P^1(y)])$;
(iii) $R^a(f)([\lambda y\ \phi(y)])=<\Gamma, \Delta>$, where "$[\lambda y\ \phi(y)]$" is not a predicate covered by (ii), $\Delta=r(f([\lambda y\ \phi(y)]))$, and $\Gamma=\{o \mid$ there is a g such that $g(x)=o$ and $v_{M+f,g}(\phi(x))=1\}$.

(iv) $R^a(f)([\lambda y_1 ... y_{n\geq 2}\ \phi(y_1 ... y_n)])=\{<o_1 ... o_n> \mid$ there is a g such that $g(x_1)=o_1, ..., g(x_n)=o_n$ and $v_{M+f,g}(\phi(x_1 ... x_n))=1\}$.

(v) $R^b(f)(<\lambda Y^1\ \phi(Y^1)>)=f(<\lambda Y^1\ \phi(Y^1)>)$

(vi) $R^b(f)(\beta^1)=<\Gamma, \Delta>$, where $\Gamma=l(f(\beta^1))$ and $\Delta=\{o \mid o=f(<\lambda Y^1\ \phi(Y^1)>)$ for some term "$<\lambda Y^1\ \phi(Y^1)>$", and there is a g such that $g(X)=f(\beta^1)$ and $v_{M+f,g}(\phi(X))=1\}$;

(vii) $R^b(f)([\lambda y_1 ... y_{n\geq 2}\ \phi(y_1 ... y_n)])=f([\lambda y_1 ... y_n\ \phi(y_1 ... y_n)])$.

(The idea is this. In the first stage R^a, we do not change the reference of an individual constant, any extension of a "simple" one-place predicate and the encoding extension of a "compound" one-place predicate, but we change the exemplification extension of every compound n-place ($n\geq 1$) according to the hypothesis f and ($CP_Z{}^{\#}$). Obviously, the result of this operation is also a hypothesis. In the second stage R^b, we do not change the reference of an individual constant and the exemplification extension of every n-place predicate got from the previous stage, but we change the encoding extension of every one-place predicate according to the hypothesis got from the result of the first stage and ($OB_Z{}^{\#}$). The result of this second operation is again a hypothesis.)

We now proceed similarly to what Gupta and Belnap did in (1993). We define $R^0(f)=f$, $R^{n+1}(f)=R(R^n(f))$ for successor ordinal n, and, for limit ordinal p:

(i) $R^p(f)(P^1)=R^p([\lambda y\ P^1(y)])=<l(I(P^1)), \Delta>$,

where $o\in\Delta(\subseteq D)$ ($o\notin\Delta$) iff there is an ordinal number $j<p$ such that, for all ordinal number m, if $j\leq m<p$, then $o\in r(R^m(f)(P^1))$ ($o\notin r(R^m(f)(P^1))$);

(ii) $R^p(f)([\lambda y\ \phi(y)])=<\Gamma, \Delta>$,

where ϕ is not a predicate term in L, $o\in\Gamma(\subseteq D)$ ($o\notin\Gamma$) iff there is an ordinal number $j<p$ such that, for all ordinal number m, if $j\leq m<p$, then $o\in l(R^m(f)([\lambda y\ \phi(y)]))$ ($o\notin l(R^m(f)([\lambda y\ \phi(y)]))$), and $o\in\Delta(\subseteq D)$ ($o\notin\Delta\subseteq D$) iff there is an ordinal number $j<p$ such that, for all ordinal number m, if $j\leq m<p$, then $o\in r(R^m(f)([\lambda y\ \phi(y)]))$ ($o\notin r(R^m(f)([\lambda y\ \phi(y)]))$).

(iii) $R^p(f)([\lambda y_1 ... y_{n\geq 2}\ \phi(y_1 ... y_n)])=\Gamma$,

where $<o_1 ... o_n>\in\Gamma(\subseteq D^n)$ ($<o_1 ... o_n>\notin\Gamma$) iff there is an ordinal number $j<p$ such that, for all ordinal number m, if $j\leq m<p$, then $<o_1 ... o_n>\in R^m(f)([\lambda y_1 ... y_{n\geq 2}\ \phi(y_1 ... y_n)])$ ($<o_1 ... o_n> \notin R^m(f)([\lambda y_1 ... y_{n\geq 2}\ \phi(y_1 ... y_n)])$).

A sequence of hypotheses $S=<R^0(f), R^1(f), ..., R^m(f), ...>$ is a revision sequence for f, iff the length of S, $lh(S)$, is either some limit ordinal or On (= the class of all ordinal numbers). Given a revision sequence S for f of length On, we define "sentence ϕ is constantly true in S" as "there is an ordinal number j such that, for all m≥j, A is true in M+$R^m(f)$". We say that a sentence ϕ is valid in (a base model) M iff, for every hypothesis *f* and every revision sequence S for *f* of length On, ϕ is constantly true in S. Similarly, we say that a sentence ϕ is invalid in M iff, for every hypothesis *f* and every revision sequence S for *f* of length On, $\sim\phi$ is constantly true in S. Sentences that are neither valid nor invalid in a model M are said to be "pathological" in M.

