HUNGARIAN PHILOSOPHICAL REVIEW

MAGYAR FILOZÓFIAI SZEMLE

2010/4 (54. évfolyam)

A Magyar Tudományos Akadémia
Filozófiai Bizottságának folyóirata

Imre Ruzsa—
A Man of Consequence

Contents

PHILOSOPHY OF MATHEMATICS

FOUNDATIONS OF SCIENCE

Foreword

This special issue of the Hungarian Philosophical Review is dedicated to the memory of our friend, colleague and teacher Imre Ruzsa (1921–2008). Ruzsa was the father of modern philosophical logic in Hungary, the founder and the first chair of the Department of Logic at Eötvös University. His professional interests centered around modal logic, intensional logic, modeling natural language in systems of intensional logic, and the foundations of logic and mathematics. He always thought of his generalization of A. N. Prior's concept of semantic value gaps to quantified, intensional and type-theoretic systems as his most important contribution to logic. He was the author of three books in English (*Modal Logic with Descriptions*, The Hague, 1982, *Intensional Logic Revisited*, Budapest, 1991, *Introduction to Metalogic*, Budapest, 1993), several monographs and textbooks in Hungarian, and many articles in leading logic journals.

The variety of the papers included in this volume represents the range of topics that Ruzsa's research covered: philosophical logic, formal semantics, the philosophy of mathematics, and foundational studies. The papers were all given at the Imre Ruzsa Memorial Conference "Logic, Language, Mathematics", held at the Philosophy Institute of Eötvös University in Budapest on September 17–19, 2009, as part of the annual conference series "Language, Understanding, Interpretation."[1] It is worth mentioning the invited speakers who gave memorable presentations that are published elsewhere: Rob Goldblatt (Victoria University, Wellington), Ági Kurucz (King's College London), Mihály Makkai (McGill University, Toronto), László Pólos (University of Durham, UK), and Anna Szabolcsi (New York University).

Among the authors the reader will find three generations of logicians: fellow researchers whose work Ruzsa's ideas influenced strongly; former students of Ruzsa, who had once been introduced into the mysteries of the logical connectives in seminar rooms of Eötvös University, and many of whom are today lead-

[1] For more information about the conference, see: http://phil.elte.hu/ruzsaconf.

ing researchers in their field, teaching across the globe; and a younger generation of logicians, Ruzsa's "grandchildren", whose work testifies to the enduring influence of Ruzsa's legacy.

The editors would like to thank all the authors for their contributions, among them Zsófia Zvolenszky, whose language editing work has also been invaluable.

András Máté and Péter Mekis

András Máté

Imre Ruzsa—A Man of Consequence

Abstract. The singular aim and task of this paper is to present an overview of the life and work of Imre Ruzsa.

1 AN UNUSUAL ROUTE TO PHILOSOPHY

Ruzsa's life was rather different from a typical academic career. He was born on the 12th of May in 1921 in Budapest but grew up in a little town in the south-eastern part of Hungary as the son of a tailor. His family couldn't send him to high school, so after finishing elementary school, he worked as an assistant to his father. At the age of seventeen, he left his father's house and worked as a tailor's assistant first in Debrecen, a city in Eastern Hungary, then in Budapest. He joined to the Social Democratic Party once in Budapest, and was admitted to the illegal Communist Party. He worked as the printer for the party newspaper and for this activity, he was convicted to eleven years in prison in 1942. The first time he heard about mathematics beyond the common arithmetical operations was from a Communist economist in the prison courtyard. Like other political prisoners, he was sent to the front lines in a forced labour company in 1944. He escaped and survived the rest of the war with false documents in Budapest.

After the war, he finished high school on an accelerated track, then in 1947, he began his university studies at the Faculty of the Humanities of Budapest University. He attended diverse lectures on philosophy and Hungarian linguistics. After the reorganization of university programs in 1950, he became a student of mathematics, physics and descriptive geometry at the new Faculty of Science. By taking a look at his university records, we can note the intellectual level of the Mathematical Institute of Budapest University in those years: in fact, to this day, all professors listed there are regarded as prominent figures of the history of modern mathematics. At the very least, let me mention the name of Lipót Fejér, Rózsa Péter and Alfréd Rényi. Ruzsa's academic records indicate that he fulfilled the requirements of this institute with flying colors. In personal

conversations, he himself remarked that his mathematical studies allowed him to keep a bit more of a distance from politics at that point. But politics and history didn't release him: in the year 1953, a few weeks before Stalin's death, he was arrested again and sentenced to five years for "war crimes and crimes against the people". He was set free in the next year and rehabilitated in 1957.

In 1956, he finished his unversity studies and began teaching mathematics in a geological polytechnic. In 1960, he was invited to the University as a lecturer of mathematical calculus. From 1962 on, he taught mathematics for philosophy students. In these years, he began his research into modal and deontic logic. Earlier, while still a student, he had studied mathematical logic with Rózsa Péter,by this time, they were at the same department—Rózsa Péter read Ruzsa's writings in the sixties and gave him extensive advice and comments, especially on the philosophy of mathematics and on the mathematical aspects of logic. Ruzsa was in contact with the other great master of mathematical logic in Hungary, László Kalmár, too, but in philosophical logic, Ruzsa had no ancestors and mentors in Hungary at all, nor did he have any opportunities to study abroad either. He simply used the library and began corresponding with Arthur N. Prior, whose ideas had the greatest influence on him.

2 EARLY WORK IN PHILOSOPHY

In the year 1965, the Philosophy Institute of the Hungarian Academy of Sciences offered Ruzsa a job as a research fellow. He accepted it but kept his position at the university as a part-time assistant professor. During the sixties, he wrote several pieces about the philosophy of mathematics: a book for teachers (1967), an article in the Hungarian Philosophical Review, a series of articles for scientists and science teachers, remarkable lecture notes on mathematics for students of philosophy (1964), as well as a book for the broader public (1968). Ruzsa had many popular writings on mathematics and often connected the popularization of mathematics with philosophy. He was awarded the Manó Beke prize for popularizing mathematics in 1971.

A second group of his early papers consists of eight survey papers about contemporary research in philosophical logic for the Hungarian Philosophical Review. It was not merely academic reasons that led him to write the last three of these about research on symbolic logic in the Soviet Union. Ruzsa needed to prove that modern logic in philosophy didn't threaten the ideological foundations of the Communist regime. He presented what was effectively an argument from authority, showing that symbolic logic was an accepted research area in the Soviet Union. Investigations by Soviet logicians at that time, like Vladimir Smirnov or Aleksandr Zinoviev, who was later to become a political dissident, and some others, were carried out in accordance with logical research at leading

Western universities. Within Hungarian philosophy, during the sixties, Ruzsa stood almost alone with his research program. It would be unjust to deny that there were some philosophers who tried to integrate some tools and ideas of modern logic into university education and research; but mostly these efforts led to no more than a confused mixture of modern and obsolete ideas. In the realm of education, it was Sándor Szalai who did the best work in terms of integrating some modern logic into the logic curriculum at Budapest University during the forties and fifties; but the effects of his work were limited because Szalai was not a logician, not even a philosopher, but a sociologist who knew a fair amount about logic. The situation was paradoxical because in mathematical logic, Hungary had an abundance of great scholars, such as László Kalmár, Rózsa Péter and their numerous students. It was Ruzsa's mission to convey their knowledge to philosophy.

The third group of papers includes the first results of his own research in logic. He focused on two topics: deontic logic and the connection between logic and probability theory. His doctoral thesis at the Hungarian Academy of Sciences (the degree was called "candidate of mathematical science") was based on the latter topic. The title was "Random models of logical systems"; it was prepared without an official supervisor and Ruzsa mentions in the documents no mentor or advisor in Hungary. Ruzsa didn't subsequently return to this topic. Another branch of his early work was, however, the very beginning of a continuous line of research for the decades to come. Deontic systems are in fact special cases of modal logic; Ruzsa's survey papers from the same time display his interest into general modal logic and Kripke semantics. The idea of semantic value gaps which turned out to be the central thought of his logical work emerged during this time from his study of Arthur Prior's work and from correspondence with him.

3 AT THE DEPARTMENT OF LOGIC

In 1970, both the structure of the departments and the curriculum for philosophy students was reorganised at Eötvös University in Budapest. Mathematics was banned from the curriculum, but Ruzsa received a new task: he joined in the teaching of logic. The newly founded Department of Logic was in charge of the course of study in logic, which consisted of two main components up until the transition period in 1989-90: two or three semesters of formal logic and two semesters of dialectical logic. The basic principle was that the true logic of Marxist-Leninist philosophy was dialectical logic and formal logic was just a subordinated preliminary study to it. Dialectical logic meant, according to the head of department, a sort of materialistically transformed Hegelian logic; it in fact required no more formal logic than a minimal knowledge of Aristotelian syllogistic.

However, this situation made it possible to teach some real logic to young philosophers as long as one resigned oneself to steering clear of questioning the superiority of dialectical logic. Ruzsa published the first version of his lecture notes in logic in 1969, even though he taught the lectures on logic only the next year, and accepted the invitation to the Department of Logic as an associated professor in 1971. This was a great turn both in Ruzsa's life and in logic education.

There was a threefold difference between earlier "formal" logicians at Budapest university and Ruzsa. Firstly, he had the requisite mathematical background to follow contemporary research and contribute to it. Secondly, he didn't bother with improving and modernising old teaching materials and curricula but wrote a completely new one built on modern logic (and improved it over the next thirty years).[1] Thirdly, he didn't go into discussions about what real dialectical logic was supposed to be. Other people in Hungary, as well as in other Eastern-block countries, tried to sell under the name "dialectical logic" some more or less modern methodology of science and were therefore drawn into conflicts with Hegelian dialectical logicians. Ruzsa didn't interfere with the affairs of dialectical logicians. Instead, he responded in sarcastical short articles when mathematical logic was attacked for sneaking antidialectic, metaphysical, neopositivistic etc. ways of thinking into Marxist philosophy, charges brought on by people who had no real knowledge about the subject. As he was a fellow at a department led by the most militant dialectical logician, he didn't expect anything more in those years than that they leave him to work and teach.

In spite of the often astonishing circumstances, the seventies were fruitful years for Ruzsa both in terms of research and teaching. In modal logic, he generalized Prior's idea of truth value gaps to semantic value gaps and on this basis, he elaborated a Kripke-style semantics for various systems of first-order modal logic. His first paper about these systems was his (1973b). His dissertation based on this research, entitled *Individuals in modal logic*, earned him the degree "Doctor of Philosophical Science" at the Hungarian Academy of Sciences. The expanded English version of the dissertation was published by Martinus Nijhoff Publishers (1981). Ruzsa was appointed full professor in 1978.

In terms of teaching logic, beyond the lectures for philosophy students, Ruzsa was given the task of teaching logic and mathematics for students in theoretical linguistics. There was a lucky coincidence between this task and his new interest in the logical modeling of natural languages. Ruzsa became acquainted with Montague semantics in mid-seventies and immediately began to investigate how the idea of semantic value gaps might be implement into Montague grammar. This idea led to more substantial changes in Montague semantics than

[1] Let us emphasize among the different versions the legendary "three-volume one" (1973).

in Kripke semantics but also proved to be even more fruitful.[2] For these investigations and for his teaching activity, he received recognition within a circle of younger linguists and some of them joined him as personal students, participants and guest speakers at his seminars in the seventies and eighties.

In 1977, an opportunity arose to expand the group of "formal logicians" within the Department of Logic with two new lecturer appointments. Ruzsa planned to orchestrate the celebration of the 100th birthday of symbolic logic (the centenary of Frege's *Begriffschrift*), and as a first task, he assigned to one of the new lecturers (namely, me) the translation of a selection from Frege's writings. This volume was published with a slight delay, in 1980, accompanied by an issue of the Hungarian Philosophical Review which contained numerous papers on Frege and modern logic and translations of Frege's articles "The Thought" and "Negation". It was only Ruzsa's own extensive programmatic article on Frege and the importance of modern logic for philosophy (1979) that was published exactly for the centenary in the Hungarian Philosophical Review. That is, the Review did not undertake to publish a special Frege-issue, as Ruzsa's original intention had been. Nevertheless, the centenary of the *Begriffsschrift* was an important step towards the formation of Ruzsa's school. The work of Frege offered a common starting point for the areas where Ruzsa's writings and educational activity gained influence over the previous years: logic, philosophy of mathematics, linguistics, philosophy of language. The works published for this occasion therefore reached all the actual and potential students of Ruzsa and sympathizers of his work.

Moreover, this time, Ruzsa could appear in public together with some of his students and members of his circle. In Hungarian philosophy, symbolic logic had often been associated with logical positivism (not only in the era of Marxism-Leninism, but earlier, too). When Ruzsa, as a modern logician, was accused of smuggling neopositivistic influences into Marxist philosophy, he found this charge was awkward and at a the same time, also amusing. For he had no special sympathy for logical positivism at all. Carnap belonged, of course, to the most widely cited authors in his monographs, but mostly it was not out of Ruzsa's agreement with Carnap's claims. Ruzsa was much more inclined towards realism; he was not a Frege-type Platonist, but his position was closer to Frege than to Carnap. This way,, the centenary celebration also offered an opportunity to present as the founding father of symbolic logic a thinker as far from neo-positivism or any sort of positivism as Frege was.

[2] His first publication on this area was (1980).

4 THE NEW DEPARTMENT

Through the eighties, the Ruzsa school flourished. There was a favourable turn of circumstances: in 1982, the unwanted marriage with dialectical logic could be broken off and a new "Group for symbolic logic and the methodology of science" was founded, headed by Ruzsa. It was an odd, uncoventional unit that was subordinated to no departments but to the institute of philosophy ("Marxism-Leninism") only; yet it didn't have the rank and rights of a department until 1984. The new department began to publish a yearbook called *Tertium non datur*. Besides the members of the department and Ruzsa's PhD students, several linguists, philosophers and other scholars wrote articles and reviews for the *Tertium*; its table of contents showed that Ruzsa and his circle were now gaining considerable influence in the humanities as well, among people interested in modern methodology.[3] In the yearbook we could now break with the earlier strategy of keeping distance from debates; by this time, it contained several sharply critical papers. In the opening volume, Ruzsa and five of his younger colleagues published a paper that dissected a logic textbook that was in use at teacher-training colleges and unified obsolete ideas of traditional logic with dialectic materialistic slogans. At times, the analysis would change into satire. We had considerable fun putting together this critique, but didn't foresee the consequences of it: during the next academic year, the textbook was withdrawn by the ministry of education. But to tell the truth, we didn't yet gather up the confidence to criticize dialectical logic.

Perhaps this is the time to say something about Imre himself as a man. It is not easy because his personality was rather hidden. Autonomy and steadfastness were among his major traits. He had chosen a path for himself and nobody could divert him from it. He wasn't interested in success, praise or money; he did what he thought was the right thing to do and that was all there was to it. He was very helpful. I think most of the colleagues who knew him are indebted to him, but only few of us can claim to have had a truly personal conversation with him. He endured the humiliating situations that occured at the Department of Logic with calm irony and on rare occasions, with sarcastic remarks—it was only by the end of the nineties that I understood how deeply he was insulted by them, when we compiled a repertory volume from the volumes of *Tertium non datur* and he wanted to devote two pages of a four-page foreword to this topic. I needed hours to convince him that Comrade Erdei (the head of that department) didn't deserve so much attention any more. Well, his good sense of humour and irony

[3] Let me illustrate this influence by quoting the names of the Hungarian authors of *Tertium*: András Bárd, Katalin Bimbó, István Bodnár M., Balázs Dajka, Katalin É. Kiss, Özséb Horányi, Márta Fehér, László Kálmán, Ferenc Kiefer, Gyula Klima, Imre Komlósi L., András Kornai, Judit Maár, Anna Madarász Zsigmond, Márta Maleczki, András Máté, Tamás Mihálydeák, Sándor Ónódy, Kornél Solt, Anna Szabolcsi, Zoltán Szabó [Gendler], Tibor Szécsényi.

often helped him in difficult situations. He was a quiet person, never a loud word even if he was angered. On the other side, during the occasional relaxed moment, he liked to make jokes. In the eighties, he wrote a "Dictionary for patho-logicians". Its entries contain "explanations" of logical notions that often mix wordplay with pin-pricks at colleagues and profound remarks. Unfortunately, most of them are basically untranslatable wordplays in Hungarian, but let me quote one that (hopefully) works in English as well:

Inconsistency: a heavy and contagious disease. Especially widespread among philologists. The reduction of texts is the only cure.

In the eighties, Ruzsa published two large monographs. The first was *Classical, modal and intensional logic* (1984). It contained less technical details but a thorough analysis of the philosophical literature on logic, especially on the logic of modalities. In this book, Ruzsa explored the philosophical motivations behind his logic with semantic value gaps and gave in-depth arguments about its advantages. The second book is the two-volume *Logical Syntax and Semantics*—volume I: (1988), volume II: (1989), in which the author gives a self-contained, comprehensive introduction to logical theory, together with the foundations of logical syntax and completing his survey with a description of a formalized fragment of Hungarian. This is the main work of Ruzsa; I shall say more in a bit about its first, metalogical chapter.

5 THE LAST YEARS

Ruzsa retired from professorship and from chairing the department in 1990. This unavoidable step together with the fact that some members of the department along with other colleagues from the Ruzsa-circle went on to pursue their careers abroad, at outstanding universities—which was otherwise very much a welcome fact—made the activities of the department somewhat more difficult and less effective. On the other side, at the end of the eighties began our cooperation with the Algebraic Logic department of the Alfréd Rényi Mathematical Institute of the Academy, which made it possible to start the Logic Graduate School, one of the very first graduate programs in Hungary. Although formally Ruzsa was not the leader of this graduate school, he did play a definitive role in its first years, up until the end of the 1990s. He published improved English versions of the two most important chapters of *Logical syntax and semantics*: *Intensional logic revisited* (1991) and *Introduction to metalogic* (1997). In 1991 he was awarded the Széchenyi Prize, the highest state honour for achievements in science. In 1998 he was appointed professor emeritus. His last larger work was a new textbook of logic (1998), even more comprehensive than the three-volume one.

Ruzsa's advanced age and failing health gradually decreased his involvement in logic and the department. But his former students who visited him over the last years had the chance to witness his spirit remaining the same throughout.

6 PHILOSOPHY OF MATHEMATICS

After this biographical outline, let me speak in some detail about Ruzsa's work in two closely connected areas that received less attention in the conference program: philosophy of mathematics and metalogic. Through the sixties, his writings on the philosophy of mathematics emerged not so much from his research interests, but mostly as responses to the interests of his readers. As he writes in the foreword of *Between Mathematics and Philosophy*, he observed that many of his students couldn't buy the lecture notes (1964) to his mathematics lectures for philosophy students (in which he discussed foundations and philosophy of mathematics in detail) because the copies were bought off by interested outsiders. He had written a book presupposing some mathematical knowledge, mainly for teachers of mathematics, but it was, again, not enough. So he published *Between Mathematics and Philosophy* (1968) for the larger public, setting forth in a popularizing style the mathematical background needed. But he didn't regard this area his field of research; his goal was merely to summarize the basics of various trends in the philosophy of mathematics from his own perspective, and convey them to the Hungarian public. So we can't speak about his philosophy of mathematics in the proper sense, I will therefore content myself with characterizing Ruzsa's point of view.