Given any base model M, it can be shown that every classical valid (or contradictory) sentence is still valid (or invalid) in M. Furthermore, it can be shown that "$<\lambda Y\ (\exists z)(Y = [\lambda y\ y = z]\ \&\ {\sim}zY)>[\lambda y\ y = z]$", "$\sim<\lambda Y\ (\exists z)(Y = [\lambda y\ y = z]\ \&\ {\sim}zY)>[\lambda y\ y = z]$", "$<\lambda Y\ Y = [\lambda y\ (\exists X)(Xy\ \&\ {\sim}yX]>[\lambda y\ (\exists X)(Xy\ \&\ {\sim}yX]$", and "$\sim<\lambda Y\ Y = [\lambda y\ (\exists X)(Xy\ \&\ {\sim}yX]>[\lambda y\ (\exists X)(Xy\ \&\ {\sim}yX]$" are all pathological in M. Thought not validating ($CP^{\#}$) and ($OB_Z{}^{\#}$) in their original forms, this revision semantics, nevertheless, validates the inference *rules*:

$$(CP_G)\qquad ([\lambda y_1...y_n\ \phi(y_1...y_n)](\alpha_{1...}\alpha_n))^n \Leftrightarrow (\phi(\alpha_{1...}\alpha_n))^{n-1},$$

and

$$(OB_G)\qquad (\beta^1<\lambda Y^1\ \phi(Y^1)>)^n \Leftrightarrow (\phi(\beta^1))^{n-1}.$$

One apparent problem about the semantics is that it validates every sentence of the form "$<\lambda Y^1\ \phi(Y^1)>\neq<\lambda Y^1\Psi(Y^1)>$", where "$\phi(Y^1)$" and "$\Psi(Y^1)$" are different formulas. Thus, "$<\lambda Y^1\ Y^1=[\lambda y\ y$ is red$]>$" and "$<\lambda Y^1\ Y^1=[\lambda y\ {\sim}y$ is not red$]>$" will always refer to different objects, even if what they refer to may fall into the exemplification and encoding extensions of the same predicates. This does not seem very implausible, but the way to fix it is not very difficult to find. If we want objects exemplifying and encoding the same properties to be identical, we can change the definition of hypotheses a bit so that:

(i') $f_0(<\lambda Y^1\ \phi(Y^1)>)\in D$ such that $f_0(<\lambda Y^1\ \phi(Y^1)>)=f_0(<\lambda Y^1\Psi(Y^1)>)$ iff, for every predicate β^1, $f_0(<\lambda Y^1\ \phi(Y^1)>)\in l(f(\beta^1))$ iff $f_0(<\lambda Y^1\Psi(Y^1)>)\in l(f(\beta^1))$ and $f_0(<\lambda Y^1\ \phi(Y^1)>)\in r(f(\beta^1))$ iff $f_0(<\lambda Y^1\Psi(Y^1)>)\in r(f(\beta^1))$;

and we can expand the two-stage revision function to contain a further stage R^c such that, in R^c, we identify referents of different individual constants with some arbitrarily chosen object among them whenever they are not distinguishable according to the result of the previous two revision stages, R^a and R^b. We then change, in R^c, the exemplification and/or encoding extension of every predicate according to the result of such an identification.

Whether the above semantics is axiomatizable and, if so, how it should be done are still open questions, but it is safe to say that a sound inferential apparatus of it will not be inconsistent.

6. Conclusion

We have seen two ways to drop the *ad hoc* restriction on (CP_Z) within the two-modes-of-predications approaches to Meinongianism. But I believe, though will not give the details here, that similar skills can also be applied when one seeks to release some similar restrictions made within the two-kinds-of-properties approach for Meinongianism. Moreover, it should be obvious that these techniques can further be applied when one is dealing with the naïve Meinongian theory or even (OB_S) plus (CP_S). This suggests that a Meinongian theory doesn't really need to make any distinction about modes of predication or kinds of properties at all; though naïve, the naïve Meinongian theory is good enough to work!

Philosophy Department
National Chung-Cheng University
#168 University Road
Ming-shiung, Chia-yi, Taiwan 621
wang@phil.ccu.edu.tw

References

Castaneda, H. N. (1989). Direct reference, the semantics of thinking, and guise theory. In J. Almog, & J. Perry, & H. Wettstein. *Themes from Kaplan* (pp. 105–44). Oxford/N.Y.: OUP.

Dunn, J. M. (1986). Relevance logic and entailment. In D. Gabbay and F. Guenthner (Eds.). *Handbook of Philosophical Logic*, Vol. 3, Chapter 3. Reidel: Dordrecht.

Gupta, A, & N. Belnap (1993). *The revision theory of truth*. London: MIT.

Gupta, A. (2001). Truth. In L. Goble (Ed.). *The Blackwell guide to philosophical logic* (pp. 90–114). London: Blackwell Publishers Inc.

Hawthorne, J. (2003). Identity. In M. J. Loux, & D. W. Zimmerman (Eds.). *The Oxford Handbook of Metaphysics* (pp. 99–130). New York: Oxford University Press.

Kripke, S. (1975). Outline of a theory of truth. In R. L. Martin (Ed.). *Recent essays on truth and the liar paradox* (pp. 53–81). London: Clarendon Press.

Martin, R. L., & Woodruff, P. W. (1984). On representing 'true-in-L' in L. In R. L. Martin (Ed.). *Recent essays on truth and the liar paradox* (pp. 47–51). London: Clarendon Press.

Parson, T. (1980). *Nonexisting Objects*. New Haven and London: Yale University Press.

Rapaport, W. (1979). An adverbial Meinongian theory. *Analysis, 39*, 75–81.

Shapiro, S. (2001). Classical logic II – Higher-order logic. In L. Goble (Ed.). *The Blackwell guide to philosophical logic* (pp. 33–54). London: Blackwell Publishers Inc.

Zalta, E. N. (1983). *Abstract Objects: An Introduction to Axiomatic Metaphysics*. Dordrecht and Boston: Reidel.

On So-Called Sentences with Category Mistakes

Jan Woleński

It is difficult to give general a definition of sentences considered to contain category mistakes (I will also refer to them as to 'problematic sentences' or 'anomalies').[1] On the other hand, we have a lot of examples, as demonstrated by the following list:

(1) This stone is now thinking about Vienna (Carnap, 1937, p. 5);
(2) Saturday is in bed (Ryle, 1938, p. 179);
(3) Quadruplicity drinks procrastination (Russell, 1940, p. 166);
(4) Caesar is a prime number (Reichenbach, 1947, p. 7)
(5) The theory of relativity is blue (Pap, 1960, p. 41);
(6) Colorless green ideas sleep furiously (Chomsky, 1964, p. 384).

In order to simplify the discussion I omit examples instantiated by set-theoretical or semantic paradoxes as well as those stemming from general philosophical views, like logical empiricists' qualification of metaphysical statements as absurd.[2] All of the above examples have the following features:

(A) They consist of ordinary and meaningful words;
(B) They are grammatically admissible;
(C) They are not figurative nor do they occur in special uses;
(D) They are not ambiguous;
(E) They are perfectly translatable into another languages;
(F) They are felt as odd or anomalous in (almost) every language;
(G) They cannot be converted into normal sentences by removing syntactic errors.

---

[1] I borrow the term 'category mistake' from Baker (1956); see also Erwin (1970, p. 41). There are also other labels, for example, 'confusion of spheres' (Carnap, 1928, pp. 53–58), 'type-differences', 'type-riddles' (Ryle, 1938, p. 179, p. 182), 'type-crossing sentences' (Drange, 1966) and 'selection errors' (McLeod, 2001, p. 9); I follow Drange's book in summing up (with some additions) properties of odd sentences. Page-references are to translations or reprints, provided that they are mentioned in the bibliography.

[2] However, category-mistake arguments are sometimes used in philosophy independently of general views. For instance, the thesis that mental phenomena are caused by physical ones was criticized for linking two fairly different categories of items.

Theses (E) and (F) are perhaps particularly important, because they suggest that the oddness of (1) – (6) is invariant across languages. According to (G) problematic sentences cannot be improved by correcting syntactic mistakes. Polish tolerates sentences with proper names in the predicate position, for instance, 'On jest Janem' (He is Jan), but it can be converted to a proper form 'Jan jest jego imieniem' (Jan is his first name). Another example is provided by ungrammatical sentences corrected by the context, for instance, 'John be tall' understood as 'John is tall'. Also (C) enlightens an interesting point of oddities. If a student says to his friends 'I need a blue relativity theory' he or she can ask for a blue (that, is with covers of that color) textbook on the relativity theory. Similarly, a policeman can say 'number 5 is very hurry' in order to inform his colleague that a followed offender, called 'number five' for brevity is trying to escape. Thus, figurative or special meanings more or less conventionally ascribed to odd sentences very often deprive problematic sentences of their oddities.

The usual, but also very preliminary explanation of what is the oddness in question points out that

(*) entities denoted by the subjects in (1)–(6) cannot fall under the expressions functioning as predicates in these sentences.