Ruzsa focuses on introducing the three classical schools in the philosophy of mathematics: logicism, intuitionism and the Hilbert-school or formalism. His main stress is on the contents of and mathematical motivations behind these trends, their connection with research in the foundations of mathematics, but he also sets forth some critical remarks. His general opinion is that all the three schools capture something from the real nature of mathematics, but each of them is one-sided; that is, he argues for some sort of eclecticism. He has the most sympathy for Hilbert's program which he demarcates from the formalist philosophy of mathematics. He agrees with this program in that foundational problems should be solved by mathematical tools, while staying away from destroying what was constructed in mathematics. He argues that Gödel's second incompleteness theorem has serious consequences for Hilbert's program but he does not consider them fatal. But he does criticize formalists for rejecting the importance of content in mathematics. Theorems of mathematics have their content, they are true propositions and it happens only for the sake of metamathematical investigations that we abstract from their content and regard them just as syntactical strings. This moderate, realistic understanding of Hilbert's

program and the sympathy for it is characteristic of the philosophical writings of Péter and Kalmár, too. But logicism is evaluated by Ruzsa in a more favorable way than by his predecessors, as he lays more stress on the philosophical, realist side of Frege's and Russell's logicism. On the other side, he criticizes intuitionists rather sharply.

Why were these writings of Ruzsa that contain little by way of original insights so popular during the sixties? Today, readers may be astonished by the occasional Marxist detours and quotations from Engels or Lenin in these works. The official prescriptions of that era were such that it was allowed to expose non-Marxist philosophical views but only when the exposition was accompanied by a thorough Marxist criticism of them. Beyond the fact that Ruzsa surveyed an area that was virtually unknown in Hungary, the novelty of his writings was that he devoted far more space to the exposition and objective analysis of the various philosophies of mathematics than to their criticism. Actually, this was a similar approach as the one found in the works of the circle of George Lukács. For example, in the book *Trends in contemporary bourgeois philosophy* by György Márkus and Zádor Tordai from 1964, we find similar efforts: the authors present the different philosophical schools and thinkers from a Marxist perspective, but the primary stress is on the exposition and analysis of the views. There was a rather sizeable distance between this attitude and the practice of Soviet Marxism, whose main concern was to classify non-Marxist thinkers as mechanical materialists, objective and subjective idealists and to discover traces of Marxist truth in their writings. It must be remarked that the Marxist detours were sincere—both from the side of Ruzsa and from the Lukácsists. They all had some rather abstract commitment to Marxism and socialism—not the actual positions of party ideologists, of course. They tried to preserve as much from Marxism as they found acceptable—there was, of course, no place for criticizing Marxism where it was not acceptable.

7 METALOGIC AND THE PHILOSOPHY OF LOGIC

After 1970, Ruzsa stopped publishing on the philosophy of mathematics. However, in his main work *Logical Syntax and Semantics* (1988, 1989) he made an important contribution to the circularity problem in foundations, that is, to the problem that logic has its semantical foundations in set theory but on the other hand, set theory is a theory which can prove its theorems within a logical framework. Ruzsa had always taught that a logical theory is useless if it has no intuitively acceptable semantical foundations. He criticized relevant logics of Zinoviev and systems of entailment for lacking such foundations and regarded Kripke semantics not only as a technical tool but as a way to make explicit the real content of modal logic. In his textbooks and lecture notes, the language

of logic is introduced in a purely semantical way. Inference rules are just mentioned but hardly anything further than that; on the elementary level, there are no formal deductions at all. Students should learn how to check the validity of a given inference by the methods of truth-tables, Venn-diagrams or semantic tableauxs; they are not expected to find out consequences of a given set of premises. The methods are semantical and their correctness is likewise confirmed by informal semantical considerations. (There is, of course, no formal semantics at this level.) This is in accord with Ruzsa's rather strong realist commitment that is present in his writings about the philosophy of mathematics and in the paragraphs and chapters concerning the philosophy of logic in his logical writings. The method to begin logic with semantically defined logical constants is present in *Logical Syntax and Semantics*, too; but quite surprising, that actually shows why Ruzsa was not a Platonist, in spite of all of his realist commitments.

The logical theory constructed there starts with introducing symbols of first-order logic—logical constants and variables—into the language of communication (metalanguage). The single difference between their introduction and the usual way is that every variable is declared, that is, it is specified what values they are allowed to take. In this way, the extended metalanguage preserves the property presupposed about the language of communication, namely, that every proposition has one and only one truth-value. Only this much is needed by way of informal semantical considerations behind the metalanguage logic. In order to prove that the axioms of this theory are true, the metalanguage is extended with class abstractions that are constructed from monadic open sentences and it is enough to introduce some minimal class theory which needs no axioms but just definitions of the empty set, the subset relation and the usual binary operations.

The concepts and assumptions needed for the theory of canonical calculi concern language as the class of expressions, that is, finite strings over a finite but nonempty alphabet (the class of letters). Using the operation of concatenation, the class of expressions can be described in an axiomatic way. Canonical calculi define inductive classes within the class of all expressions as strings deducible by a given finite set of rules. (Rules contain mostly a distinguished letter not contained in the alphabet: the arrow, and we are also allowed to use other auxiliary letters.) This very simple machinery suffices for the following:

- to represent calculi by strings of the original alphabet;
- to produce hypercalculi that define the class of all calculi;
- to introduce Gödel numbering, using as "numbers" the strings formed solely from an arbitrary element of the alphabet;
- to prove that there are certain subclasses of the language that can be defined in the metalanguage but are not inductive (although their complements with respect to the set of all expressions are).

This last claim is in fact a Gödel-type theorem now. The following step is the introduction of Markov-algorithms that is natural and easy in this language because the formalisms of canonical calculi and Markov-algorithms are very similar. Roughly speaking, the single difference between the two is that in executing an algorithm, the next step is always determined; in executing a deduction within a canonical calculus, we are free to choose the next step among the allowed ones. Enumerability and decidability by algorithms are defined as usual, and it is easy to show that enumerable sets of expressions are the same as the inductively definable ones. A set is decidable iff both the set itself and its complement is enumerable. With respect to these facts, the theorem mentioned above has as a simple corollary a Church-type theorem: there are enumerable but undecidable sets of expressions.

Real first-order logic follows only after this theory of canonical calculi and algorithms. We can inductively define the language of first-order logic and the set of provable formulas. Within this first-order logic, the theory of canonical calculi (CC) can be formalized and we can prove via metalanguage argumentation that all the theorems of CC are true. In fact, this is the only statement for which we need to use metalanguage logic and set theory. In other words, metalanguage logic has to be accepted on the basis of intuitive semantical background considerations only as far as it is applied to classes of expressions, that is, strictly finite objects. The only place for infinity is that we need a weak form of the induction principle in our metalanguage argumentation. Metalanguage set theory is basically no more than an inventory of abbreviations; its theorems are in fact truths of metalanguage logic.

Everything else turns out surprisingly simple: it follows from the previous theorems that CC is not decidable and every theory which is an inductive class of theorems containing CC is negation-incomplete. Real set theory is a first-order theory defined inductively, and we can use set theoretical propositions in constructing semantics for first-order logic only if we can prove them within this first-order set theory. This whole construction is the answer to the question of how the priority of semantics should be understood: we should accept some semantical considerations before we can construct the syntax of our logic, but these considerations are reduced to a minimum that fulfils the Hilbertian requirement of finiteness. In the formal construction, the priority belongs to syntax and deducibility; there is no Platonic heaven of mathematical objects that we know about without knowing an axiomatic theory of them. Most of the details of Ruzsa's construction of the foundations of logic are not his own inventions; but the construction as a whole is both well-considered and well-founded on the philosophical side and elegant on the mathematical side.[4]

[4] This paper was supported by the Hungarian Nationak Scientific Research Foundation OTKA, project No. 68043.

REFERENCES

All the items in Hungarian if not indicated otherwise.

Ruzsa, Imre, 1964, *Mathematics for philosophy students*. Lecture notes, two volumes. Budapest, Tankönyvkiadó.

– , 1967, *On some philosophical problems of mathematics + Mathematical logic* (the latter with János Urbán). Budapest, Tankönyvkiadó.

–, 1968, *Between mathematics and philosophy*. Budapest, Gondolat. (German: *Die Begriffswelt der Mathematik*. 1976, Berlin, Volk und Wissen.)

–, 1969, *Elementary logic*. Lecture notes. Budapest, Tankönyvkiadó.

–, 1973, *Symbolic logic*. Lecture notes, three volumes (with co-authors). Budapest, Tankönyvkiadó.

–, 1973b, Prior-type modal logic (in English). *Periodica Mathematica Hungarica*. First part: 51-69., second part: 183-201.

–, 1979, A hundred years of symbolic logic: the ouvre of Gottlob Frege. *Magyar Filozófiai Szemle*. 590-613.

–, 1980, Intensional logic without intensional variables. In I. Ruzsa (ed.), *Modal and intensional logic*. Budapest, OM Marxizmus-Leninizmus Oktatási Főosztálya.

–, 1981, *Modal logic with descriptions* (in English). The Hague, M. Nijhoff.

–, 1984, *Classical, modal and intensional logic*. Budapest, Akadémiai Kiadó.

–, 1988, *Logical syntax and semantics I*. Budapest, Akadémiai Kiadó.

–, 1989, *Logical syntax and semantics II*. Budapest, Akadémiai Kiadó.

–, 1991, *Intensional logic revisited* (in English). Budapest, published by the author.

–, 1997, *Introduction to metalogic* (in English). Budapest, Áron Publishers.

–, 1998, *Introduction to modern logic* (with András Máté). Budapest, Osiris.

FERENC CSABA

Whose Logic is Three-Valued Logic?

Abstract. 'It would be unfair to judge that I use a three-valued logic or that I abandon the principle of *tertium non datur*", writes Imre Ruzsa in his (Ruzsa 1991, 11.). For Ruzsa a truth value gap (the "third" truth value) arises only from the "defects" of our expressions (for example when a definite description does not denote anything) and not because there are "gaps" in reality. In the first part of the paper we explain in some detail how the truth value gaps arise and how they are transmitted in Ruzsa's system. In the second part we will argue that there may be sentences which in a sense reflects real gaps, in other words, that the third truth value is a real truth value.

1 THE SEMANTICS OF SEMANTIC VALUE GAPS

One of Imre Ruzsa's main achievements in logic is his system of intensional logic with semantic value gaps. A semantic value gap arises when a well formed expression of our (natural or artificial) language fails to denote anything. The simplest case is perhaps a definite description without a denotation (e.g. 'the present king of France'). In Ruzsa's system there *are* denotations even in such cases—these are the artificial entities "filling" the gaps. The individual denoted by 'the present king of France' is not a real individual: Ruzsa's choice is the set U, the set of "real" (actual and possible) individuals, simply because evidently $U \notin U$. A bit more precisely: the type ι of individuals has the domain $D(\iota) = U \cup \{U\}$, and $\Theta(\iota) = U$ is the type's *zero entity*—the "object" denoted by e.g. the empty descriptions.

Systems of intensional logics have in general two kinds of semantic values: the extensions—in Ruzsa's terminology, factual values—and the intensions. In the type o of sentences the factual values are the truth values, the intensions are functions from worlds and times (technically, from the set $I = W \times T$, where W is the set of possible worlds and T is the linearly ordered set of time moments) to

the factual values. The truth values are represented by $3 = \{0, 1, 2\}$, $\Theta(o) = 2$ being the zero entity representing the truth value gap. If a sentence p has 2 as factual value (in a world w in a given moment t) than we say it has no "real" truth value (in the world w in the moment t). The "real" truth values are of course 0 (representing falsity) and 1 (representing truth).

From the basic types ι and o, we get the other (functor) types. For example, predicates are expressions of type $o(\iota)$. The domain of this type is the set of all functions $f : D(o) \longrightarrow D(\iota)$ for which $f(U) = 2$. The zero entity of this type is the constant function having 2 for all the arguments. A predicate is partial if its interpretation $\sigma(P) : D(\iota) \longrightarrow 3$ maps more than one individual to 2.

In what follows $|A|_{vi} =$ denotes the factual value of A at the *index* $i = \langle w, t\rangle \in W \times T$ according to the valuation v. If i is an index, $d(i) \subseteq D(\iota)$ is the set of the "actual" individuals at i, that is, the actual individuals in the world w at the time moment t. If x is a variable of type ι, than the value $v(x)$ is always an element of $D(\iota)$; if $v(x) \notin d(i)$, then $|x|_{vi} = \Theta(\iota)$ and similarly for constants of type ι. It could happen that a value of variable in the world w at the moment t is an individual not belonging to the domain of w at t—in such cases the factual value of the variable is the zero entity of the type.

Definite descriptions are handled as it is expected. If F is of type $o(\iota)$ then the factual value of $\mathbb{I}F$ ('the F') is $|\mathbb{I}F|_{vi} = u_i$ if $\{u \in d(i) : |F|_{vi}(u) = 1\} = \{u_i\}$, and in all other cases $|\mathbb{I}F|_{vi} = \Theta(\iota)$. If there is exactly one F in the world w at the moment t, then $|\mathbb{I}F|_{vi}$ is *this* object; and if the set of the Fs is empty or has more than one element, $|\mathbb{I}F|_{vi}$ is the zero entity.

The identities are (of course) expressions of type o. If A and B are of the same type α, then

$$|A = B|_{vi} = \begin{cases} 2 & \text{if } |A|_{vi} = \Theta(\alpha) \text{ or } |B|_{vi} = \Theta(\alpha) \\ 1 & \text{if } |A|_{vi} = |B|_{vi} \neq \Theta(\alpha) \\ 0 & \text{otherwise} \end{cases}$$

According to this rule, if on one side of an identity stands an expression having the zero entity of its type as its factual value then the identity's factual value will be automatically 2. It has the (somewhat strange) consequence that non-existent individuals cannot be identical even with themselves. For example the sentence 'the present king of France' = 'the present king of France' falls in the truth value gap.

Ruzsa—following an idea of Tarski—defines the propositional connectives in terms of λ and $=$. The technical details (see (Ruzsa 1991, 39—41)) do not concern us, only the truth tables governing the connectives. The truth tables for negation, conjunction, and alternation are the following:

p	$\sim p$
1	0
0	1
2	2

$p \wedge q$	1	0	2
1	1	0	2
0	0	0	2
2	2	2	2

$p \vee q$	1	0	2
1	1	1	2
0	1	0	2
2	2	2	2

The tables for the conditional and the biconditional (the latter is simply the $=$ in the type o):

$p \supset q$	1	0	2
1	1	0	2
0	1	1	2
2	2	2	2

$p \equiv q$	1	0	2
1	1	0	2
0	0	1	2
2	2	2	2

These tables are the weak Kleene tables. The connectives working according to them always transmit the truth value gap from the part to the whole. This is not true for the strong Kleene connectives for which the truth tables are the following:

p	$\neg p$
1	0
0	1
2	2

$p \,\&\, q$	1	0	2
1	1	0	2
0	0	0	0
2	2	0	2

$p \vee q$	1	0	2
1	1	1	1
0	1	0	2
2	1	2	2

$p \supset q$	1	0	2
1	1	0	2
0	1	1	1
2	1	2	2

$p \equiv q$	1	0	2
1	1	0	2
0	0	1	2
2	2	2	2

The strong Kleene conjunction, alternation, and conditional does not transmit the truth value gap: for example, if p is true then the value of $p \vee q$ is 1 (true) even if the truth value of q is 2. The strong Kleene connectives are in better harmony with "the logic of empirical investigations", the conception that Ruzsa calls *epistemic*. In such a logic 2 denotes the value "yet unknown".

Why did Ruzsa decide in favor of the weak versions? The question has (at least) three answers. One is a bit personal: the epistemic conception reminds him of the "so called" intuitionist logic; and this logic, according to his conception, does not even deserve the name 'logic'.[1] The second, and less personal, answer is that using Ruzsa's system's temporal operators and introducing an epistemic operator, one could probably succeed in modeling some aspects of the "epistemic conception".

[1]He once told us the following story. In a conference (probably in the sixties) when he used the arrow for the conditional, the chair asked him: "So you are an intuitionist, aren't you?' At the very moment he decided to use the horseshoe symbol: let there be no mistake.

But Ruzsa's most prominent reason is the importance of the transmittal of the semantic value gaps, a phenomenon we have already seen in the definitions of the domain of functor types and that of the factual values of the identites. A general theorem of his system states that this phenomenon holds in general.[2]

Summing up: in extensional contexts, a semantic value gap is a special "illness" for which the treatment is: not allowing it to disappear. A truth value gap is really a *gap*, arising from clashes of language and reality; it is impossible for the "real world" to have gaps.

> My *credo* is simply this: A sentence may or may not have a truth value. If it has one then it expresses a statement which is either true or false. The lack of a truth value is not a third truth value. (Ruzsa 1991, 11.)

2 ABSOLUTELY UNDECIDABLE SENTENCES

In the epistemic conception the semantic value gaps are due to the gaps in our logic. By contrast, Ruzsa's approach is ontological. As he puts it:

> In (. . .) informal reasonings, the sources of value gaps are located in the realm of facts, in the formal semantics they [are] located in the interpretations of the (formal) language. (Ruzsa 1991, 12.)

The facts of which Ruzsa speaks are facts of the world *and* facts of our language, gaps only arise when something is mistaken in our expressions. Are there sources of truth value gaps "in the world" in which our language plays no significant role? In other words: are there gaps in reality? Before trying to answer this question, let's go back once more to the epistemic conception. According to our knowledge of it, every (well-formed, unambiguous) sentence p must fall in one of the following seven cases:

(1) p is true, and we know (proved, verified) that it is true
(2) p is true; we do not know that yet, but we will
(3) p is true, but we will never know that it is true
(4) p is absolutely undecidable (even God cannot determine its truth value)
(5) p is false, but we will never know that it is false
(6) p is false; we do not know that yet, but we will
(7) p is false, and we know that it is false

In cases (1), (2), (6) and (7) there are no difficulties. Moreover, we have good candidates that are of case (3) or of case (5). For example, let p be the sentence

> The value of the digit in the $10^{10^{10^{10}}}$th place of the decimal expansion of $\pi - 3$ equal to zero. [3]

[2]If A (of type α) is an extensional component of B (of type β), then $|A|_{vi} = \Theta(\alpha) \Rightarrow |B|_{vi} = \Theta(\beta)$, for the details see (Ruzsa 1991, 33.)

[3]Cf. (Feferman 2006).

The truth value of this sentence can be determined in principle by a mechanical check—but this check is far beyond our computational powers. Nevertheless, we can say that this sentence has a determinate truth value and God knows what it is.

What about case (4)? Does the existence of absolutely undecidable sentences threaten God's omniscience? We can say with Michael Dummett: not at all. God knows the answer to every question that has answer, and He knows of every question whether it has an answer. If there really are questions which has no answer even for Him then the divine logic must be three-valued, and instead of a truth value gap, there will be a genuine third truth value.[4]

In comparison with God, in this respect (too) we are in a more uncomfortable position: we cannot in principle distinguish cases (4), (5), and (6). The reason lies in what can be called the "Hauptsatz" of undecidable propositions: if p is an absolutely undecidable statement, then we cannot prove "constructively" that it is really the case. The proof (due to Martin-Löf[5]) relies heavily on the constructivist (verificationist) conception of negation: proving $\neg p$ amounts to showing that any attempt to prove p will eventually be blocked (in mathematics, by a contradiction; in general, by some serious difficulty). Proving that p is undecidable amounts to a proof that any attempt to prove p, as well as any attempt to prove $\neg p$, will eventually be blocked. But it is nothing but a proof of $\neg p$, and $\neg\neg p$, respectively. We arrive at a contradiction (a serious difficulty).

So we have to rely on our intuitions.

3 FINITE KNOWLEDGE OF THE INFINITE

We cannot prove of a sentence that it is absolutely undecidable, but we can perhaps imagine what such a sentence could be. Feferman's (in fact, the intuitionists') example about the decimal expansion of π gives a clue. If we—with our limited means—cannot end a process that is too long, it is conceivable that an infinite process is such that even an infinite mind cannot go through it.

But we must be careful. Even we, finite beings know very much about the natural numbers, we can prove for example that for every natural umber n, 7^n is divisible with 6, that there are infinitely many primes and so on. We have methods which make the infinite finite, that is, methods (first of all, mathematical induction) by which we can prove, say, that for some property F,

[4]See e.g. (Dummett 2006, 108—109).

[5](Martin-Löf 1995). Martin-Löf actually "proves" that there are no undecidable propositions. His proof relies on the intuitionist conceptions of proposition, truth, falsity, and knowledge. For someone not in the intuitionist camp, his conclusion can be formulated as follows: there may be absolutely undecidable propositions, but we cannot produce one about which we can prove that it is really absolutely undecidable. Cf. (Feferman 2006, 147.).

there are infinitely many numbers for which it holds. Augustine and many of his followers believe that for God it is so with *every* property F.

> As for their other assertion, that God's knowledge cannot comprehend things infinite, it only remains for them to affirm, in order that they may sound the depths of their impiety, that God does not know all numbers. For it is very certain that they are infinite. . . Does God, therefore, not know numbers on account of this infinity; and does His knowledge extend only to a certain height in numbers, while of the rest He is ignorant? Who is so left to himself as to say so?
>
> . . . if everything which is comprehended is defined or made finite by the comprehension of him who knows it, then all infinity is in some ineffable way made finite to God, for it is comprehensible by His knowledge.(Augustine 1993, Book XII., Chapter 18.))

According to this conception, the property 'n is one member of a twin prime-pair' is for God as simple as for us the property 'n is a prime number': He can decide whether it holds for infinitely many numbers or not. He can perhaps "see" some higher-order structure which decide the matter (as in the case of Fermat's last theorem there are structures revealed by the theory of analytic functions that decide that a property holds for all natural numbers bigger than 2). But is it really the case? Is every property of natural numbers such that it is in principle possible "making it finite"—deciding in finite steps, whether it holds for infinitely many numbers or not? If there is a property F for which even God has no other choice in order to decide whether it holds for infinitely many numbers than to check "all" the numbers one after another, then the sentence 'there are infinitely many numbers n for which $F(n)$ holds' is a good candidate for being absolutely undecidable.

We can argue that—*pace* Augustine—in this case even God cannot determine the truth value of this sentence (but He would know *that*). Such a sentence would then be per definitionem absolutely undecidable, and as such, it would signal the presence of a real gap "in the world".

4 INFINITE TASKS

What makes it impossible even for God to run through an infinite series of computations? The strongest argument can be extracted from the paradoxes of super-tasks. The classic example of these paradoxes is Thomson's lamp.

> There are certain reading-lamps that have a button in the base. If the lamp is off and you press the button the lamp goes on, and if the lamp is on and you press the button the lamp goes off.
>
> Suppose now that the lamp is off, and I succeed in pressing the button an infinite number of times, perhaps making one jab in one minute, another

> jab in the next half-minute, and so on. . . After I have completed the whole infinite sequence of jabs, i.e., at the end of two minutes, is the lamp on or off? It seems impossible to answer this question. It cannot be on, because I did not ever turn it on without at once turning it off. It cannot be off, because I did in the first place turn it on, and thereafter I never turned it off without at once turning it on. But the lamp must be either on or off. This is a contradiction.[6]

One lesson from the paradox is simply this: carrying out a super-task is *conceptually* impossible. For "at the end" of the infinite series of tasks, there could be a discontinuity which cannot be explained. Russell famously called it only "medically impossible" running through the whole expansion of π, see (Russell 1953, 143.). By contrast, Michael Dummett argues that "the reason why we cannot survey an infinite totality is not the deficiency of human capabilities: it is that it is *senseless* to imagine an infinite task completed" (Dummett 2006, 70–71; the italic is Dummett's).

Arithmetic can be a natural realm of super-tasks. If there is a property F of natural numbers for which the truth value of $F(n)$ for each number n can be determined by finite computation but there are "absolutely" no general method for determining whether $F(n)$ holds for infinitely many numbers or not then even a "Divine Arithmetician" who can carry out a computation "infinitely quickly" cannot determine the truth value of the statement *there are infinitely many n for which F holds* or, of the statement *there are infinitely many n for which F does not hold.* For to decide these statements, She has to check every number n, that is, by running through an infinity of tasks. And this is impossible—even for the Divine Arithmetician. The reason is that the same kind "discontinuity" would arise as in the case of the Thomson's lamp. For suppose the "prover" has a white paper. After checking $F(0)$, she paint it black; then after checking $F(1)$ she paint it again white; and so on. (If she can decide whether $F(n)$ is true or false then manipulating the paper is only a simple extra.) What color will be the paper after checking all of the natural numbers? There is no answer—the super-task cannot be carried out.[7]

[6]Thomson (1954), cited in (Sainsbury 2009, 12.). The paradox resembles that of the staccato run, a variant of Zeno's Racetrack paradox. In the staccato version the runner - say, Achilles - runs for half a minute, then pauses for half a minute, then runs for a quarter of a minute, then pauses for a quarter of a minute, and so on ad infinitum. At the end of two minutes he will have stopped and started in this way infinitely many times. Each time he pauses he could perform a task of some kind. Then at the end of two minutes he will have performed infinitely many of these tasks. On the staccato run and other paradoxes of the infinite, see (Moore 1990).

[7]It is a (super-)task for the philosophers of time to explain exactly what makes it impossible for (even) the Divine Arithmetician to run through an infinite series of tasks. For if the continuum of time has the structure—say—of the real numbers than for "someone" who could count with no speed limit, it may be possible to determine whether $F(0)$ holds or not in a minute, $F(1)$ in half a minute, $F(2)$ in a quarter of a minute and so on. . .

If p is such a sentence then it may happen that God does not know whether it is true or false. (But even in this case He knows *that*.) And in this case we can call 2 a *real* truth value.

5 WHAT WOULD RUZSA SAY

Without any doubt this argument would not affect Ruzsa's philosophical position. If there was a knock-down argument against his—Platonist—view from the standpoint of the constructivists and the intuitionists, it would not be a philosophical argument like the preceding one. And I can imagine Imre Ruzsa stamping his foot saying: "after all, there are infinitely many n for which $F(n)$ is true or there are only finitely many such n, there is no third possibility".*

REFERENCES

Augustine (Aurelius Augustinus), 1993, *City of God*. Transl. Marcus Dods. New York, The Modern Library.

Dummett, M., 2006, *Thought and Reality*. Oxford, Clarendon Press.

Feferman, S., 2006, Are there absolutely unsolvable problems? *Philosophia Mathematica* 14 (III), 134–152.

Martin-Löf, P., 1995, Verificationism then and now. In DePauli-Schimanovich, W. *&al.* (eds.), *The Foundational Debate*. Dordrecht, Kluwer, 187–196.

Moore, A. W., 1990, *The Infinite*. Oxford, Routledge.

Russell, B., 1953, The Limits of Empiricisim. *Proceedings of the Aristotelian Society* 36.

Ruzsa, I., 1991, *Intensional Logic Revisited*. Budapest, published by the author.

Sainsbury, R. M., 2009, *Paradoxes*. 3rd ed. Cambridge, Cambridge University Press.

Thomson, J. F., 1954, Tasks and super-tasks. *Analysis* 15, 1–13.

*The author happily acknowledges the support of the Philosophy of Language Research Group of the Hungarian Academy of Sciences.

Tamás Mihálydeák

On Models of General Type-Theoretical Languages

Abstract. In the present paper we consider general type theoretical languages as the representations of the functor–argument decomposition and compositional semantics relying on it and find some theorems making explicit the theoretical presuppositions of general type theoretical languages and their total or partial semantics. After defining the notion of semantic categories in the spirit of Husserl, we characterize Tarskian and Husserlian models both in total and partial semantics and prove their characteristic theorems.

1 PERSONAL FOREWORD

I am greatly indebted to Professor Imre Ruzsa for the opportunity to work with him for almost two decades. After graduation I began to work as a research assistant at Kossuth University, Debrecen in 1979 and I wrote a letter to professor Imre Ruzsa. In spite of the fact that we had never met and did not know each other personally, he answered soon. The first personal meeting changed my scientific life profoundly. I have no opportunity to tell the whole story, but I should like to emphasize that I should be quoting his books and papers[1] in almost each sentence of the present paper, which is dedicated to the memory of Professor Imre Ruzsa.

2 INTRODUCTION

From the theoretical point of view, type theoretical languages (with a lambda operator) represent function abstraction and function application and rely on functor–argument decomposition, which goes back to Frege.

[1]I mention here only three of them: (Ruzsa 1989), (Ruzsa 1991), (Ruzsa 1997).

In Frege's view, one of the most important inventions of Begriffsschrift is the replacement of the subject–predicate decomposition by the functor–argument one. He wrote the following: "The very invention of this Begriffsschrift, it seems to me, has advanced logic. . . [L]ogic hitherto has always followed ordinary language and grammar too closely. In particular, I believe that the replacement of the concept *subject* and *predicate* by *argument* and *function* will prove itself in the long run. It is easy to see how taking a content as a function of an argument gives rise to concept formation. . . . The distinction between subject and predicate finds no place in my representation of a judgement."[2] (Frege 1879/1997, 51, 53.)

One of the most general theoretical representations of the functor–argument decomposition is the well–known type theory (or the different systems of type–theoretical language and/or logic[3]).

Generally, syntactic categories have to be distinguished from semantic ones. At the same time, our formal systems fulfill the following fundamental principle of formal type–theoretical semantics:

> [The mirror principle:] Associated with every syntactic category C is a counterpart semantic category C^*, whose *mathematical type* 'mirrors' the *grammatical type* of C. And, in particular, every expression of syntactic category C is interpreted by an object of semantic category C^*.
> (Dunn and Hardegree 2001, 142.)

On the basis of the mirror principle, in what follows, we are speaking about types, and using them to define and denote different syntactic categories and the corresponding sets of possible semantic values.

3 GENERAL FORMAL SYSTEM

At first, the system of types has to be defined. The system of types relies on primitive type(s). Generally we have only one requirement: the symbol o must be a primitive type. From the theoretical point of view, the main reason for this is that the symbol o is taken as the type of the most fundamental expressions of our formal language. Expressions of type o are called formulae. Formulae directly correspond to a special sort of conceptual content or information. It means that formulae are the structures of complete information or closed (and whole) conceptual content. In a given interpretation, formulae are intended to represent complete information called proposition in the literature.

There is another, mainly semantic reason for type o having been declared to be primitive. From the semantic point of view, Frege's context principle or as

[2]I use the expression 'functor' instead of 'function' in order to differentiate an incomplete expression of a language from its semantic value.

[3]It goes back to (Church 1940).

(Hodges 2001a) says, Frege's Dictum can be taken as a general leading idea. In *The Foundations of Arithmetic* Frege wrote the following, usually quoted as the context principle:

> never to ask for the meaning of a word in isolation, but only in the context of a proposition; (Frege 1884/1980, x.)
> It is enough if the proposition taken as a whole has sense; it is this that confers on its parts also their content. (Frege 1884/1980, 71.)

According to the context principle, an expression has sense (meaning) only in the sentence in which it occurs. Sometimes we need more than one primitive type (usually individual names constitute another primitive type). The main difference between primitive and non-primitive types is that the semantic domains of primitive types have to be given via definition, while the domains of non-primitive types are originated from them. Non-primitive types are usually called functor types.

Definition 1. *Let PT be an arbitrary set of symbols, the set of primitive types, such that $o \in PT$. Then the set $TYPE_{PT}$ is defined inductively as follows:*

(1) $PT \subseteq TYPE_{PT}$;
(2) $\alpha, \beta \in TYPE_{PT} \Rightarrow \langle \alpha, \beta \rangle \in TYPE_{PT}$.

Remark 1. *Here o is the type of formulae from the syntactic point of view, and the type of their possible semantic values from the semantic point of view. $\langle \alpha, \beta \rangle$ is the type of functors which, when they are filled in by an argument of type α, yield an expression of type β in syntax (in the formal language), and it stands for the type of function from objects of type α to objects of type β in semantics.*

The type–theoretical language is the most general one concerning the functor–argument decomposition. It has only two syntactic operations: filling a functor with an argument (function application from the semantic point of view) and lambda abstraction. The latter produces a way to create a functor from an expression.

Definition 2. *A type–theoretical language is an ordered quadruple*

$$L = \langle LC, Var, Con, Cat \rangle$$

satisfying the following conditions:

(1) *LC is the set of theoretical constants.*[4] $LC = \{\lambda, (,)\}$
(2) $Var = \cup_{\alpha \in TYPE_{PT}} Var(\alpha)$ *and $Var(\alpha)$ is a denumerably infinite set of symbols*[5].

[4] A theoretical constant has the same semantic value (or sense) in every interpretation as a logical constant does in a logical system.

[5] $Var(\alpha)$ is the set of variables of the type α.

(3) $Con = \cup_{\alpha \in TYPE_{PT}} Con(\alpha)$, *where* $Con(\alpha)$ *is a denumerably set of symbols.*[6]
(4) *All mentioned sets of symbols are assumed to be pairwise disjoint ones.*
(5) $Cat = \cup_{\alpha \in TYPE_{PT}} Cat(\alpha)$, *where the sets* $Cat(\alpha)$ *are defined by the inductive rules (a). . . (c) as follows:*[7]
  (a) $Var(\alpha) \cup Con(\alpha) \subseteq Cat(\alpha)$;
  (b) $C \in Cat(\langle \alpha, \beta \rangle), B \in Cat(\alpha) \Rightarrow$ '$C(B)$' $\in Cat(\beta)$;
  (c) $A \in Cat(\beta), \tau \in Var(\alpha) \Rightarrow$ '$(\lambda \tau A)$' $\in Cat(\langle \alpha, \beta \rangle)$;

The (total or partial) functor–argument frame is the compositional mirror of a type–theoretical language. It can be said that the functor–argument frame gives *possible* semantic values.

Definition 3. *A total functor–argument frame* F *is the system of sets* $\langle Dom_F(\gamma) \rangle_{\gamma \in TYPE_{PT}}$ *such that*

(1) *If* $\gamma \in PT$, *then* $Dom_F(\gamma)$ *is an arbitrary nonempty set.*
(2) $Dom_F(\langle \alpha, \beta \rangle) = Dom_F(\beta)^{Dom_F(\alpha)}$ *for all* $\langle \alpha, \beta \rangle \in TYPE_{PT}$

Definition 4. *A partial functor–argument frame* PF *is the system of sets* $\langle Dom_{PF}(\gamma) \rangle_{\gamma \in TYPE_{PT}}$ *such that*

(1) *if* $\gamma \in PT$, *then* $Dom_{PF}(\gamma)$ *is an arbitrary set with a distinguished member* Θ_γ, *which is called the null entity of type* γ, *such that* $Dom_{PF}(\gamma) \setminus \{\Theta_\gamma\} \neq \emptyset$;
(2) $Dom_{PF}(\langle \alpha, \beta \rangle) = Dom_{PF}(\beta)^{Dom_{PF}(\alpha)}$ *for all* $\langle \alpha, \beta \rangle \in TYPE_{PT}$ *and* $\Theta_{\langle \alpha, \beta \rangle} = g$ *where* $g \in Dom_{PF}(\langle \alpha, \beta \rangle)$ *and* $g(u) = \Theta_\beta$ *for all* $u \in Dom_{PF}(\alpha)$.

Interpretive function and assignment associate the constants and the variables of the type–theoretical language with their semantic values. In a model, which consists of a frame, an interpretive function and an assignment, semantic rules can be defined to determine the semantic values of compound expressions with respect to the given model.

Definition 5. *A (total or partial) model* M *on* G *is an ordered triple* $\langle G, \varrho, v \rangle$ *where*

(1) G *is a (total or partial) functor–argument frame;*
(2) ϱ, v *are functions with domains* Con *and* Var *respectively*[8] *such that*
  (a) *if* $a \in Con(\alpha)$, *then* $\varrho(a) \in Dom_G(\alpha)$;
  (b) *if* $\tau \in Var(\alpha)$, *then* $v(\tau) \in Dom_G(\alpha)$.

Remark 2. *A model* M *on* G *is total or partial if* G *is a total or partial functor–argument frame respectively.*

[6] Con is the set of non–theoretical symbols of L. The semantic value of an expression belonging to the set Con is given by an interpretation.

[7] Cat is the set of all well–formed expressions of L. The set $Cat(\alpha)$ is the α–category of L ($\alpha \in TYPE_{PT}$).

[8] ϱ is an interpretive function, v is an assignment.