Accordingly, 'being a prime number' cannot be predicated of human beings, the days of week are not able to be in beds, etc. However, oddity is not a clear qualification and, what is more important, it has no straightforward logical meaning. Hence, (*) is explicated by assertions stating that (1)–(5) are meaningless, absurd, nonsensical, inconsistent, contradictory or false. The locus classicus of the position that the sentences in question are meaningless (absurd, etc.) is as follows (Pap, 1960, p. 41; this view was also more or less shared by Carnap, Ryle and Russell):[3]

> "The theory of relativity is blue", "the number 5 weighs more than the number 6", "his mind eats fish": these and millions more predications would unhesitatingly be dismissed as meaningless, not false, in spite of their *syntactic* correctness, by plain people.

However, there is a problem how to develop (*) and Pap's diagnosis in order to obtain a satisfactory account of causes of meaninglessness in question. Is it generated by syntactic, semantic, pragmatic or ontological factors? In order to see the difficulties I will review a sample of proposed solutions.[4]

[3] Chomsky argues that (6) is less grammatical than typical English sentences. Since his analysis remains on syntactic level, I will not enter into the Chomskyan theory of degrees of grammaticalness, although I will analyze (6).

[4] I do not pretend to be exhaustive in my survey.

Roughly speaking, the proposals in question can be divided into two groups. The first includes informal attempts, the second appeals to formal-logical constructions.[5] I begin with the former. Husserl (see Husserl, 1921, pp. 67–68) tried to solve the question with his famous distinction of *Unsinn* (a syntactic or grammatical incorrectness) and *Widersinn* (an absurd, nonsense, that is, an expression going against the linguistic sense). Thus, each of (1)–(6) is absurd, although the sequence of words 'the theory of relativity has or' is qualified as syntactically incorrect. According to Ewing and Baker (see Ewing, 1937 and Baker, 1956) problematic sentences lead to contradictions, that is, obvious falsities or evident absurdities. Ryle (see Ryle, 1938–39) argues that predicates occurring are applied to objects which do not belong to the same category or type; this circumstance produces type rules for grammar. Strawson (see Strawson, 1959, p. 101) sees the cause of oddity in the multi-applicability of some predicates, although the results are rather not contradictory, but assume stupid questions. Chomsky (see Chomsky, 1964) suggests special rules of selection, which would decide, for example, that being abstract excludes being animate; thus, ideas as abstract cannot sleep, because this kind of state is reserved to animate beings only. Drange (see Drange, 1966, Chapter 7) considers problematic sentences to be unthinkable and explains the unthinkability property as equivalent to the necessity of falsehood.

Unfortunately, these diagnoses and the resulting salvations are very vague. Although the concept of *Unsinn* is clear and can be explained by a reference to grammatical or even logical rules, the situation with *Widersinn* seems different. It is too easy to identify obvious falsities or absurdities with contradictions (Ewing), we do not know whether types (Ryle), categories (Ryle) or modes of application (Strawson) have logical, semantic or ontological import. The impression of vagueness increases when we ask what is a difference between absurdity and falsity or how to establish rules of selection (Chomsky). It seems that informal solutions rely too much on ordinary intuitions. This is clear, when we look at Ryle's rule that substitutions across various types lead to anomalies as in the case replacing 'day' by 'in bed' in (2). This rule is well-known in the theory of syntactic categories, but assumes a clear formal account of types, although particular outcomes can vary depending on the logical system. For example, Leśniewski's ontology admits proper names as predicated, but it is prohibited in first-order logic. Moreover, we have a serious problem with negations of problematic sentences in some solutions, particularly that offered by Ewing and Drange. I will return to this question later.

[5] Three comments are in order here. Firstly, this division is related to my tasks consisting in using formal semantics in expounding my own offer. Secondly, the borderline between informal and formal proposals is not sharp. In fact, any solution employs some formal tools, although not always taken from logic, because grammatical concepts are also used. Thirdly, both kinds of accounts are mutually related. For example, Husserl anticipated Pap's idea.

Formal-logical solutions are either conservative over classical logic or revisionist with respect to the standard system. Russell, Carnap, Reichenbach and Pap wanted to keep classical logic. They combine this move with some insights taken from the theory of logical types. The main idea consists in ascribing, to use Pap's apt wording, the range of significance to every predicate. This range can be considered as the logical type associated with a given predicate and decides what can be substituted for it; the theory of logical types justifies the rule used by Ryle and surely influenced his proposal. Russell, Carnap and Reichenbach, following the line of *Principia Mathematica*, adopted the view that ranges of significance of predicates are limited (restricted) to certain classes. If a statement is coherent with the ranges of its predicate, it is counted as meaningful and thereby true or false, otherwise it must be qualified as neither true nor false. Thus, examples (1) – (5) (I leave (6) for later consideration) as meaningless. Consequently, negations of meaningful sentences are meaningful and denials of meaninglessness sentences are meaningless.[6] Thus we have the following very important principle

(**) (a) if A is meaningful, its negation $\neg A$ is meaningful as well;
(b) if A is meaningless, its negation $\neg A$ is meaningless as well.