If $M = \langle F, \varrho, v\rangle$ *is a total model on* F*, then*

$$Dom_M(\alpha) = Dom_F(\alpha).$$

If $PM = \langle PF, \varrho, v\rangle$ *is a partial model on* PF*, then*

$$Dom_{PM}(\alpha) = Dom_{PF}(\alpha) \setminus \{\Theta_\alpha\}.$$

If $M\ (= \langle G, \varrho, v\rangle)$ *is a total or partial model,* $\xi \in Var(\gamma)$ *and* $u \in Dom_G(\gamma)$*, then the model* $M^u_\xi\ (= \langle G, \varrho, v[\xi : u]\rangle)$ *is like* M *except that* $v[\xi : u](\xi) = u$*.*

Definition 6. *A total or partial model* $M\ (= \langle G, \varrho, v\rangle)$ *assigns each expression* A *of type* α *a semantic value* $[\![A]\!]_M$ *according to the following semantic rules:*

(1) *if* $a \in Con(\gamma)$*, then* $[\![a]\!]_M = \varrho(a)$*;*
(2) *if* $\xi \in Var(\gamma)$*, then* $[\![\xi]\!]_M = v(\xi)$*;*
(3) *if* $A \in Cat(\langle\alpha, \beta\rangle)$ *and* $B \in Cat(\alpha)$*, then* $[\![A(B)]\!]_M = [\![A]\!]_M([\![B]\!]_M)$*;*
(4) *if* A *is an expression of type* β *and* $\xi \in Var(\alpha)$*, then* $[\![\lambda\xi A]\!]_M = g$*, where* g *is a function from* $Dom_G(\alpha)$ *to* $Dom_G(\beta)$ *such that* $g(u) = [\![A]\!]_{M^u_\tau}$ *for all* $u \in Dom_G(\alpha)$*.*

Proposition 1. *If* M *is a total model and* $A \in Cat(\alpha)$*, then* $[\![A]\!]_M \in Dom_M(\alpha)$*. If* M *is a partial model, then* $[\![A]\!]_M \in Dom_M(\alpha) \cup \{\Theta_\alpha\}$*.*

Definition 7. *If* M *is a total or partial model, then* A *is meaningful with respect to* M*, in symbols* $A \in Cat^M_{mf}$ *if* $A \in Cat(\alpha)$ *for some type* α *and* $[\![A]\!]_M \in Dom_M(\alpha)$*.*

Remark 3. *If* M *is a total model, then all* $A \in Cat$ *are meaningful, i.e. there is no difference at all between the notions of well–formedness and meaningfulness. We can only make a real differentiation between them in the case of partial models.*

Theorem 1. *If* $A \in Cat$*,* $M_1 = \langle G, \varrho, v_1\rangle$ *and* $M_2 = \langle G, \varrho, v_2\rangle$ *are two (total or partial) models of* L *with the same frame* G *and interpretive function* ϱ *such that* $v_1(\tau) = v_2(\tau)$ *for all* $\tau \in V(A)$[9]*, then* $[\![A]\!]_{M_1} = [\![A]\!]_{M_2}$*.*

Proposition 2. *If* $A \in Cat$ *is a closed expression, then* $[\![A]\!]_M$ *is independent from* v *i.e.* $[\![A]\!]_M = [\![A]\!]_{M^u_\tau}$ *for all* $\tau \in Var(\gamma)$ *and* $u \in Dom_F(\gamma)$*.*[10]

To prove lambda–conversion law, we need the Law of replacement 2 and Lemma 1. The first one says that in semantics, we only take into consideration semantic values and don't pay any attention to the expression itself—except its type—whose semantic value is given. It doesn't matter how we get a

[9]The definitions of subterms, free variables, open and close expressions and the substitutability are usual ones. The set $V(A)$, is the set of free variables of the expression A.

[10]In the case of closed expressions we can speak about models as ordered pairs of frames and interpretive functions.

semantic value, what form of the compound expression gets the semantic value. We may formulate the property in the law of replacement by means of universal replacement of expressions belonging to the same type with the same semantic value. >From the logical–philosophical point of view, the law of replacement is a special type–theoretical formulation of a version of the principle of compositionality called the substitutivity principle, which goes back to Leibniz.

> [The Substitutivity Principle:] If two expressions have the same meaning, then substitution of one for the other in a third expression does not change the meaning of the third expression. (Szabó 2000, 490.)

I must emphasize that the law of replacement can only be considered as a restricted version of the substitutivity principle, the unrestricted form of the substitutivity principle holds only in Husserlian models dealt with in Section 6. The next definition introduces the notion of 1–compositionality. 1–compositional systems fulfill a restricted version of the substitutivity principle, and Corollary 1 of Law of replacement 2 says that our general system is compositional in the sense of 1–compositionality.

Definition 8. *Let M be a model of L. We say that M is 1–compositional if for all well–formed expressions A, B, C $(A, B, C \in Cat)$ and variable τ $(\tau \in Var)$ such that $(\lambda\tau C)(A), (\lambda\tau C)(B) \in Cat^M_{mf}$*

$$[\![A]\!]_M = [\![B]\!]_M \Rightarrow [\![(\lambda\tau C)(A)]\!]_M = [\![(\lambda\tau C)(B)]\!]_M$$

Theorem 2 (Law of replacement). [11]

If $A \in Cat$, $B, C \in Cat(\gamma)$, and M is a (total or partial) model of L, then

$$[\![B]\!]_M = [\![C]\!]_M \Rightarrow [\![A]\!]_M = [\![A[C \downarrow B]]\!]_M.$$

Corollary 1. *If M is a (total or partial) model of L, then M is 1–compositional.*

Lemma 1. *If B is substitutable for variable τ in A, M is a (total or partial) model, and $[\![B]\!]_M = u$, then $[\![A^B_\tau]\!]_M = [\![A]\!]_{M^u_\tau}$.*

Theorem 3 (Lambda–conversion law). *If $A \in Cat$, $\tau \in Var(\beta)$, $B \in Cat(\beta)$ and B is substitutable for τ in A, then $[\![(\lambda\tau A)(B)]\!]_M = [\![A^B_\tau]\!]_M$ for all (total or partial) models M.*

4 PROPERTIES OF TOTAL AND PARTIAL MODELS

Let us turn our attention to different, total or partial models.[12] We need some notions to compare and combine different models. In the following

[11] If $A \in Cat$ and $B, C \in Cat(\gamma)$, then $A[C \downarrow B]$ $(\in Cat)$ is obtained by replacing a subterm occurrence (i.e. not preceded immediately by λ) of B by C.

[12] The proofs of theorems in Section 4.5 can be found in (Mihálydeák 2010, 127–131.).

definitions let $L\ (= \langle LC, Var, Con, Cat\rangle)$ be a type–theoretical language and $M\ (= \langle G, \varrho, v\rangle)$ be its total or partial model.

Definition 9.

(1) *If $\approx$ is an equivalence relation on the set $Cat'\ (\subseteq Cat)$, then $\approx$ is a synonymy for L. The set Cat' is the field of synonymy $\approx$.*

(2) *Syntactic synonymy for L is the synonymy $\cong_L$ generated by the syntax of L, i.e. $A \cong_L B$ if and only if there is a type γ such that $A, B \in Cat(\gamma)$.*

(3) *Synonymy generated by the model M is a synonymy $\approx_M$ for L with the field Cat^M_{mf} such that $A \approx_M B \Leftrightarrow [\![A]\!]_M = [\![B]\!]_M$.*

(4) *Closed synonymy (or c–synonymy) generated by the model M is a synonymy $\approx_{Mc}$ for L with the field $\{A : A \in Cat, A \text{ is closed}\} \cap Cat^M_{mf}$ such that $A \approx_{Mc} B \Leftrightarrow [\![A]\!]_M = [\![B]\!]_M$.*

(5) *A synonymy $\approx$ for L is semantic if there is a model M of L such that $\approx_M$ equals $\approx$.*

The next proposition shows that in a general type–theoretical compositional framework, syntactic synonymy can be treated as a degenerate semantic one.

Proposition 3. *The syntactic synonymy for L is semantic (in a degenerate sense).*

Remark 4. *In what follows, a model of L generating the synonymy $\cong_L$ is denoted by M_L and called 'syntactic' model.*

Definition 10.

(1) *Two models M_1, M_2 of a language L are said to be equivalent (closed equivalent, c–equivalent) if $\approx_{M_1}$ equals $\approx_{M_2}$ ($\approx_{M_1 c}$ equals $\approx_{M_2 c}$), i.e. their generated synonymies (c–synonymies) are equivalent.*

(2) *Given two synonymies $\approx$ and $\approx'$ for L, we say that $\approx$ is compatible with $\approx'$ if for all expressions $A, B\ (\in Cat)$ in the field of both synonymies, $A \approx B \Leftrightarrow A \approx' B$*

(3) *Given two synonymies $\approx$ and $\approx'$ for L, we say that $\approx$ is closed compatible with (or c–compatible with) $\approx'$ if for all closed expressions $A, B (\in Cat)$ in the field of both synonymies $A \approx B \Leftrightarrow A \approx' B$*

(4) *We say that two models M_1, M_2 of L are compatible (closed compatible) if their generated synonymies $\approx_{M_1}, \approx_{M_2}$ are compatible (c–compatible).*

Proposition 4. *If the models M_1, M_2 of L are equivalent, then M_1 and M_2 are compatible and c–compatible.*

Proposition 5. *If the models $M_1,\ M_2$ of L are equivalent, then M_1 and M_2 are c–equivalent.*

Proposition 6. *If $M\ (= \langle G, \varrho, v\rangle)$ is a model of L, $\tau \in Var(\gamma)$ and $u \in Dom_G$, then the models M and M^u_τ are c–equivalent.*

Proposition 7. *If the models $M_1,\ M_2$ of L are compatible, then $M_1,\ M_2$ are c–compatible.*

Proposition 8. *Let the models $M_1,\ M_2$ of L be total. $M_1,\ M_2$ are compatible if and only if $M_1,\ M_2$ are equivalent.*

In order to investigate the connection between total and partial semantic systems, we need a 'total' or 'pseudo partial' part of a partial frame PF, which is denoted by PF^t.

Definition 11. *Let PF be a partial frame. The total part PF^t of the partial frame PF is the system of sets $\langle Dom^t_{PF}(\gamma)\rangle_{\gamma\in TYPE_{PT}}$ such that*

(1) *if $\gamma \in PT$, then $Dom^t_{PF}(\gamma) = Dom_{PF}(\gamma) \setminus \{\Theta_\gamma\}$;*
(2) *if $\gamma = \langle\alpha,\beta\rangle$ then $Dom^t_{PF}(\gamma) \subseteq Dom_{PF}(\gamma)$ such that for all $f \in Dom^t_{PF}(\langle\alpha,\beta\rangle)$ $f(u) \in Dom^t_{PF}(\beta)$ if $u \in Dom^t_{PF}(\alpha)$ and $f(u) = \Theta_\beta$ otherwise.*

Remark 5.

(1) *For the sake of brevity, we use the notation 'Dom^t_F' in the case of a total frame F. Of course, in this case $Dom^t_F(\gamma) = Dom_F(\gamma)$ for all $\gamma \in TYPE_{PT}$.*
(2) *If $M\ (= \langle G, \varrho, v\rangle)$ is a total or partial model, then $Dom^t_M(\gamma) = Dom^t_G(\gamma)$ for all $\gamma \in TYPE_{PT}$.*

Definition 12. *An expression A of type γ is total with respect to M if $[\![A]\!]_M \in Dom^t_M(\gamma)$.*

Proposition 9.

(1) *If a non–logical constant A of a primitive type is meaningful with respect to a model M of L, then A is total, i.e. if $A \in Con(\gamma)$ where $\gamma \in PT$, and $A \in Cat^M_{mf}$, then $[\![A]\!]_M \in Dom^t_M(\gamma)$.*
(2) *If $A \in Cat(\langle\alpha,\beta\rangle)$ and $B \in Cat(\alpha)$ are total with respect to M, then $A(B)$ is total with respect to M.*

Definition 13.

(1) *If $\approx, \approx'$ are synonymies for L, we say that $\approx'$ extends $\approx$ (or it is an extension of $\approx$) if the field of $\approx'$ includes that of $\approx$ and the two synonymies are compatible.*
(2) *If M_1, M_2 are models of L, we say that M_2 extends M_1 (or that it is an extension of M_1) if $[\![A]\!]_{M_2} = [\![A]\!]_{M_1}$ for all $A \in Cat^{M_1}_{mf}$.*
(3) *If M_1, M_2 are models of L, we write $M_2 \geq M_1$ to mean that $\approx_{M_2} \supseteq \approx_{M_1}$.*

Remark 6. *If $M_2 \geq M_1$, then the domain of M_2 includes that of M_1, but within that domain, M_1 may make more distinctions than M_2 does.*

Proposition 10. *The models M_1 and M_2 of L are equivalent if and only if both $M_2 \geq M_1$ and $M_1 \geq M_2$.*

Proposition 11. *If M_2 extends M_1, then $M_2 \geq M_1$. (In this case M_2 makes exactly the same distinctions in the field of M_1 as M_1 does.)*

Proposition 12. *A total model is maximal in the sense that all of its extensions are equivalent to it.*

Proposition 13. *A total model M of L is minimal in the sense that there is no total model M' such that M extends M' and M and M' are not equivalent.*

Corollary 2. *If a total model M extends M' such that M and M' are not equivalent, then M' is a partial model of L.*

5 TARSKIAN MODELS

In Section 4 we investigated the properties of models by means of their generated synonymies. In his well–known paper (Tarski 1936/1983) Tarski introduces a new classification. The classification and therefore the associated synonymy is—at least in some cases—between syntactic synonymy and synonymies generated by non–degenerate models of our language.

Definition 14. *If L is a type–theoretical language, M is a model of L and A, B are well–formed expressions (or grammatical terms, i.e. $A, B \in Cat$), then we say that A, B belong to the same semantic category with respect to M (they have the same M–category), in symbols $A \sim_M B$, if for every expression C $(\in Cat)$ and a variable τ $(\in Var)$*

$$(\lambda\tau C)(A) \in Cat^M_{mf} \Leftrightarrow (\lambda\tau C)(B) \in Cat^M_{mf}.$$

In a very general sense, the next proposition has been mentioned by Tarski. In our case it sounds as follows:

Proposition 14. *If M is a (total or partial) model of L, then $\sim_M$ is a synonymy with the field of Cat.*

Theorem 4. *$A \sim_M B \Rightarrow A \cong_L B$ (and so $\cong_L \supseteq \sim_M$), where M is a (total or partial) model of L.*

Corollary 3. *If A, B are well–formed but not meaningful expressions with respect to a partial model M, i.e. $A, B \in Cat \setminus Cat^M_{mf}$, then*

$$A \sim_M B \Leftrightarrow A \cong_L B$$

By means of the notion of semantic category, Tarski lays down a very important principle called the first principle of the theory of semantic categories,[13] which is, as he says, very natural "from the standpoint of ordinary usage of language" (Tarski 1936/1983, 216.). In our terminology the principle sounds informally as follows:

[13]Its original version can be found in (Tarski 1936/1983, 216.).

> [The first principle of the theory of semantic categories:] Two expressions of our language have the same semantic category if there is an expression of our language such that it produces meaningful expressions when combined with them.[14]

The following definition formulates the first principle of the theory of semantic categories formally, and gives the notion of a Tarskian model:

Definition 15. *We say that the model M of L is Tarskian if it is the case that if there is a meaningful expression C and a variable τ such that $(\lambda\tau C)(A)$ and $(\lambda\tau C)(B)$ are both meaningful, then A and B have the same M–category.*

Remark 7. *A model M of L is Tarskian if and only if it fulfills Tarski's first principle of the theory of semantic categories.*

Theorem 5 (Characteristic theorem of Tarskian models)**.** *The model M of L is Tarskian, if and only if the synonymies $\sim_M$ and $\cong_L$ are equivalent, i.e. $\sim_M$ equals $\cong_L$.*

Remark 8. *According to the Characteristic theorem of Tarskian models 5, all Tarskian models of L have the same system of semantic categories and this system is equivalent to the system of syntactic categories.*

Proposition 15. *If the model M of L is total, then the synonymies $\sim_M$ and $\cong_L$ are equivalent, i.e. $\sim_M$ equals $\cong_L$.*

Theorem 6. *If M is a total model of L, then M is Tarskian.*

Corollary 4. *Non–Tarskian models are partial.*

6 HUSSERLIAN MODELS

In Section 5 we dealt with the connection between syntactic and semantic categories. The next step we have to take is the investigation of the bridge between the system of semantic categories and the classification generated by the equivalence relation $\approx_M$.

Definition 16.

(1) *Let M_1, M_2 be models. We say that M_1 and its generated synonymy $\approx_{M_1}$ are M_2–Husserlian if $A \approx_{M_1} B \Rightarrow A \sim_{M_2} B$ for all $A, B \in Cat$.*
(2) *We say that a model M of L is Husserlian if it is M–Husserlian. (That is $A \approx_M B \Rightarrow A \sim_M B$ for all $A, B \in Cat$.)*
(3) *We say that a model M $(= \langle G, \varrho, v\rangle)$ of L is strictly Husserlian if M' $(= \langle G, \varrho, v'\rangle)$ is Husserlian for all assignments v'.*
(4) *We say that the generated synonymy $\approx_M$ of a model M is Husserlian (strictly Husserlian) if the model M is Husserlian (strictly Husserlian).*

[14] A version of the principle is quoted by (Hodges 2001b, 11.).

The notion of a Husserlian model creates a connection between generated synonymy and M–category. It requires that two expressions with the same semantic value with respect to M have to belong to the same M–category, and so according to Theorem 4 they have to have the same type. More precisely:

Proposition 16. *If a model M of L is Husserlian and $A \approx_M B$ for some $(A, B \in Cat)$, then $A \cong_L B$ i.e. there is a $\gamma \in TYPE_{PT}$ such that $A, B \in Cat(\gamma)$.*

Corollary 5. *If a model M of L is Husserlian, then $\cong_L \supseteq \approx_M$, i.e. $M_L \geq M$.*

Corollary 6. *Let M_1, M_2 be models of L. If M_1 is M_2–Husserlian, then it is M_L–Husserlian.*

Theorem 7. *Let M be a Tarskian model of L. The model M is Husserlian if and only if $\cong_L \supseteq \approx_M$, i.e. $M_L \geq M$.*

Corollary 7. *Let M be a total model of L. The model M is Husserlian, if and only if $\cong_L \supseteq \approx_M$, i.e. $M_L \geq M$.*

Law of replacement 2 says that an expression can substitute for another one without changing the semantic value of the compound expression, if the semantic value of the first expressions equals that of the second one. In the law, there is a special condition usually regarded as not too important. The condition requires that the two expressions have to belong to the same syntactic category. Without supposing it, the law of replacement holds only in Husserlian models. That is why I said that Law of replacement 2 is only a restricted version of the substitutivity principle (see in Section 3), a version of the principle of compositionality. Its unrestricted type–theoretical formulation is the following Husserlian law of replacement.

Theorem 8 (Husserlian law of replacement)**.** *If $A, B, C \in Cat$ and M is a Husserlian model of L, then*

$$[\![B]\!]_M = [\![C]\!]_M \Rightarrow [\![A]\!]_M = [\![A[C \downarrow B]]\!]_M.$$

Theorem 9 (Conversion of Husserlian law of replacement)**.** *If for all $A, B, C \in Cat$*

$$[\![B]\!]_M = [\![C]\!]_M \Rightarrow [\![A]\!]_M = [\![A[C \downarrow B]]\!]_M,$$

then M is a Husserlian model of L.

Proof. The proof is indirect. Suppose that the model M is not Husserlian. Then there are $B, C \in Cat$ such that $B \approx_M C$ ($[\![B]\!]_M = [\![C]\!]_M$) and $B \not\sim_M C$. Therefore there is some $D \in Cat, \tau \in Var$, such that $(\lambda\tau D)(B) \in Cat^M_{mf}$ and $(\lambda\tau D)(C) \notin Cat^M_{mf}$. According to Law of replacement 2, it is impossible that $B,\ C \in Cat(\gamma)$ for some $\gamma \in TYPE_{PT}$ because in contrary $[\![(\lambda\tau D)(B)]\!]_M = [\![(\lambda\tau D)(C)]\!]_M$. Therefore there are $\alpha, \beta \in TYPE_{PT}$ such that $\alpha \neq \beta$ and $B \in Cat(\alpha),\ C \in Cat(\beta)$. Let $A =$ '$(\lambda\xi\xi)(B)$' where $\xi \in Var(\alpha)$. $A \in Cat$ and $A[C \downarrow B] \notin Cat$ and so $[\![A]\!]_M \neq [\![A[C \downarrow B]]\!]_M$. □

Definition 17. *A model M of L fulfills the substitutivity principle if for all $A, B, C \in Cat$*

$$[\![B]\!]_M = [\![C]\!]_M \Rightarrow [\![A]\!]_M = [\![A[C \downarrow B]]\!]_M.$$

The next theorem shows that the substitutivity principle is a strong version of the principle of compositionality. In our theoretical framework all models are compositional, but a model fulfills the substitutivity principle if and only if it is Husserlian.

Theorem 10 (Characteristic theorem of Husserlian models)**.** *A model M of L is Husserlian if and only if it fulfills the substitutivity principle.*

Definition 18. *A model M of L is strongly compositional if it fulfills the substitutivity principle.*

Remark 9. *Characteristic theorem of Husserlian models 10 says the property of being strongly compositional is equivalent to being Husserlian.*

Corollary 8. *If M is a Tarskian model of L and $\cong_L \supseteq \approx_M$, then it fulfills the substitutivity principle.*

Theorem 11. *A model M of L is strictly Husserlian if and only if the sets $Dom_M(\gamma)$ $(\gamma \in PT)$ are pairwise disjoint ones.*

Proof. I have to note that the sets $Dom_M(\gamma)$ $(\gamma \in PT)$ are pairwise disjoint ones if and only if the sets $Dom_M(\gamma)$ $(\gamma \in TYPE_{PT})$ are pairwise disjoint ones.

At first we prove that if M $(= \langle G, \varrho, v \rangle)$ is strictly Husserlian, then the sets $Dom_M(\gamma)$ $(\gamma \in TYPE_{PT})$ are pairwise disjoint ones. The proof is indirect. Suppose that M is strictly Husserlian and there is a semantic value u such that $u \in Dom_M(\alpha) \cap Dom_M(\beta)$ where $\alpha \neq \beta$. Let $\tau_1 \in Var(\alpha)$ $\tau_2 \in Var(\beta)$ and v' be an assignment such that $v'(\tau_1) = u = v'(\tau_2)$. If $M' = \langle G, \varrho, v' \rangle$, then $[\![\tau_1]\!]_{M'} = [\![\tau_2]\!]_{M'}$ but $\tau_1 \ncong_L \tau_2$ and according to Proposition 16 M' is not Husserlian. So M is not strictly Husserlian.

Secondly it is enough to prove that if M is a model of L and the sets $Dom_M(\gamma)$ $(\gamma \in TYPE_{PT})$ are pairwise disjoint ones, then M is Husserlian. The proof is indirect. Suppose that $A \approx_M B$ and $A \nsim_M B$ where $A, B \in Cat^M_{mf}$. Then $A \ncong_L B$, and so there are $\alpha, \beta \in TYPE_{PT}$ such that $A \in Cat(\alpha)$, $B \in Cat(\beta)$ and $\alpha \neq \beta$. According to Proposition 1 $[\![A]\!]_M \in Dom_M(\alpha)$, $[\![B]\!]_M \in Dom_M(\beta)$. Since $[\![A]\!]_M = [\![B]\!]_M$, $Dom_M(\alpha) \cap Dom_M(\beta) \neq \emptyset$. □

Definition 19. *A (total or partial) frame G is strictly Husserlian if the sets $Dom_G(\gamma)$ $(\gamma \in PT)$ are pairwise disjoint ones.*

Corollary 9. *If M is a model on a strictly Husserlian frame then the model M of L is strictly Husserlian.*

Corollary 10. *The degenerate model M_L, which generates the syntactic synonymy $\cong_L$, is strictly Husserlian, and so the synonymy $\cong_L$ is strictly Husserlian.*

Theorem 12. *A model M is Husserlian if and only if there is a strictly Husserlian model M' such that $M' \geq M$.*

Proof. According to Corollary 10, the model M_L is strictly Husserlian. If M is Husserlian, then according to Corollary 5, $M_L \geq M$.

Let M' be a strictly Husserlian model such that $M' \geq M$. Then according to Corollary 5 $M_L \geq M'$ and so $M_L \geq M$. It means that if $A \approx_M B$, then $A \cong_L B$ i.e. there is $\gamma \in TYPE_{PT}$ such that $A, B \in Cat(\gamma)$. Therefore $(\lambda\tau C)(A) \in Cat$ if and only if $(\lambda\tau C)(B) \in Cat$ for any $C \in Cat$ and $\tau \in Var$. According to Law of replacement 2 $[\![(\lambda\tau C)(A)]\!]_M = [\![(\lambda\tau C)(B)]\!]_M$ and so $(\lambda\tau C)(A) \in Cat^M_{mf} \Leftrightarrow (\lambda\tau C)(B) \in Cat^M_{mf}$, i.e. $A \sim_M B$. □

REFERENCES

Church, A., 1940, A formulation of the simple theory of types. *Journal of Symbolic Logic 5*. 56–68.

Dunn, J. M. and G. M. Hardegree, 2001, *Algebraic Methods in Philosophical Logic*, Vol. 41 of *Oxford Logic Guide*. New York, Oxford University Press.

Frege, G., 1884/1980, *The Foundations of Arithmetic. A Logic-Mathematical Enquiry into the Concept of Number*. Oxford, Basil Blackwell, second revised edition. Translated by J. L. Austin, from Grundlagen der Arithmetik. Eine logisch-matematisch Untersuchung über den Begriff der Zahl. Breslau, W. Koebner.

Frege, G., 1879/1997, Begriffsschrift, a formula language of pure thought modelled on that of arithmetic. In M. Beaney (ed.), *The Frege Reader*. Oxford, Blackwell. 47–78. Selections (Preface and part I). Translated by M. Beaney from Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens. Halle, L. Nebert.

Hodges, W., 2001a, A context principle. Manuscript.

Hodges, W., 2001b, Formal features of compositionality. *Journal of Logic, Language and Information 10*. 7–28.

Mihálydeák, T., 2010, On Tarskian models of general type-theoretical languages. In C. Drossos, P.Peppas and C. Tsinakis (eds.), *Proceedings of the 7th Panhellenic Logic Symposium*. Patras, Patras University Press. 127–131.

Ruzsa, I., 1989, *Logikai szintaxis és szemantika* (in Hungarian). Budapest, Akadémiai Kiadó.

Ruzsa, I., 1991, *Intensional Logic Revisited*. Budapest, published by the author.

Ruzsa, I., 1997, *Introduction to Metalogic*. Budapest, Áron Publishers.

Szabó, Z. G., 2000, Compositionality as supervenience. *Linguistics and Philosophy 23*. 475–505.

Tarski, A., 1936/1983, The concept of truth in formalized languages. In J. Corcoran (ed.), *Logic, Semantics, Metamathematics*. Indianapolis, Hackett Publishing, second edition. 152–278.

Thomason, R. H., 1999, Type theoretic foundations of context, Part 1: Contexts as complex type-theoretic objects. In P. Bouquet, L. Serafini, P. Brézillon, M. Benerecetti, and F. Castellani (eds.), *Modeling and Using Contexts: Proceedings of the Second International and Interdisciplinary Conference, CONTEXT'99*. Berlin, Springer–Verlag. 352–374.

Thomason, R. H., 2001, Contextual intensional logic: type–theoretic and dynamic considerations. Manuscript.

Zsófia Zvolenszky

Ruzsa on Quine's Argument Against Modal Logic

Abstract. Through the 1970s and 1980s—the days when ELTE philosophy was named Marxism–Leninism—Imre Ruzsa prepared logic books and articles with sharp, comprehensive, up-to-date surveys of the most recent international developments in logic and the philosophy of language. For decades to come, the chapters of his *Classical, Modal and Intensional Logic* would be just about the only Hungarian-language sources available on W. V. O. Quine's famous argument against modal logic, on Saul Kripke's modal semantics that seemed to bypass the Quinean objections, and on Kripke's arguments about the semantics of natural language: that proper names are rigid designators. Based primarily on John Burgess's subsequent work, we can complete the picture of modal logic that Ruzsa painted in his survey by shedding light on additional important connections: crucial links not so much between Quine's argument and Kripke's formal work (as Ruzsa and others had thought), but instead between the Quinean argument and Kripke's thesis about proper names being rigid designators.

Various stripes of modality—senses of 'must' and 'can', necessity and possibility—are traditionally distinguished by logicians, linguists, and philosophers. Let us list a couple of them:

– Deontic modality—what is necessary/possible given laws or norms; that is, what the laws/norms require/permit. For example, "It is necessary (given public transportation regulations) that I buy a ticket to ride the tram"; more colloquially put: "I must buy a ticket to ride the tram".
– Epistemic modality—what is necessary/possible given what is known. For example, "It is necessary (given what I know) that the Opera building is in the next block"; more colloquially put: "The Opera building must be in the next block".

There is also the category of alethic modality, concerning *truth*—what is necessarily and possibly true. Within this, we can draw further distinctions; let us focus on necessity, leaving possibility aside (as is traditionally done):

– Necessary truth as logical truth (sometimes called 'strict modality')—truth given some system of logic, in other words, truth given the meanings of the logical vocabulary of a selected system. "I either buy a ticket or I don't buy a ticket" is an example of a logical truth, for it is true in virtue of what 'or' and 'not' mean.
– Necessary truth as analytic truth—truth given the meanings of the words contained in the sentence. "All single people are unmarried" is an example of an analytic truth.
– Necessary truth as physical or natural necessity—truth given the laws of physics/laws of nature. "Trams travel slower than the speed of light" is an example of a truth of physics.
– Indeed, in his state-of-the-art 1984 survey volume *Classical, Modal and Intensional Logic* (written in Hungarian), Imre Ruzsa distinguished each of these stripes of modality (Ruzsa 1984, 119–121., 156–160.). What is conspicuously missing from Ruzsa's (and his contemporaries') list is yet another sense of necessity within the alethic category: the notion of counterfactual or metaphysical necessity, brought into the limelight by Saul Kripke's 1970 lecture series "Naming and Necessity" (subsequently published as Kripke 1980):
– Necessary truth as counterfactual (or metaphysical) necessity—truth across all counterfactual circumstances. "Aristotle is (was) human" is a plausible example of a metaphysically necessary truth. Although it is epistemically as well as logically and analytically possible for Aristotle to be a cat, it is not *counterfactually or metaphysically* possible that he is a cat.

Ruzsa's *Classical, Modal and Intensional Logic* stood alone in various ways, providing just about the only Hungarian-language coverage of numerous landmarks in philosophy of language and logic for almost two decades:

(I) W. V. O. Quine's arguments against modal logic (1943–1962)
(II) Kripke's formal results: semantics for modal logic (1959–1963)
(III) Kripke on the semantics of natural language, specifically, his theory that proper names are so-called rigid designators. (1970)

As for (I), it was not until 2002 that a collection of Quine's essays was published in Hungarian, including his definitive formulation of his attack on modal logic "Reference and Modality" (Quine 1953, discussed in detail below). Until then, there were just three articles by Quine available in Hungarian: "Two Dogmas of Empiricism" (Quine 1951/1973) as well as two smaller chapters from Quine's attacks on modal logic (Quine 1963, 1947 both in Copi–Gould 1964/1985). Ruz-

sa's 30-page section entitled "Modality and Quantification: Logic 'Conceived in Sin'" was thus, for quite some time, *the* Hungarian source to consult on Quine's attacks on modal logic (Ruzsa 1984, 164–193).

As for (II), to this day, none of Kripke's formal work has been translated into Hungarian, and Ruzsa's 20-page section entitled "Kripke's Modal Semantics" remains the definitive secondary source to turn to in Hungarian (Ruzsa 1984. 227–248, see also Ruzsa 1988, XX). In addition, Ruzsa went on to develop his own Quine-proof system of modal logic (Ruzsa 1984, 290–345).

As for (III), not until the late 1990s was there any Hungarian coverage or translation of Kripke's *Naming and Necessity* available in Hungarian, apart from Ruzsa's 13-page section on Kripke's rigidity thesis (Ruzsa 1984. 302–315). Kripke argues that proper names like 'Aristotle' exhibit distinctive behavior within a certain rather straightforward kind of modal context: they are rigid designators, that is, they refer to the same individual with respect to every counterfactual situation. The rigidity thesis yields a powerful argument against Frege's descriptivist theory of proper names, which associates proper names with definite descriptions—such as 'the teacher of Alexander the Great'—that are non-rigid: after all, in a counterfactual situation in which someone else taught Alexander the Great, this definite description picks out someone other than Aristotle. The turn of the 20th-21st centuries brought the Hungarian translation of Kripke's "Identity and Necessity" paper, which also discusses the rigidity thesis (Kripke 1971/2004; see also the brief excerpts collection Kripke 1980/1997). Around the same time, important, albeit brief coverage of the rigidity thesis appeared in (Sainsbury 1997, 85–89) and (Farkas–Kelemen 2002, 135–145). The Hungarian translation of *Naming and Necessity*, along with an 87-page companion article was published fairly recently (Kripke 1980/2007, Zvolenszky 2007). Again, for almost two decades, Ruzsa's 1984 book provided one of very few sources on Kripke's work on the semantics of natural language.

My goal in this short paper is to highlight, beyond (I)–(III), two more aspects of the debate between Quine and Kripke, neither of which have been properly recognized by Ruzsa or his contemporaries:

Supplementing (I): (a) Quine's lasting argument against modal logic, and his challenge to locate an alternative notion of necessity unaffected by his arguments (especially in Quine 1953, 1960, 1963).

Supplementing (III): (b) The role of Kripke's explication of the notion of metaphysical necessity (1970).

These complete the picture painted by Ruzsa's pioneering survey in *Classical, Modal and Intensional Logic*.

* * *

From the 1940s through the 1960s, Quine put forth various arguments against modal logic and did not properly distinguish them, which made interpreting him no easy task. One of these arguments—(a)—stands, posing a challenge that was not met until Kripke's observations about counterfactual necessity—(b)—appeared on the scene. Yet this went unrecognized until much later—from the late 1990s, particularly by John Burgess (1997) and Stephen Neale (2000):

(a) Preliminary formulation: Quine's lasting argument: Certain formulas of modal logic lack sense, they cannot be interpreted.

Let us see how we might arrive at such a suspect, uninterpretable formula. Imagine a traveler who knows all too well that the Isonzo river is identical with the Isonzo. She might still be surprised upon arriving at the river Soča (advertised in brochures as the whitewater rafting paradise of Slovenia), when she learns that it is one and the same river as the Isonzo, the scene of numerous battles in World War I that she had read about in history books. (Indeed, I myself was in for that surprise when travelling to Slovenia: that the Soča is one and the same as the Isonzo constituted a discovery). Thus if we interpret □ as, say, epistemic necessity, then (1) is indeed true given what our traveler knows, while (2) is false. Similarly, if we interpret □ as analytic necessity—as Quine does—(1) is true given the meanings of the words featured (all of which are familiar to our traveler), while (2) is false (given her subsequent discovery):

(1) It is necessarily true that the Isonzo is identical with the Isonzo.
□ Isonzo = Isonzo true

(2) It is necessarily true that the Soča river is identical with the Isonzo.
□ Soča = Isonzo false

The truth value assignments for (1) and (2) remain unaltered even if we interpret □ as logical necessity, truth in virtue of the meanings of the logical vocabulary. Indeed, it will help our exegesis to introduce the category of *linguistic necessity* to cover both analytic and logical necessity: for both concern truth in virtue of the meanings of certain expressions; the difference is only whether we consider the meanings of all vocabulary items or just the logical ones. Crucially, in formulating his argument (a), Quine's concern was with linguistic necessity (what he called strict necessity), although he rarely made this explicit, especially in his later work.

We can generalize over (1) to arrive at one of the suspect formulas:

(3) There is a thing x, such that x is identical with the Isonzo.
$\exists x \Box (x = \text{Isonzo})$

Interpretive trouble ensues: What is this river which, according to (3), is necessarily identical with the Isonzo? According to (1), from which (3) was inferred, it is the Isonzo, that is, the Soča; but to suppose this would conflict with the fact that (2) is false. In a word, to be necessarily [in the linguistic sense] identical

with the Isonzo is not a trait of the river, but depends on the manner of referring to the river. (adapted from Quine 1953, 148)

(3) is an instance of quantifying in, that is, binding the variable *x* within the scope of the modal operator □ by the quantifier ∃*x*, which is outside the scope of □. This is the sort of construction that spells interpretive trouble when it comes to linguistic necessity, according to Quine. He did not think he has given a general argument against quantifying into any modal context whatsoever (as many interpreters at the time thought)—he says this much in the following passage (see also Quine 1963):

> What has been said of modality in these pages relates only to strict [that is, linguistic] modality. For other sorts, for example, physical necessity and possibility, *the first problem would be to formulate the notions clearly and exactly*. Afterwards we could investigate whether such modalities, like the strict ones, cannot be quantified into without precipitating an ontological crisis. The question concerns intimately the practical uses of language. … In discussions of physics, naturally, we need quantifications containing the clause '*x* is soluble in water', or the equivalent in words; but … we should then have to admit within quantifications the expression … 'necessarily if *x* is in water then *x* dissolves'. Yet we do not know whether there is a suitable sense of 'necessarily' into which we can so quantify. (Quine 1953, 158-159.; emphasis added)

Here, Quine poses a challenge: quantifying in spells interpretive trouble for linguistic notions of necessity; when considering how interpretation would go with alternative notions of necessity (physical necessity, for example), first, those notions should be clarified, then the question of interpreting quantifying in can be raised. Accordingly, we can expand (a):

(a) Quine's lasting argument: When considering the (then-)established notion of necessity, that of linguistic necessity, certain modal logic formulas (those involving quantifying in) lack sense, they cannot be interpreted.

Quine's associated challenge: Clarify an alternative notion of necessity, and if the need for interpreting quantifying in arises with respect to that notion, then check that there is no interpretive trouble there.

In what follows we will unpack Quine's lasting argument (following primarily Burgess 1997), and see how Kripke responds to Quine's associated challenge by bringing in the notion of metaphysical necessity. But before that, let us introduce a preliminary distinction between *de dicto* and *de re* statements:

a *de dicto* ("about the sentence") statement:

(4) Necessarily, all single people are unmarried.