This principle is justified by an obvious observation that inserting the sign of negation before an expression cannot change its status as meaningful (meaningless). Of course, (**) can be generalized to

(***) meaningfulness (meaninglessness) of an expression E is invariant with respect to proper operating logical constants on E.

The clause of proper operating is dictated by the fact that arbitrary operating can lead to the syntactic incorrectness of an expression, like in the case inserting 'or' after 'John is'. Now we can return to the problem of negations of problematic sentences, according to the solutions of Ewing and Drange. Assume that A is an anomaly and necessarily false. Extending (***) to the modal status of a sentence, we obtain that A is necessarily true. Although this conclusion can be defended, it must meet an objection that any meaning of 'necessary sentence', except 'logically true sentence' is very obscure and has to be carefully explained.

Pap proposed a different solution. For him, the range of significance of a given predicate is unrestricted, unless the theory of logical types (**TLT** for brevity) decides differently. We touch here the problem of how **TLT** should be understood. Originally, it was invented to solve antinomies, set-theoretical, for example, the Russell paradox or semantic, for example, the Liar paradox.

[6] Sommers (see Sommers, 1982, p. 297) introduced the related idea of nested contrariety.

Leaving aside that the former paradoxes are eliminated by the simple **TLT**, but the latter require the ramified **TLT**, we can say that the construction offered by Russell and Whitehead in *Principia* was well motivated by (almost) pure logical reasons.[7] It is difficult to see that an extension of **TLT** to sentences (1)–(5) preserves its logical status. All these examples are not genuine paradoxes, but just anomalies or linguistic oddities. Thus, if **TLT** qualifies the expressions '$X \in X$' (where X is a set) as nonsense, because no set belongs to itself, (1)–(5) cannot be considered in the same way. In fact, proponents of extending **TLT** to category-mistake sentences supplement logical reasons but additional grounds; in particular, they appeal to experience, the possibility of verification, etc. Pap disagrees with such moves and suggests that problematic sentences are meaningful and false. Thus, according to (**) their negations are meaningful and false. His example, that is (5), is analyzed in such a way. Since the theory is relativity is not an object which could be blue, (the sentence (5) is empirically false, but the sentence 'the theory of relativity is not blue' is true. 'Could be' is understood as an empirical assertion, not as logical impossibility.

Now I pass to the revisionist solutions. Halldén (see Halldén, 1949) maintains that three-valued logic is unavoidable for coping with anomalies. Goddard and Routley (see Goddard & Routley, 1973) represent a kind of two-dimensonalism. According to them, two-valued logic is enough for the distinction of truth and falsehood, but three-valued logic should be the formal frame for the division of sentences into significant (meaningful) and non-significant (meaningless). Still another possibility consists in regarding problematic sentences as true-value gaps, similarly to the Liar paradox and the like (see Martin, 1970, although it is not directly applied to anomalies but proposes a category cum truth-value gaps solution to the Liar). I cannot enter here into the formal details of particular revisionist constructions.[8] Let it be sufficient to point out that the standard way is to take 'non-significance' as the third value (neutrum). Accordingly, anomalies are valued as neutral, and the same concerns their negations, according to (**). The first objection points out that there are many three-valued logics and it is unclear which should be selected to be the proper logic of significance. A special problem arises with compound sentences consisting of significant and non-significant parts in various combinations. I will limit my analysis to implications evaluated by Łukasiewicz's matrices for three valued logic with symbols t (truth), f (falsehood) and n (neutrum). Take the following examples:

---

[7] The qualification 'almost' is caused by a well-known discussion concerning the status of the axioms of infinity and reducibility in **TLT**.

[8] Let me add that Halldén's system did not enjoy a major acceptance. On the other hand, Goddard and Routley probably became dissatisfied with their proposals, because they never wrote the second volume of their study. Hence, I will not refer to Halldén or Goddard and Routley even in later illustrations.

(7) if Caesar is a prime number, then the theory of relativity is blue;
(8) if 2 + 2 = 5, then Caesar is a prime number;
(9) if Caesar is a prime number, then 2 + 2 = 4;
(10) if 2 + 2 = 4, then Caesar is a prime number;
(11) if Caesar is a prime number, then 2 + 2 = 5.

The evaluations are (justifications are in brackets): t for (7) ($n \Rightarrow n = t$), t for (8) ($n \Rightarrow n = t$), t for (9) ($n \Rightarrow t = t$), f for (10) ($t \Rightarrow n = n$) and f for (11) ($n \Rightarrow f = f$). Although the first evaluation seems quite plausible as realizing the intuition that an implication with anomalies in the antecedent and consequent is itself anomalous, the rest is fairly questionable. It seems that ordinary intuitions will qualify all (7) – (11) as problematic sentences; the same qualification can be also suggested for considering anomalies as truth-values gaps. However, one can argue that the oddity of (9) is of a similar kind as in the case of the sentence:

(12) if Caesar is a catholic priest, then 2 + 2 = 4.

I do not deny that we can perhaps imagine a possible world in which Caesar is a catholic priest much better than the situation in which he is a prime number, but it seems that ordinary conditionals with obviously false antecedents are equally odd. This is an interesting point, because it shows that the sentence 'Caesar is a catholic priest' functions differently for itself than in its role as the antecedent of (12). Even when (9) and (12) were qualified as unequal with their oddity, the thesis that all compound sentences with anomalous parts are neutral or represent truth value gaps entails that a special logic is unnecessary for them. Still we have a problem with some inference patterns. Consider the following reasoning:

(****) if Caesar is a prime number, the theory of relativity is true;
if the theory of relativity is blue, then Saturday is in bed;
____
if Caesar is a prime number, then Saturday is in bed.

According to the decision that all compounds are neutral, we should disqualify (****) as a correct inference or decide that n is a distinguished value in such cases. On the other hand, logicians very often explain the validity of inferences as independent of the meanings of expressions occurring in premises and conclusions. Thus we should not banish patterns like (****) from the scope of logic.

My view is that the problem cannot be solved in natural language in a fully universal manner. I do not deny that we have quite definite intuitions concerning problematic sentences and commonly qualify them as plain oddities without any hesitation. On the other hand, our feelings are fairly changeable, depending on various more or less accidental circumstances, for example, the development of science and technology, or accepted world-views. Consider the sentences:

(13) computers think about themselves;
(14) it is possibly to establish whether two physical events are simultaneous, independently of the frame of reference.

Now (14) was perfectly meaningful around 1900, but everybody would consider (13) an oddity about 50 years ago, but the situation is different today. Although we can also say that (13) and (14) are true in some possible worlds, physical objects, according to the theory of special relativity, are not items which can be considered as simultaneous independently of their frames of reference. Thus, ordinary linguistic intuitions concerning the oddity of problematic sentences vary and occur with different degrees of intensity.

If semantics pretends to universality, something must be done with problematic sentences, because they are composed from usual names and predicates which are perfectly comprehensible, when they occur outside of problematic contexts. I will follow Pap's suggestion that anomalies are meaningful, which means that some are true, others false, but there are no sentences without logical values. I will not appeal to logical types or ranges of significance of predicates, but to typical contemporary devices of formal semantics. I make the following general assumptions:

(I) we are working with formalized languages;
(II) syntactic incorrectness is decided by recursive rules;
(III) all qualifications with respect to truth, falsehood and meaningfulness are related to interpreted formal (well-specified) languages;
(IV) matters of truth and falsehood are semantic;
(V) matters of meaningfulness and meaninglessness are pragmatic;
(VI) matters of truth/falsehood and meaningfulness/meaninglessness taken together are a fusion of semantics and pragmatics.
(VII) if one says that a sentence is odd, although grammatically correct, one understands it (the Ingarden rule; see Ingarden, 1936)[9];

In general, these assumptions favor semantics as prior to syntax. The relation between semantics and pragmatic will be considered later.

Let **L** be a formal interpreted first-order language. We can think about it as a logical representation of a portion of natural language. Because **L** is interpreted its expressions are equipped with some interpretations. In particular, we have

[9] Otherwise speaking, the Ingarden rule states that if one says that a syntactically correct sentence is an anomaly, it impossible without understanding it as such. This rule can be viewed as an interpretation of Husserl's concept of *Widersinn.*

some domain **D** as the universe for quantifying, distinguished objects as denotations of constants of **L** and subsets of **D** as denotations of predicates (for simplicity, we assume that **L** has monadic predicates only). Consider a simple language with logical constants, one term 'Caesar' and two predicates 'is a prime number' and 'is a human being'. The interpretation is given by ($\mathbf{c} = \mathbf{d}$('Caesar'), $\mathbf{U_1} = \mathbf{d}$('is a human being'), where $\mathbf{U_1} \subseteq \mathbf{U}$, $\mathbf{U_2} = \mathbf{d}$('is a prime number'), where $\mathbf{U_2} \subseteq \mathbf{U}$)

(#) $\quad \mathbf{M} = <\mathbf{U}, \mathbf{c}, \mathbf{U_1}, \mathbf{U_2}>$.