"The following is necessary: all singles are unmarried."

a *de re* ("about the thing") statement:

(5) All single people are necessarily unmarried.

"All singles bear the modal attribute of being necessarily unmarried."

Consider, for a moment, the counterfactual sense of necessity. According to it, (4) is true, for in all counterfactual circumstances, everyone who is single is unmarried. Meanwhile, (5) is false: for those who are in fact single might, in an alternative scenario, have gotten married instead—they are not single in all counterfactual situations.

Now we can spell out step by step Quine's request for interpreting quantifying in, this time with linguistic necessity at hand:

Step 1. First we need to make sense of the open formula '$\Box$ (x = Isonzo)'.

Step 2. This requires making sense of *de re* modal claims.

Step 3. The *de dicto* claims at hand are (1) and (2), and their *de re* counterparts are (1r) and (2r):

(1) $\Box$ Isonzo = Isonzo	true
(1r) It is true of the Isonzo that it is necessarily identical with the Isonzo.	
$\exists x$ (x = Isonzo & $\Box$ x = Isonzo)	?
(2) $\Box$ Soča = Isonzo	false
(2r) $\exists x$ (x = Soča & $\Box$ x = Isonzo)	?

But the notion of linguistic necessity—about truth given the meanings of expressions—provides guidance for interpreting *de dicto* modal claims only; there is no direct guidance for making sense of *de re* modal claims. (For what might that *river* be that is *analytically* or *logically* identical with the Isonzo, given that (1) and (2) differ in truth value?)

Step 4. We have two strategies for interpreting (1r) and (2r), but both turn out unacceptable.

Step 5. The first strategy for interpreting *de re* modal claims is:

the unselective strategy: the *de dicto* statement yields its *de re* counterpart—for any proper name whatsoever.

This yields an unacceptable outcome: we have objects with contradictory properties: the river Isonzo a.k.a. Soča is at once analytically identical with the Isonzo (qua Isonzo) and not analytically identical with it (qua Soča). The cost of avoiding this is high: we have to give up on the idea that the truth of *de dicto* modal claims may in part depend on the words and names used. But linguistic necessity is supposed to be about truth in virtue of the meaning of certain expressions, so this option is unacceptable.

Step 6. The second strategy for interpreting *de re* modal claims is:

the selective strategy: *de dicto* modal claims yield their *de re* counterparts in selected cases only—with respect to standard names.

For example, if 'Isonzo' counts as a standard name while 'Soča' does not, then we cannot get (2r) from (2). But then we would have to make arbitrary decisions about which natural-language proper name to regard as standard: 'Cicero' or 'Tully'? 'Burma' or 'Myanmar'?

Step 7. With linguistic necessity, the standard names featured in the selective strategy lead to an arbitrary form of essentialism:

> "Evidently, the reversion to Aristotelian essentialism … is required if quantification into modal contexts is to be insisted on. An object, of itself and by whatever name or none, must be seen as having some of its traits necessarily and others contingently, despite the fact that the latter traits follow just as analytically from some ways of specifying the object as the former traits do from other ways of specifying it." (Quine 1953, 155)

In other words, with standard names chosen arbitrarily, we end up with arbitrary choices for what is and what is not analytically true of an object. So the distinction between essential and accidental properties of objects—this is what essentialism is committed to—will be arbitrarily drawn.

For linguistic necessity, this seven-step argument does conclusively show that interpreting *de re* modal claims spells trouble whichever interpretive strategy we follow, making the first half of (a) a lasting argument indeed. The second half of (a), Quine's challenge is: we have (yet) to locate a notion of necessity which allows us to make sense of *de re* modal statements without running into unacceptable consequences. It is to this challenge that (Kripke 1980/2007) provides a response:

(b) Kripke's response to Quine's challenge: with the notion of counterfactual (metaphysical) necessity spelled out, interpreting *de re* modal claims is no longer problematic.

The following seem like plausible candidates for true *de re* modal claims: Cicero was necessarily human, but was only contingently born outside Rome; there is a counterfactual situation in which he was born in Rome, but there is no counterfactual situation in which *he* fails to be human. With this counterfactual notion of necessity at hand, our interpretation of *de re* modal claims is directly given; there is no need for either the selective or the unselective strategy of piggybacking on *de dicto* modal claims.

Ruzsa, along with contemporary commentators of Quine, thought that Quine's argument against modal logic (a, that is) targeted all stripes of modality. Hence, they thought that providing a framework for accommodating formulas with quantifying in—Kripke's formal work from the 1950s and 1960s (for example, Kripke 1963)—suffices to show that quantified modal logic is viable after all. (Indeed, commentators were in a difficult position because alongside his lasting argument, Quine also gave other, more general arguments against inter-

preting quantified modal logic, without properly distinguishing them from one another; for problems raised for some of the general arguments, see for example Kaplan 1986 and Fine 1989, 1990.) Ruzsa and others also considered Quine's charge that quantified modal logic comes with a high price tag—embroilment in essentialism, that is, commitment to a distinction between essential and accidental properties of objects (in Step 7)—to arise for quantified modal logics of all stripes. Yet again, there is a crucial detail to realize about Quine's argumentative strategy: his claim that essentialism is untenable is doubly embedded within his argument: first, it is featured within one of the interpretive strategies for making sense of *de re* modal claims (the one based on standard names); and second, we get an arbitrary, and hence objectionable form of essentialism specifically in the case of linguistic necessity, precisely because of the need to rely on standard names. In short, the lasting argument of Quine's does not claim that across the board, there is a problem with interpreting *de re* modal formulas; nor does it claim that across the board, essentialism is objectionable. And the response for his challenge calling for an alternative notion of modality where the interpretive problem is resolved, is in fact met not in Kripke's formal work, but in Kripke's observations about the semantics of natural language, when, in propounding his rigidity thesis, he also clarified the notion of counterfactual necessity (b, that is). (a) and (b) are then the missing links that complete the otherwise admirably detailed and illuminating picture of state-of-the-art modal logic and modal semantics that Imre Ruzsa relayed to Hungarian readers back in 1984.

REFERENCES

Burgess, John, 1997, Quinus ab omni nævo vindicatus. In A. A. Kazmi (ed.), *Meaning and Reference: Canadian Journal of Philosophy Supplement* 23. 25–65.

Copi, Irving M. and Gould, James A (eds.), 1964/1985, *Kortárs tanulmányok a logikaelmélet kérdéseiről* (*Contemporary Readings in Logical Theory*. In Hungarian, trans. D. Bánki, B. Dajka, I. Faragó-Szabó, K. G. Havas, L. Hársing, A. Máté, K. Solt and L. Urbán). Budapest, Gondolat. Includes translations of Quine 1947 and 1963.)

Fine, Kit, 1989, The problem of *de re* modality. In J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*. Oxford, Oxford University Press, 197–272.

Fine, Kit, 1990, Quine on Quantifying in. In C. A. Anderson and Joseph Owens (eds.), *Proceedings of the Conference on Propositional Attitudes*. Stanford: CSLI, 1–26.

Kaplan, David, 1986, Opacity. In L. E. Hahn and P. A. Schilpp (eds.), *The Philosophy of W. V. Quine*. La Salle, IL, Open Court, 229–289.

Kripke, Saul, 1963, Semantical considerations in modal logic. *Acta Philosophica Fennica* 16, 83–94.

Kripke, Saul, 1971/2004, Azonosság és szükségszerűség (Identity and Necessity. in Hungarian, trans. F. Csaba). In K. Farkas and F. Huoranszki (eds.), *Modern metafizikai tanulmányok*. Budapest: ELTE Eötvös, 39–68.

Kripke, Saul, 1980/1997, Névadás és szükségszerűség (részletek az első és a második előadásból). (Short excerpts from *Naming and Necessity* in Hungarian, trans. M. Wappel). *Helikon* 43, 410–426.

Kripke, Saul, 1980/2007, *Megnevezés és szükségszerűség*. (*Naming and Necessity*. In Hungarian, trans. T. Bárány). Zs. Zvolenszky (ed.), Budapest. Akadémiai.

Farkas, Katalin and Kelemen, János, 2002, *Nyelvfilozófia*. (*Philosophy of Language*. In Hungarian.) J. Bárdos (ed.), Budapest, Áron.

Neale, Stephen R., 2000, On a milestone of empiricism. In P. Kotatko and A. Orenstein (eds.), *Knowledge, Language and Logic: Questions for Quine*. Dordrecht, Kluwer 237–346.

Quine, Willard V., 1947, The problem of interpreting modal logic. *Journal of Symbolic Logic* 12, 43–48. (Hungarian translation in Copi–Gould 1964/1985.)

Quine, Willard V., 1951/1973, Az empirizmus két dogmája. (Two Dogmas of Empiricism. In Hungarian, trans. I. Faragó-Szabó.) *Magyar Filozófiai Szemle* 47, 225–239.

Quine, Willard V., 1953/2002, Reference and Modality. (Reference and modality. In Hungarian, trans. E. Boldizsár.) In Quine 2002, 225–251.

Quine, Willard V., 1960, *Word and Object*. Cambridge, MA, MIT Press.

Quine, Willard V., 1963, (1962). Reply to Professor Marcus. In M. Wartofsky (ed.), *Boston Studies in the Philosophy of Science*. Dordrecht, D. Reidel, 97–104. (Hungarian translation in Copi–Gould 1964/1985.)

Quine, Willard V., 2002, *A tapasztalattól a tudományig*. (*From Experience to Science*. Gábor Forrai (ed.) Budapest, Osiris.

Ruzsa, Imre, 1984, *Klasszikus, modális és intenzionális logika*. (*Classical, Modal and Intensional Logic*. In Hungarian.) Budapest, Akadémiai.

Ruzsa, Imre, 1988, *Logikai szintaxis és szemantika* vols. 1–2. (*Logical Syntax and Semantics*. In Hungarian.) Budapest, Akadémiai.

Sainsbury, Mark, 1995/1997, Filozófiai logika. (Philosophical logic. In Hungarian, trans. K. Farkas.) In A. C. Grayling (ed.), *Filozófiai kalauz*. (*Philosophy: A Guide through the Subject*. In Hungarian.) Budapest, Akadémiai, 73–140.

Zvolenszky, Zsófia, 2007, *Megnevezés és szükségszerűség* – Négy évtized távlatában (Four decades of *Naming and Necessity*. In Hungarian.) In Kripke 1980/2007, 151–218.

Anna Brożek

On the so-called embedded questions

Abstract. The analysis of dependent questions plays an important role in the general theory of questions. Dependent questions are expressions which are parts of compound questions and are isomorphic with some independent questions (scil. questions *sensu stricto*). One may meet the tendency to explicate the sense of independent questions by the sense of dependent ones, e.g. the sense of questions such as "Where is Budapest situated?" is explicated by the sense of sentences such as "*A* knows where Budapest is situated", where the second contains the first as a part. The analysis of dependent questions is often the point of departure for constructing set-theoretical or possible worlds semantics for independent questions. In my opinion, these tendencies are abortive and lead to irrelevant explications of the sense of questions *sensu stricto*. But on the other hand, semiotic functions of the so-called dependent questions as parts of compound expressions require deeper analysis. My paper contains a proposal of such an analysis.

1 SUPPOSITIONS

The following are the points of departure for my paper:
(A) Following Kazimierz Ajdukiewicz (1956/1978), I distinguish stating from expressing; for instance, if a person *P* utters the sentence '*p*' then this sentence *states* the occurrence of a certain state of affairs (namely that *p*), and *expresses* the conviction of *P* that *p*.
(B) I adopt the concept of name introduced by Stanisław Leśniewski: "a name" means an expression that can occur as a subject of a subject-predicate sentence, or as part of a predicate in a subject-predicate sentence of the form "… is …"; this means that both "Imre Ruzsa" and "the inventor of the system of intensional logic with semantic value gaps" are names, since the first is the subject and the second is part of the predicate within the sentence "Imre Ruzsa is the inventor of the system of intensional logic with semantic value gaps. "

(C) I distinguish designating from denoting:
(a) a name *N* designates an object *A* iff *N* can be truly predicated about *A* or *A* can be indicated by *N*;
(b) the denotation of *N* is the set of all things designated by *N*—its designata.

2 *SENSU STRICTO* QUESTIONS VS. NOMINALIZED QUESTIONS

The communicative sense of every *sensu stricto* question is composed of three elements: cognitive, incognitive and volitional. For instance, if a person *P* asks seriously:

(1) Where was Imre Ruzsa born?

then *P* expresses:

(a) that *P* is convinced that Imre Ruzsa was born somewhere;
(b) that *P* does not know where Imre Ruzsa was born;
(c) that *P* wants to know where Imre Ruzsa was born.

These three components of sense distinguish questions as a specific class of expressions.

It is convenient to describe this situation using the metaphor of a picture of a situation. A person seriously uttering a question has a mental picture of a situation with an epistemic gap, and wants to fill that gap.

Questions have the following general form:

(2) ?*x* (*Fx*)

i.e. "For which *x* it is a fact that *Fx*?" This scheme was first proposed by Kazimierz Ajdukiewicz in 1923 (13 years before Rudolf Carnap, who is usually credited with introducing it). Later, Tadeusz Kubiński (1970) used Ajdukiewicz's formulation in the construction of his systems of erotetic logic (*i.e.* his logic of questions), noting the analogy between the role of the questionmark in (2) and the role of quantifiers in declarative sentences.

Example (1) is a completive question; in what follows, I shall use examples of such questions only. However, my remarks may easily be expanded to other kinds of questions (selective and confirmative ones), since all questions—after appropriate preparation—come under the scheme (2). The only difference is in the scope of the unknown and the way of defining it.

3 REDUCTIONS

Logical theories of questions usually simplify the sense of questions: they reduce it to an exclusively cognitive, exclusively incognitive, or exclusively procognitive element. In my opinion, none of these three elements should be omitted when

constructing a materially adequate theory of questions.
One could argue that since I am able to list the elements of the sense of questions, I should agree that questions of the form (1) uttered by me may be reduced to the conjunction of declarative sentences of the form:

(3) (I know that $\exists x\,(Px)) \wedge \sim\exists x$ (I know that $Px) \wedge \forall x\,(Px \rightarrow$ I want to know that Px)

where the variable x ranges over (names of) places and P is the property of being-a-place-of-birth-of-Imre-Ruzsa. There are at least two reasons why (3) is not an adequate paraphrase of (1). Firstly, the expression "I want to know that Px", being a component of (3), is semantically defective. We encounter it sometimes in ordinary situations, but only in the sense "I want you to tell me that Px", which is not of course the proper sense of "to know". Secondly, the sense of (3) is essentially different from the sense of (1): one may express this difference by saying that (3) states what (1) expresses. To state that one possesses experiences motivating one to pose a question is not the same as actually to pose that question. One may experience everything that is stated in (3) without asking (1) at all.

Both these reasons for rejecting the paraphrase (3) are important in the case of so-called embedded questions.

4 EMBEDDED QUESTIONS: MISUNDERSTANDINGS

Let us use the term "embedded questions" to denote the set expressions isomorphic to *sensu stricto* questions but being proper parts of declarative sentences (not merely quoted in them).

The first misunderstanding connected with embedded questions is that one may reduce *sensu stricto* questions to declarative sentences containing embedded questions (*scil*. that one may explicate the sense of *sensu stricto* questions through the sense of embedded ones).
Such reductions are proposed, e.g., in the imperative-epistemic tradition in the theory of questions where exclusively embedded questions are used—as a certain step—within paraphrases of questions (see Åquist 1965).

For instance, at the point of departure in one of the versions of this concept, (1) is paraphrased as follows:

(4) Let it be the case that I know where Imre Ruzsa was born.

In the next step, sentences like (4) are paraphrased in such a way that they do not contain embedded questions: they are equal to sentences containing the predicate "know that" which has been well analyzed by logicians, e.g.:

(5) $\forall x$ (Imre Ruzsa was born in $x \rightarrow$ let it be the case that I know that Imre Ruzsa was born in x).

The paraphrase (5) ignores the aforementioned distinction between expressing and stating or describing: (5) describes components of the sense expressed in (1). In addition, paraphrase (5) violates our linguistic intuitions by introducing the expression "I want to know that" which (as was previously observed) seems to be incorrect.

The second misunderstanding connected with the concept of embedded questions is that the point of departure of so-called erotetic semantics should be (or at least could be) the semantics of embedded questions (see Lahiri 2002). Such a view is incorrect simply because embedded questions are not *sensu stricto* questions. In fact, according to such an approach, one constructs nothing over and above a semantics of declarative sentences containing embedded questions.

However, the problem of the sense of embedded questions is intriguing.

5 EMBEDDED QUESTIONS AS NAMES

Let us now analyze the problem of what the function of embedded questions in declarative sentences is. Consider the sentence:

(6) Ferenc knows where Imre Ruzsa was born.

This contains as a component an expression isomorphic to (1), i.e. the expression "Where was Imre Ruzsa born?" (the only and usually ignored difference is inversion). Thus we notice an analogy between (6) and the sentence:

(7) Ferenc knows that Imre Ruzsa was born in Budapest.

since (7) contains the sentence:

(8) Imre Ruzsa was born in Budapest.

as a component. There are many possible analyses of (7); the most popular of these takes the expression "knows that" as the main predicate with two arguments: name-argument and sentence-argument. In another interpretation—a less popular but more accurate one—the predicate "know" takes two name-arguments, the second argument being a name of a suitable situation. In Polish, one may even say:

(9) Ferenc wie to, że Imre Ruzsa urodził się w Budapeszcie.

and see explicitly the «*reificator*» "to, że" (Eng. "that") of the sentence occurring after "wie" (Eng. "knows"). In word-for-word translation, the start of sentence (9) has the form "Ferenc knows this [fact] that…".

Let us analyze sentence (6) analogously. We accept that the predicate "know" in (6) possesses two name-arguments (and not name-argument and question-argument). Again, we may say in Polish:

(10) Ferenc wie to, gdzie urodził się Imre Ruzsa.

The initial phrase of (10)—"Ferenc wie to, gdzie" —has the structure of the type "Ferenc knows this [fact] where…". It is hypothesized that embedded questions are always preceded by an explicit or implicit reificator (or nominalizator). There

are several arguments in favor of such an analysis of (at least some) embedded questions.

Firstly, the following expression, being a paraphrase of (6), possesses explicitly two name-arguments:

(11) Ferenc knows the place of Imre Ruzsa's birth.

Secondly, in Polish (and probably some other languages), embedded questions which occur at the beginning of the sentence possess an obligatory reificator, e.g.

(12) To, gdzie urodził się Imre Ruzsa, ciągle pozostawało dla Ferenca tajemnicą.

Maybe the lack of reificator *inside* the sentence is caused only by the specific connectivity of some verbs. In English, an explicit reificator of the type "this [fact]" does not appear:

(13) Where Imre Ruzsa was born was still a mystery for Ferenc.

But the position of the embedded question at the beginning of the sentence, and its specific word order, make its name-like character more clear.