Clearly, sentence (i) 'Cesar is a human being' true, because $\mathbf{c} \in \mathbf{U_1}$, but sentence (ii) 'Ceasar is a prime number' is false, because $\mathbf{c} \notin \mathbf{U_2}$. Respectively, the negation of (i) is false, but the negation of (ii) is true.

If we want to keep classical logic, there is only one way to deal with problematic sentences, namely by regarding them as false.[10] Accordingly, all atomic anomalies are false, their negations are true, and compound sentences with problematic parts are evaluated by standard rules for connectives and quantifiers. Now the question arises whether anomalies are 'normally' false or 'extra' (for example, obviously) false. The answer is that they are normally false from the semantic point of view, because their negations are normally true; in particular, we have no doubts that sentence (iii) 'it is not the case that Caesar is a prime number', is true. How to interpret the oddity of (ii)? A simple route is to say that the interpretation of this sentence is anomalous. Since **L** is interpreted, we must admit that

(#') $\quad \forall \mathbf{t}\, \exists \mathbf{U_i} \subseteq \mathbf{U}(\mathbf{d}(\mathbf{t}) \in \mathbf{U_i})$.

This means that every term has a denotation which belongs to the denotation of some predicate (recall that first-order logic does not admit empty terms, that is, void proper names). Divide interpretations of predicates $\boldsymbol{P_1}$ = 'is a human being' and $\boldsymbol{P_2}$ = 'is a prime number' into standard ($\mathbf{I^S}$) and non-standard ($\mathbf{I^{NS}}$). Let $\mathbf{U_1}$ be the set of human beings and $\mathbf{U_2}$ be the set of prime numbers. We say that $\mathbf{I^S}(\boldsymbol{P_1}) = \mathbf{U_1}$ and $\mathbf{I^S}(\boldsymbol{P_2}) = \mathbf{U_2}$. We can invent various non-standard interpretation of the predicates $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$. In particular, we can define $\mathbf{I^{NS}}(\boldsymbol{P_1}) = \mathbf{U_1} - \{\mathbf{d}(\text{'Caesar'})\}$ and $\mathbf{I^{NS}}(\boldsymbol{P_2}) = \mathbf{U_2} \cup \{\mathbf{d}(\text{'Caesar'})\}$. Keeping the assumption that $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ are the only predicates of our language, (#') and the principle of bivalence entail

(#'') (a) $\forall \mathbf{t}(\mathbf{d}(\mathbf{t}) \in \mathbf{I^S} \vee \mathbf{d}(\mathbf{t}) \in \mathbf{I^{NS}})$;
(b) $\forall \mathbf{t} \neg(\mathbf{d}(\mathbf{t}) \in \mathbf{I^S} \wedge \mathbf{d}(\mathbf{t}) \in \mathbf{I^{NS}})$.

[10] This intuition is shared by some mathematicians. For example, Exner and Rosskopf (see Exner & Rosskopf 1959, p. 128) consider the sentence 'The Eiffel tower is a man' as false.

This means that every object belongs to an interpretation, standard or non-standard, but no object belongs to an interpretation which is both standard and non-standard.

The adopted non-standard interpretation of $\boldsymbol{P_2}$ validates the sentence 'Caesar is a prime number'. Speaking otherwise, a sentence seems anomalous if it occurs as atomic under a non-standard interpretation, which consists in adding Caesar to prime numbers in this case. From a purely semantic point of view, there is absolutely nothing wrong with the equality $\mathbf{I}^{\mathrm{NS}}(\boldsymbol{P_2}) = \mathbf{U_2} \cup \{\mathbf{d}(\text{'Caesar'})\}$. It is just as good a set as $\mathbf{U_2}$, because any set is constituted by its members and nothing more. The only objection can come from pragmatics. In fact, the equality in question combines semantics, that is a mapping from $\boldsymbol{P_2}$ to $\mathbf{U_2} \cup \{\mathbf{d}(\text{'Caesar'})\}$ with pragmatics, indicated by the phrase $\mathbf{I}^{\mathrm{NS}}(\boldsymbol{P_2})$. I do not want to decide whether every non-standard interpretation is odd. Perhaps some are not, for example, those related to alternative worlds admitted by scientific theories. However, I see no way to describe sentences as problematic, anomalous or odd without taking into account pragmatic marks. Now one can say that we come back to informal solutions of category mistakes in sentences. Yet this remark is only partly correct, because this informal feature is correlated with formal semantic machinery. What is the basis for dividing interpretations into standard and non-standard? Well, the simplest answer is naturalistic: we learn and create languages as evolutionary devices and this causes our linguistic intuitions to select what is standard and non-standard.