Thirdly, embedded questions do not perform the communicative function of questions (mentioned in section 1). The person uttering (6) does not reveal the desire to fill a gap in a picture of a situation. The situation is similar with embedded sentences. The communicative function of sentences consists in expressing convictions. But a person uttering (7) does not express the conviction that Imre Ruzsa was born in Budapest—only the conviction that Ferenc knows that Imre Ruzsa was born in Budapest.

In what follows, I assume that at least some embedded questions are nominalized questions. I also assume that a question that has undergone nominalization does not perform the same functions as a *sensu stricto* question (just as a nominalized sentence does not perform the same functions as the *sensu stricto* sentence). Nominalized questions are names and—like every name—they have referential functions, *scil.* they designate something. The problem is to say what the designata of nominalized questions are.

6 THE DESIGNATA OF NOMINALIZED QUESTIONS

Consider the following sentences containing nominalized sentences:

(14) Ferenc knows that Imre Ruzsa was born in Budapest.

(15) Ferenc was convinced that Imre Ruzsa was born in Budapest.

(16) That Imre Ruzsa was born in Budapest influenced his life.

(17) That Imre Ruzsa was born in Budapest encourages Ferenc to take part in the conference *Logic, Language, Mathematics* devoted to the author of *Modal Logic with Descriptions*.

What does the name *N*: "that Imre Ruzsa was born in Budapest" in sentences (14)–(17) refer to? Generally speaking, one usually assumes that nominalized sentences designate states of affairs (i.e. elements of reality or objects abstracted from reality), or judgments (i.e. elements of thoughts or objects abstracted from thoughts). The following are possible approaches to this problem (two uniform and one mixed approach):

(a) every nominalized sentence refers to situations;

(b) every nominalized sentence refers to judgments;

(c) nominalized sentences are ambiguous: in one sense they refer to situations; in the second sense to judgments.

At first glance, it seems that in (14) and (15), the name *N* refers to a state of affairs, whereas in (15) and (17) it refers to a judgment. This implies the mixed solution: it is hard to defend any homogeneous one. I omit this problem since it does not relate to the main theme of my investigations.

Let us stress once again: nominalized sentences perform different semiotic functions than non-nominalized sentences do. *Sensu stricto* sentences describe the occurrence of states of affairs and are used to express convictions. Nominalized sentences designate states of affairs or judgments; they are not used to express convictions.

The situation appears similar to the case of nominalized questions: they perform a different semiotic function than *sensu stricto* questions.

Consider the question:

(18) Who was born in Budapest?

Let us keep in mind that somebody who seriously poses such a question possesses a gapped picture of a situation; this situation involves the relation _ was-born-in_, with the gap in the first argument, the second being known (it is Budapest). Somebody who seriously utters (18) wants to fill this gap.

Let us call a person's particular experiences, composed of these three components (cognitive, incognitive and volitional), "inquiries". An inquiry understood in such a way—as expressed in questions—is a counterpart of the convictions expressed in sentences.

In reality, there are no *gapped* states of affairs. But our pictures of real situations possess gaps. However, questions are correlated with some specific *full* situations—situations which one asks for, pictures of which we aim to possess when we pose questions. Sentences stating the existence of these states of affairs constitute accurate (i.e. true and direct) answers to questions. Let us call states of affairs which are correlates of true answers of a given question "supplementations" of that question. It should be stressed that some questions—in particular, improperly posed questions—do not have supplementations, since they do not possess accurate answers.

A supplementation of a question of the type '?x (Px)' is identical with such a state of affairs whose occurrence is stated by a true substitution of the formula

'*Px*'. For instance, the fact that Imre Ruzsa was born in Budapest is one of the supplementations of the question (18) (this question has of course many other supplementations).

Nominalized sentences are not suitable to express conviction; nominalized questions are likewise not suitable to express inquiries. However they are suitable to indicate inquiries or supplementations.

Three possibilities may be considered:

(a) every nominalized question refers to an inquiry;

(b) every nominalized question refers to supplementation;

(c) nominalized questions are ambiguous: in one sense, they refer to inquiries, in another sense, to supplementations.

7 INQUIRIES AS CORRELATES OF NOMINALIZED QUESTIONS

Consider the set of famous logicians, the nominalized question "[that] who was born in Budapest", and its role in sentences:

(19) Ferenc asked [about] who was born in Budapest.

(20) Ferenc knows who was born in Budapest.

(21) Ferenc wanted to know who was born in Budapest.

(22) Who was born in Budapest was a mystery.

(23) Who was born in Budapest influenced the fate of the city.

(24) Who Budapest's citizens are proud of depends on who was born in Budapest.

What would it mean to say that the nominalized questions in (19)–(24) refer to inquiries? *Sensu stricto* (not nominalized) questions communicate the desire to fill a gap in the picture of a situation. It seems that in the question-state indicated (by a nominalized sentence), the volitional element is not included; the only indicated elements are the cognitive and incognitive ones. In other words, nominalized sentences designate *gapped* pictures of situations.

Such a solution is implied first of all in contexts in which nominalized questions are arguments of predicates such as "ask", "wonder", "inquire", "guess", etc.

8 "FULL" STATES OF AFFAIRS AS CORRELATES OF NOMINALIZED QUESTIONS

In some contexts, it seems that nominalized questions refer to supplementations. In particular, the name "that-*q*" designates the state of affairs designated by "that *p*", where '*p*' is a true answer to '*q*'.

It was observed long ago that in the case of the verb "know", the following dependencies hold:

(25) If Ferenc knows where Imre Ruzsa was born and Imre Ruzsa was born in Budapest, then Ferenc knows that Imre Ruzsa was born in Budapest.

The same applies in the case of such verbs as "say" (as a synonym of "inform", and not "utter"), "be surprised", "it is clear that". But notice that such a solution (i.e. considering nominalized questions as names of supplementations) can also be applied in cases (19)–(24). For instance, it is a certain state of affairs which is unknown (say, it is a mystery) (example (22)); they are certain facts such that a relation of dependency holds between them (example (24)), etc.

9 PRAGMATIC PROPERTIES OF NOMINALIZED SENTENCES AS NAMES OF STATES OF AFFAIRS

Let us agree that the following sentence is true:

(26) Imre Ruzsa was born in Budapest.

Now, consider the name:

(27) [that] Imre Ruzsa was born in Budapest.

Let us agree also that the name (27) designates a certain fact; let us call this fact '*f*'. Now, consider the following names:

(28) [that] somebody was born in Budapest

(29) [that] Imre Ruzsa was born somewhere

Both these names designate *f*; but (28) designates additionally other states of affairs.

Consider, finally, the following names:

(30) [that] who was born in Budapest

(31) [that] where Imre Ruzsa was born

What are their designata?

If we agree that nominalized questions designate supplementations (in my sense), then the designata of (30) and (31) are the same as in the case of (28) and (29). It seems that (30) designates *f* and other states of affairs, whereas (31) designates only *f*.

It is not surprising that two names designate the same object. But what is the difference between the two types of names?

Note that we use nominalized questions in specific situations, i.e. when we cannot indicate the supplementation precisely or when we do not want to indicate the filling of a gap. This may be easily seen from the following examples:

(32) Ferenc knows who was born in Budapest (but I do not know).

(33) I know who was born in Budapest (but I shall not say).

Moreover, nominalized questions are used when we want to express general dependencies:

(34) How successful our conference is depends on what the weather is like.

(35) Whether I understand Imre Ruzsa's works depends on what the language in which they were published is.

These sentences have a more general sense than sentences with the functor "if-then":

(36) If the weather is dreadful, then our conference will turn out well.

(37) If the language in which Imre Ruzsa's works were published is Hungarian, then I do not understand them.

Again—in contrast with (34) and (35)—in the case of (36) and (37), we do not know or do not reveal consciously what exactly this relation consists in.

10 SUMMARY

Let me summarize my views.

- Firstly, embedded questions are not *sensu stricto* questions.
- Secondly, questions in embedded contexts are (explicitly or implicitly) nominalized questions.
- Thirdly, nominalized questions are ambiguous: in one sense, they designate question-states, while in another, they designate supplementations.
- Fourthly, nominalized questions have denotations similar to some nominalized sentences, but they perform more sophisticated pragmatic functions.

REFERENCES

Ajdukiewicz, K., 1923, O intencji pytania 'Co to jest *P*?' (On the intention of the question 'What is *P*'). *Ruch Filozoficzny* 8, 152b-153a.

Ajdukiewicz, K., 1956/1978, Conditional statement and material implication (Okres warunkowy a implikacja materialna.) In: K. Ajdukiewicz, *The Scientific World-Perspective and Other Essays*, Dordrecht, Reidel, 222–238.

Åquist, L., 1965, *A New Approach to the Logical Theory of Interrogatives*. Upsala, Almquist & Wilksell.

Kubiński, T., 1970a. *Wstęp do logicznej teorii pytań*. (*Introduction to the Logical Theory of Questions*.) Warszawa, PWN.

Lahiri, U., 2002, *Questions and Answers in Embedded Contexts*. Oxford, Oxford University Press.

Gyula Klima

Natural Logic, Medieval Logic and Formal Semantics

Abstract. This paper provides a comparative analysis of the issue of natural logic: the "formalizational approach", prevalent in contemporary logic, and the "regimentational approach", prevalent in medieval logic, as exemplified by the 14th-century nominalist philosopher, John Buridan. The differences between the two are not as great as they may first appear: a little tweaking of standard quantification theory can take us surprisingly close to Buridan's logic. However, as the conclusion of the paper points out, there still are some fundamental differences between the resulting "reconstructed Buridanian logic" and Buridan's logic itself, discussed in detail in the author's recent monograph.

NATURAL LANGUAGE AND THE IDEA OF A "FORMAL SYNTAX" IN BURIDAN

The idea of the universality of logic is based on the conviction that despite the immense diversity of human languages, there are certain invariant features of human reasoning, carried out in any natural language whatsoever, that allow the formulation of universal logical laws, applicable to any language. It is precisely for expressing these universal, invariant aspects of human reasoning that in modern logic we construct an artificial language, which is then conceived to serve as a more direct linguistic expression of those invariant conceptual structures that are variously expressed by various natural languages.

But this is not the only possible way to achieve the desired transparency of conceptual structure through the transparency of syntax. The way the 14th-century nominalist philosopher, John Buridan (and medieval logicians in general) achieved this was by using, *not* a full-fledged artificial language, but an artificially "regimented" Latin. We can get a nice, yet relatively simple illustration of what this "regimentation" of Latin consists in if we take a closer look at how Buridan introduces the idea that every simple categorical proposition of Latin can be

reduced to the "canonical" subject-copula-predicate form. After briefly stating the division of propositions into categorical and hypothetical, and the description of categorical propositions as those that consist of subject and predicate as their principal parts, Buridan remarks:

> … a verb has to be analyzed into the verb 'is' as third adjacent, provided that the proposition is assertoric [*de inesse*] and in the present tense [*de praesenti*], and into the participle of that verb, as for example, 'A man runs' is to be analyzed into 'A man is running', and similarly, 'A man is' into 'A man is a being'.[1]

English speakers might at once notice that the proposed transformation does not always yield equivalent sentences, given the tendency in English to use the simple present tense to signify habitual action as opposed to the continuous present tense, consisting of the copula and the appropriate participle, which is used to express present action. For instance, if I say 'I smoke', I may simply want to express that I am a smoker, a person who has the habit of smoking, but this does not mean that I am actually smoking, which would properly be expressed by the sentence 'I am smoking'. In fact, in accordance with Buridan's theory of predication, according to which the affirmative copula expresses the identity of the *supposita*, that is, the referents of the terms flanking it, a more appropriate rendering of his proposed transformation would be 'I am [identical with] someone smoking'.

But Buridan might answer that this is merely a difference in the different syntactical "clues" a different language uses to indicate a different sort of underlying conceptual construction. The simple present tense of English, when it is used to signify habitual action, should then not be analyzed into a participle and a simple assertoric copula, but *perhaps* (somewhat unidiomatically) into a participle and an adverbially modified copula, as in 'I am usually smoking',[2] where we just express in the surface syntax of this sentence an adverbial modifier that is unmarked in the simple tense (as is the implicit copula), but which is nevertheless present in the corresponding mental proposition. In any case, it is in this spirit that Buridan answers four questions he raises in connection with the issue of the "canonical form" of categorical propositions:

[1] SD 1.3.2.

[2] Alternatively, one might say that the best explication of 'I smoke' expressing the habit is 'I am a smoker', where the nominal definition of 'smoker' may explicate the habit, as in 'x is a smoker iff x has the habit of smoking'. But as Buridan often remarks, "examples are not to be verified", i.e., it does not matter whether we provide "the right analysis" here, as long as it serves to illustrate the point.

> But then some questions arise. The first concerns what such a copula signifies. The second is whether that copula is a principal part of a categorical proposition. The third question is whether the proposition 'The one lecturing and disputing is a master or a bachelor' is categorical or hypothetical; for it seems that it is hypothetical, since it has two subjects and two predicates. The fourth question is the same concerning the proposition 'A man who is white is colored'; for it seems that it is hypothetical, since here we have two subjects, two predicates and two copulas; and also because it seems to be equivalent to 'A man is colored, who is white' which is apparently hypothetical.[3]

In his reply, Buridan provides the rationale for the canonical subject-copula-predicate structure in terms of what modern linguists would certainly recognize as "deep structure", and what for Buridan is the conceptual structure of the corresponding mental proposition:

> To the first question we should reply that a spoken proposition has to signify a mental proposition [..]. A mental proposition, however, involves a combination of concepts [*complexio conceptuum*], and so it presupposes in the mind some simple concepts, to which it adds a complexive concept, by means of which the intellect affirms or denies one of those [presupposed simple] concepts of the other. Thus, those presupposed concepts are the subject and the predicate in a mental proposition, and they are called the matter of the mental proposition, for they are presupposed by the common form of a proposition, just as matter is presupposed by the substantial form in [the process of] generation. And then it is clear that the subject and the predicate of the spoken proposition signify in the mind the subject and the predicate of the mental proposition. The copula 'is' signifies an affirmative complexive concept, whereas the copula 'is not' signifies a negative complexive concept; and the intellect is unable to form that complexive concept except when it has formed those which are the subject and the predicate, for it is impossible to have the combination [*complexio*] of the predicate with the subject without the predicate and the subject. And this is what Aristotle meant[4] when he said that 'is' signifies a certain composition which cannot be understood without the components.[5]

What fundamentally justifies sticking to the idea of this "canonical form" according to Buridan is that no matter how a mental proposition is expressed

[3] SD 1.3.2.
[4] Aristotle, *On Interpretation*, 1, 16b24.
[5] SD 1.3.2.

in the ("surface") syntax of a spoken language, the concept of the copula is there in the mental proposition. Therefore, indicating it in the syntax of the spoken proposition merely explicates the presence of the complexive concept of the copula in the corresponding mental proposition. Indeed, this explication is *always* justified because, as Buridan now explains in his answer to the second question, that complexive concept *has to* be present in *any* mental proposition:

> To the second question we should reply that the copula is truly a principal part of the proposition, because there could not be a categorical proposition without it; and also because it can be compared to a form of the subject and the predicate, and the form is a principal part of a composite.[6]

Thus, given that the copula is the "formal", principal part of a categorical proposition, i.e., it is that complexive concept (proposition-forming functor) in the mind without which the concepts corresponding to the terms would not constitute a proposition, it is obvious that no matter how complex those terms and the corresponding concepts are, if they are joined by one copula, then they form one proposition. This is precisely the basis of Buridan's answer to the third question:

> To the third question we should reply that that proposition is categorical; for it does not contain two categoricals, as there is only one copula here; neither are there several subjects, nor several predicates here, for the whole phrase 'the one lecturing and disputing' is a single subject […], namely, a conjunctive subject, and the whole phrase 'master or bachelor' is likewise a single predicate, although disjunctive.[7]

As this remark clearly illustrates, Buridan would allow complex terms in either the subject or the predicate positions of otherwise simple, categorical propositions. In fact, given the potentially unlimited complexity of the terms of categorical propositions, these propositions may exhibit a *very* complex structure *within* their terms, despite the simplicity of the general subject-copula-predicate scheme. For it is not just the (iterable) "Boolean" operations of disjunction, conjunction and negation that can yield potentially infinite complexity in these terms, but also the fact that any proposition can be turned into a term (by forming a "that-clause") or into a determination of a determinable term (in the form of a relative clause). For example, if we take the proposition 'A man is running', it can easily be transformed into the term 'That a man is running', which can then

[6] SD 1.3.2.

[7] Ibid. Note that in Buridan's usage, 'hypothetical' in this context simply means 'complex', as opposed to the widespread modern usage that makes it equivalent to 'conditional'.

be the subject of another proposition, e.g., 'That a man is running is possible' or a part of another more complex term in another proposition, as in 'That a man is running is believed by Socrates'. Again, taking the proposition 'A man is white', and inserting a relative pronoun after its subject, we get another complex term 'A man who is white', which can then be the subject in the proposition 'A man who is white is colored'.

Now if we look at this proposition in this way, namely, as having a complex subject term built up from a head noun as the *determinable* determined by a relative clause, then it should be obvious why Buridan gives the following answer to the problem raised in connection with this proposition:

> To the fourth question we should reply that there is one predicate here, namely, 'colored', which by the mediation of the copula is predicated of the whole of the rest as of its subject, namely, of the whole phrase: 'man who is white'; for the whole phrase: 'who is white' functions as a determination of the subject 'man'. And the case is not similar to 'A man is colored, who is white', for there are two separate predicates here, which are predicated separately of their two subjects, and there is not a predicate here which would be predicated by the mediation of one copula of the whole of the rest. And although these [propositions] are equivalent, they are not equivalent if we add a universal sign. For positing the case that every white man runs and there are many others who do not run, the proposition 'Every man who is white runs' is true, and is equivalent to: 'Every white man runs'; but the proposition 'Every man, who is white, runs' is false, for it is equivalent to: 'Every man runs and he is white'.[8]

Buridan's response to the objection in terms of distinguishing two interpretations of the relative clause indicated by different word order is particularly revealing of his practice of using a "regimented Latin" to make logical distinctions. Indeed, the difference between the syntactical devices used in English and Latin to make the same distinction is also very instructive concerning the advantages and disadvantages of developing logical theory in a "regimented" natural language, as opposed to doing the same using an artificial language, as we usually do nowadays.

Let us take a closer look at the syntax and the semantics of the propositions distinguished here, both in English and in Latin:

(1) Homo qui est albus est coloratus
(2) A man who is white is colored
(3) Homo est coloratus qui est albus
(4) A man, who is white, is colored

[8] SD 1.3.2.

(5) Omnis homo qui est albus currit ↔ (5') Omnis homo albus currit
(6) Every man who is white runs ↔ (6') Every white man runs
(7) Omnis homo currit qui est albus ↔ (7') Omnis homo currit et ille est albus
(8) Every man, who is white, runs ↔ (8') Every man runs, and he is white

Every other line here is the English translation of the Latin of the preceding line. Yet, the syntactical devices by which the Latin and the English sentences bring out the intended conceptual distinction are obviously different (word order vs. punctuation). Nevertheless, the important thing from our present point of view is that these different devices can (and do) bring out the *same* conceptual distinction.