Finally, let me return to sentence (6). It requires a more formal translation, for example, as follows:

(15) for any x (if x is a colorless green idea, then x sleeps furiously).

Since the predicate 'is a colorless green idea' denotes the empty set under the standard interpretation, (15) is vacuously true in this interpretation, independently of whether sleeping can be furious or not. Yet (15) can be false under non-standard interpretations, if one accepts that there are colorless green ideas, but furious sleeping is impossible. Hence, (15) is not tautological. As a result, we have that compounds with anomalies can be true. This is the cost of the proposed interpretation, although fairly comparable to the paradoxes of truth-functional connectives.

Institute of Philosophy
Jagiellonian University
Kraków, Poland
wolenski@theta.uoks.uj.edu.pl

References

Baker, A. J. (1956). Category Mistakes. *Australasian Journal of Philosophy XXXIV*, 18–33.

Carnap, R. (1932). *Der logische Aufbau der Welt*. Berlin: Weltkreis-Verlag. English translation (1967). *The Logical Structure of the World and Pseudoproblems in Philosophy*, translated by R. A. George. London: Routledhe & Kegan Paul.

Carnap, R. (1937). Testability and Meaning. *Philosophy of Science IV* (pp. 1–40).

Chomsky, N. (1961). Degrees of Grammaticality. In J. A. Fodor, & J. J. Katz (Eds.). *The Structure of Language. Reading in the Philosophy of Language. A Philosophical Approach to Language based on Selected Articles Delling with the Theories, Methods and Results of Linguistics* (pp. 384–389). Englewood Clifts: Prentice-Hall.

Drange, T. (1966). *Type Crossings*. The Hague: Mouton.

Erwin, E. (1970). *The Concept of Meaninglessness*. Baltimore: The John Hopkins Press.

Ewing, A. C. (1937). Meaninglesness. *Mind XLVI* (pp. 347–364). Reprinted in P. Edwards, & A. Pap (Eds.) (1957). *A Modern Introduction to Philosophy. Readings from Classical and Contemporary Sources* (pp. 770 –779). New York: The Free Press.

Exner, R. M., & Rosskopf, M. F. (1959). *Logic in Elementary Mathematics*. New York: McGraw-Hill.

Goddard, L., & Routley, R. (1973). *The Logic of Significance and Context, vol. I.* Edinbourgh: Scottish Academic Press.

Halldén, S. (1949). *The Logic of Nonsense*. Uppsala: Uppsala Universitets Årrskrift.

Husserl, E. (1921). *Logische Untersuchungen, B. 2, Zweite Auflage*.Halle: Niemeyer. English translation (1970). *Logical Investigations*, translated by J. N. Findlay. London: Routledge and Kegan Paul.

Ingarden, R. (1936). Der logistische Versuch einer Neugestaltung der Philosophie. Eine kritische Bemerkung. In *Actes du Huitième Congrès Internationale à Prague, 2–7 Septembre 1934* (pp. 203–208). Prague: Orbis.

Martin, R. L. (1970). A Category Solution to the Liar. In R. L. Martin (Ed.). *Paradox of the Liar* (pp. 91–112). Reseda: Ridgeview.

McLeod, S. K. (2001). *Modality and Anti-Metaphysics*. Aldershot: Ashgate.

Pap, A. (1960). Types and Meaninglesness. *Mind, 69*, 41–54.

Russell, B. (1940). *An Inquiry into Meaning and Truth*. London: George Allen and Unwin.

Reichenbach H. (1947). *Elements of Symbolic Logic*. New York: The Macmillan Company.

Ryle, G. (1937–1938). Categories. *Proceedings of the Aristotelian Society XXXIX* (pp. 281–298). Repr. in G. Ryle, *Collected Papers, vol. 2, Collected Essays 1929–1968* (pp. 170–184). Bristol: Thoemmes.

Sommers, F. (1982). *The Logic of Natural Language*. Oxford: Clarendon Press.

Strawson, P. F. (1959). *Indiviudals. An Essay in Descritptove Metaphysics*. London: Methuen.

The Logica Yearbook 2006

Editoři Ondřej Tomala a Radek Honzík

Technická redaktorka Marie Vučková

Návrh obálky Marta Bílková

Typografie a sazba Martin Pokorný, Praha, Matúškova 804,

a Jana Andrlová, Mělník, 5. května 111

Vydal Filosofický ústav AV ČR, Jilská 1, 110 00 Praha 1

ve svém nakladatelství

FILOSOFIA

nakladatelství

Filosofického ústavu

AV ČR

ΦΙΛΟΣΟΦΙΑ

jako svou 225. publikaci

Vytiskl PB tisk Příbram

Vydání první

Stran 304

Praha 2007

First e-book edition

Prague 2019