As should be clear, the fundamental difference in all the contrasted cases is whether the relative clause is construed as a *restrictive relative clause*, forming part of the complex subject term, or it is construed as a *non-restrictive relative clause*, making a separate claim referring back to the simple subject of the main clause.

The "regimentation" of the syntax of a natural language, therefore, is the *explication*, and occasionally even the *stipulation*, of *which* syntactical structures of the given language are supposed to convey *which* conceptual constructions. The governing principle of Buridan's regimentation of his technical Latin seems to be what may be called the *principle of scope-based ordering*. This principle is most clearly at work in the "Polish notation" in modern formal logic (where the order of application of logical connectives is indicated by their left-to-right ordering), but something similar is quite clearly noticeable in Buridan's rules of logical syntax in general.

To be sure, Buridan never goes as far as to organize Latin according to the rules of a formal syntax in the way a modern artificial language is constructed.[9] And for all his views about the conventionality of language, even he would shy away from re-rewriting the rules of Latin grammar to fit the requirements of the syntax of an artificial language. Rather, he uses the existing grammatical, structural features of Latin (sometimes stretching, and sometimes bending them a little) to make conceptual distinctions. However, once such a distinction is somehow made, using some such existing syntactical device, Buridan's regimentation of Latin consists in his insistence on the point that this syntactical device should be consistently regarded as expressing this conceptual distinction, at least when we use language for the purposes of logic (as opposed to, for example, using it to do poetry).

[9] I tried to do this once for a *tiny* fragment of Latin with an explicitly listed finite vocabulary for the purposes of illustration, and even that resulted in an extremely complex, unwieldy system. See (Klima 1991).

REGIMENTATION VS. FORMALIZATION

Thus, even if doing logic by means of a full-fledged artificial, formal language did not even emerge as a theoretical alternative for Buridan, given the fact that in our time this is the dominant approach to logic, we should pause here a little to reflect on the theoretical as well as the practical advantages and disadvantages of these two different approaches.

One apparent disadvantage of Buridan's "regimentational" approach in comparison to the modern "formalizational" approach is that an informal system can never be as exact as a formal one, given all the possible ambiguities and arbitrariness of an informal approach. By contrast, in the formal approach, the rules of interpretation in a formal semantics and the manipulations with formulae in a formal syntax are fixed by the highest standards of mathematical exactitude, which can never be matched by any sort of informal approach. Therefore, it seems that Buridan's approach suffers from an inherent *inexactitude* that can be overcome only by the formalizational approach.

Again, Buridan's approach renders the construction of logical theory in a fundamental sense *unfinishable*. Given the immense variety and variability of the syntactical forms of a natural language, a logical theory based on its regimentation will never cover *all* syntactically possible constructions in a natural language. By contrast, since in an artificial language we have an explicit and effective set of construction rules, we can formulate logical laws that apply to all possible well-formed formulae of that language without having to worry about possible formulae that may not be covered by these laws.

Furthermore, Buridan's approach seems to be plagued by what may be termed its *linguistic provincialism*. If logical rules and distinctions are formulated in terms of the regimentation of the existing syntactical devices of a particular natural language, then, given the obvious syntactical diversity of natural languages, this approach seems to threaten the universality of logical theory. Indeed, following the lead of the syntax of a particular natural language may even provide "false clues" concerning what we may mistakenly believe to be the universal conceptual structure of Mentalese. By contrast, the formalizational approach provides equal access for speakers of all languages to *the same* "conceptual notation", which directly reflects the structure of the common mental language of all human beings endorsed by Buridan. So, apparently, even Buridan's logic would be much better off if it were also couched in an artificial, formal language.

Finally, if we use the natural language embodying our logic in our *reflections* on *the same* natural language, then we are obviously running the risk of Liar-type paradoxes, which are bound to emerge under the resulting conditions

of *semantic closure*, first diagnosed as such by Alfred Tarski.[10] By contrast, an artificial language embodying our logical theory can serve as the *object language* of the considerations concerning the syntax and semantics of this language which are to be carried out in a distinct *meta-language*. In this way we avoid the risk of paradox, since keeping the object language apart from the meta-language eliminates semantic closure.

Perhaps, these would be the most obvious reactions against Buridan's "regimentational" approach coming from someone comparing it to the modern "formalizational" approach. Nevertheless, these considerations may not be sufficient to establish once and for all the "absolute superiority" of the modern approach over Buridan's. For if we take a closer look at the modern practice, we can see that it is not much better off concerning these issues.

It must be conceded at the beginning that the mathematical exactitude of a formal logical system is unmatched by any "natural" logic (i.e. a logical system based on a certain regimentation of reasoning in some natural language). But the exactitude in question concerns only the formal system in and of itself. Concerning the formal system, we may have absolutely rigorous proofs of consistency or inconsistency, completeness or incompleteness, etc., which we may never have concerning an "unfinishable" system of "natural" logic. However, as soon as we use a formal logical system to represent and evaluate natural language reasoning, the uncertainties and ambiguities of interpretation return with a vengeance, as anyone who has ever tried to impart "formalization skills" in a symbolic logic class can testify. "Formalization" is the largely intuitive process of translating natural language sentences to formulae of a formal language based on the linguistic competence of the speakers of the natural language in question and their understanding of the import of the symbols of the formal language. Therefore, this process involves just as much inexactitude, uncertainty and ambiguity as does working with "unregimented" natural language expressions in general.

This difficulty can be overcome by constructing a formal syntax for an interesting portion of a natural language, in the vein of the approach of Richard Montague and Imre Ruzsa,[11] which can then serve as the basis for an exact and effective translation procedure into the artificial language of a formal logical system. In this way, having a formally constructed (not to say, "regimented") part of a natural language at our disposal, the problem of the inexactitude of

[10] Cf. (Tarski 1944). The gist of the idea of semantic closure is that a language that contains its own truth-predicates and has the means of referring to its own sentences is semantically closed, which is quite obviously the case with natural languages. According to Tarski, in a semantically closed language, Liar-type paradoxes ('This sentence is false' – is this true or false?) are bound to arise. For a more recent, generalized version of Tarski's argument, see (Priest 1984).

[11] Cf. (Montague 1973), (Montague 1974), (Ruzsa 1989).

the otherwise merely intuitive formalization process can certainly be avoided. However, given that the formal syntax can only cover a sufficiently interesting, yet manageable, part of a natural language, this approach brings out most clearly the inherently "unfinished" character of the enterprise as far as the representation of *all possible forms of natural language reasoning* is concerned. Thus, the formalizational approach can overcome the problems of inexactitude only by carving out a manageable part of natural language reasoning, thereby making explicit the "unfinished" character of the enterprise. Buridan's regimentational approach, in comparison, simply acknowledges from the start that it can only explicate and regulate certain manageable types of natural language reasoning, and it does this with the requisite degree of exactitude, yet without introducing the explicit, full-fledged formal syntax of an artificial language that would allegedly universally reflect the structure of Mentalese underlying all natural linguistic structures.

Since the process of formalization as it is commonly practiced is based on the linguistic competence of the speakers of particular natural languages, it involves just as much "linguistic provincialism" as does the regimentational approach. Actually, it is quite instructive to observe the differences between different Montague-style approaches to formalization motivated by different languages, especially if they are also motivated by certain logical considerations that are "most natural" in those languages.

But we can also say that the syntax of standard predicate logic as we know it was also motivated by some fairly "provincial" linguistic considerations, namely, considerations concerning the language of mathematics, rather than any actual natural language. This is probably the historical reason for the notorious "mismatch" between the syntax of predicate logic on one hand, and the syntax of various natural languages on the other, which otherwise agree among themselves in those of their syntactic features that predicate logic systematically fails to match. Consider again sentences (5)-(8) listed above:

(5) Omnis homo qui est albus currit ↔ (5') Omnis homo albus currit
(6) Every man who is white runs ↔ (6') Every white man runs
(7) Omnis homo currit qui est albus ↔ (7') Omnis homo currit et ille est albus
(8) Every man, who is white, runs ↔ (8') Every man runs, and he is white

In modern predicate logic, these sentences have to be represented in terms of the basic vocabulary of the formal language of this logic. In that language, besides the logical constants (which Buridan would recognize as syncategorematic terms, such as negation, conjunction, conditional, etc.), we have primitive symbols referring to individuals, namely, individual names (intuitively corresponding

to proper nouns) as well as variables (roughly corresponding to pronouns),[12] and predicates (corresponding to common terms). All complex expressions are built up from these primitive symbols by means of an explicit set of rules that effectively determine which strings of these symbols are to be regarded as well-formed. Frege's original rationale for this type of construction was that he regarded all common terms as functional expressions: on this conception, a common term, such as 'man', denotes a function from individuals to the two truth-values, the True and the False. Thus, the term itself is essentially predicative; it needs to be completed with a referring expression picking out an individual to yield a complete sentence that denotes one of these truth values. Therefore, since for Frege *all* common terms denote functions of this sort, *all common terms are essentially predicative*. Accordingly, in the sentences above, even their grammatical subject terms need to be construed as predicates of individuals, which are picked out by variables bound by the quantifier word 'every' or 'omnis'. It is for this reason that universal sentences in this logic are to be represented as universally quantified conditionals. Since the subject terms of these sentences are not regarded as having the function of restricting the range of individuals to be considered in determining whether the sentence is true, these sentences will have to be interpreted as concerning all individuals in the universe, stating of them all conditionally that *if* they fall under the subject, *then* they also fall under the predicate.

Thus, (6) and (8), and the corresponding Latin sentences as well, would on this approach be parsed as expressing the same as

(6'') For everything (it holds that) if it is a man and it is white, then it is running

(8'') For everything (it holds that) if it is a man, then it is white and it is running

In other words, using the variable x in place of the pronoun,

(6''') For every x, if x is a man and x is white, then x is running

(8''') For every x, if x is a man, then x is white and x is running

And these, using the symbols of predicate logic, directly yield the matching formulae:

(6'''') $(\forall x)[(Mx \,\&\, Wx) \supset Rx]$

(8'''') $(\forall x)[Mx \supset (Wx \,\&\, Rx)]$

[12] The problems of representing anaphoric pronouns with bound variables of quantification theory generated a whole new field of research in the eighties, primarily inspired by Peter Geach's reflections on "donkey-sentences", coming from medieval logic, and especially from Buridan. For a summary account of those developments and their comparison to Buridan's ideas, see (Klima 1988).

However, given Buridan's radically different conception of the semantic function of common terms, he would provide a very different parsing for (6) and (8) (or rather for (5) and (7)). For on his conception, common terms have the function of signifying several individuals indifferently (as opposed to singular terms that would signify one individual as distinct from any other), and correspondingly they *supposit*, i.e., stand for (some of) these individuals in the context of a proposition in which the term is actually used for this purpose. Therefore, on Buridan's reading, (6) and (8) (or rather (5) and (7)) do not make a conditional claim about all individuals in the universe, but rather a categorical claim about a restricted range of individuals, namely, those marked out by the subject term, i.e., the *suppositа* of the subject.

In fact, as anyone checking her own linguistic intuitions in English can testify, Buridan's analysis, coming from a "provincial" natural language, namely, Latin, matches much better the intuitions of speakers of another "provincial" natural language, namely, English. For English speakers would also find it "more natural" to understand the corresponding sentences as being categorical claims about a restricted range of individuals, rather than conditional claims about absolutely everything. To be sure, further reflection on the implications of this sort of analysis may further influence one's judgment on what "the correct" analysis of these sentences ought to be, but at least it should be clear that the Fregean analysis is definitely *not the only possible* or even the "most natural" one.

Thus, the Fregean analysis, being only one possible theoretical option, turns out to be just as provincial as Buridan's approach based on a particular natural language. Nevertheless, one may still object that at least for the Fregean analysis we have a working formal system with all the advantages of mathematical exactitude going for it, whereas we have nothing comparable for Buridan's approach. But this is simply not true.

BURIDAN'S LOGIC AS A LOGIC OF RESTRICTED QUANTIFICATION

As I have argued in several earlier papers, a simple, conservative extension of predicate logic can go a long way toward capturing in an exact form much of medieval logic in general and Buridan's logic in particular. Once we enhance the language of standard predicate logic with *restricted variables*, and provide the appropriate formal interpretation for their semantic evaluation in a formal semantic system, the resulting system at once becomes capable of capturing an enormous amount of traditional logic, and especially Buridan's version of it. We do not have to go into the technical details of constructing that formal system[13]

[13] For the technically-minded reader, a semantic system of this sort is available in (Klima 2001).

to explain its basic intuitive idea and its important philosophical implications concerning the relationships between this "enhanced predicate logic", classical predicate logic, and Buridan's informal logic.

The "basic intuitive idea" can be articulated in the following principles of construction:

(1) Restricted variables function as variables in classical predicate logic, i.e., they are quantifiable terms that fill in the argument places of predicate letters.

(2) Restricted variables have the general form of 'v.Av', where v is what is referred to as the operator variable of the restricted variable, and 'Av' as the matrix of the restricted variable, which is a well-formed formula open in v (i.e., having at least one occurrence of v that is not bound by a quantifier). The operator variable may itself be a restricted variable, in which case we can refer to it as a "nested" restricted variable (a restricted variable "nested" in another); other restricted variables occurring in the matrix of a restricted variable are spoken of as "embedded" in that restricted variable.

(3) Restricted variables pick their values in a value-assignment *not* from the entire domain of interpretation ("universe of discourse"), but from the extension of their matrix, i.e., from the set of individuals of which the matrix is true (under a certain value-assignment of variables).

(4) If the extension of the matrix of a restricted variable is empty, then the restricted variable has no value (which in the formal system can be represented by assigning an arbitrary value to it, outside the domain of interpretation, a so-called "zero-entity", a technical device I owe to Imre Ruzsa). When a restricted variable has no value (i.e., technically, its value is outside the domain of discourse), then its value cannot fall within the extension of any predicate, i.e., all simple affirmative predications containing this variable in the argument of a predicate letter will come out as false.

Having these "principles of construction" in place, we can obtain a system that (i) reflects more faithfully the syntax and semantics of natural languages than standard predicate logic,[14] (ii) naturally extends itself to a generalized quantification theory, (iii) it not only matches, but surpasses standard predicate logic in expressive power, and (iv) provides an analysis of categorical propositions perfectly in tune with Aristotelian logic, validating all relations of the traditional *Square of Opposition* and the traditionally valid syllogistic forms.

Let us now take these four points in turn, and see exactly how the system constructed in accordance with (1)-(4) can obtain these results.

(i) Predicate logic formulae using unrestricted quantification exhibit a compositional structure involving propositional connectives that are nowhere

[14] For a precise characterization of the notion "faithfulness" involved in this intuitive claim, see (Klima 1988).

to be found in the corresponding natural language sentences (be they English, Latin or even Hungarian, etc.). If we take a look, e.g., at (6'''')-(6) above, the structural mismatch is obvious. But the same sort of mismatch becomes even more striking if we change the quantifier from universal to particular (or "existential"), which requires that the main conditional be replaced with a conjunction in the resulting formula, whereas no such change is apparent in the syntax of the corresponding natural language sentence. Indeed, the variation of the natural language determiner does not require any change at all in the rest of the sentence, whereas changing the corresponding quantifier always requires a change in the propositional connectives of the formula following it, if a corresponding formula can be produced at all.

Therefore, there is no single propositional connective that could fill the place of the question mark in the following semi-formal schemata:

For every *x*	}	
For some *x*	}	
For the *x*	}	*Fx* ? *Gx*
For most *x*	}	
For five *x*	}	

so that we would get correct representations of the following sentence-schemata, which obviously exhibit a uniform structure (just as would the corresponding Latin, etc.):

(1) Every	}	
(2) Some	}	
(3) The	}	*F*('s) is/are *G*('s)
(4) Most	}	
(5) Five	}	

Among these schemata, (1) and (2) can be represented in predicate logic only with formulae involving different propositional connectives, (3) and (5) demand complex formulae to provide their correct truth-conditions (such as the Russellian formula: '$(\exists y)\{[Fy \ \& \ (\forall x)(Fx \rightarrow x = y)] \ \& \ Gy\}$' for (3)), and for (4) there is demonstrably no quantificational formula that would provide its correct truth-conditions.[15]

(ii) By contrast, in the system of predicate logic enhanced with restricted variables (as well as with the requisite set of quantifiers), the following formula schema provides an intuitive formalization of (1)-(5): '$(Qx.Fx)(Gx.)$'. This states that Q *x* that is an *F* is a *G*, or in the plural form, that Q *x*'s that are *F*'s are *G*'s, where Q stands for any of the appropriate determiners or "quantifier words"

[15] For the proof, if "most" is understood as "more than half of the", see (Barwise and Cooper 1981, 214-215.)

of English (and *mutatis mutandis* the same goes for any other natural language). This immediately establishes the claim that this system naturally extends itself to a generalized quantification theory.[16]

(iii) People who argue for the superiority of modern predicate logic over "traditional", Aristotelian logic often refer to (various versions of) De Morgan's famous example as proof that the Aristotelian analysis of categorical propositions, and correspondingly Aristotelian syllogistic, is incapable of handling reasoning involving relational terms. Intuitively, the following looks like a valid inference: 'Every man is an animal; therefore, every man's head is an animal's head'. However, there is no way of parsing this inference along traditional lines so it would fit into a valid Aristotelian syllogistic form.

Medieval logicians, taking their cue from Aristotle's *Prior Analytics*, treated such inferences under the heading *de syllogismis ex obliquis*, i.e., "on syllogisms involving oblique terms", which is to say, terms in cases other than the nominative case, such as the genitive "man's" in the conclusion of De Morgan's example.[17] To be sure, "standard syllogistic" treats the terms of a syllogism as unbreakable units (just as propositional logic treats atomic sentences as such units), although it allows complex terms as substituends of such units. Therefore, when the validity of an inference turns on the conceptual connections between parts of such complex terms, "standard syllogistic" is indeed inapplicable (just as uniform quantification theory, involving only monadic predicates, is unable to handle inferences with multiply quantified sentences.) So, to account for such inferences, Buridan and others distinguished between the *terms of the syllogism* and the *terms of the propositions*, where the *terms of the syllogism* (in particular, the middle term) can be parts of the *terms of the propositions*, and provided further syllogistic rules in terms of this distinction, referring to the intrinsic complexity of the terms of the propositions involved.

Correspondingly, the predicate logic with restricted variables inspired by Buridan provides a compositional semantics for formulae that represent the internal structure of propositions with complex terms. Therefore, this logic has no more difficulty in handling such inferences than standard predicate logic does. There are, however, some important and instructive differences between the two.

[16] For good surveys of the booming research on generalized quantifiers in the mid-eighties, see (Van Benthem and Ter Meulen 1985), and (van Benthem 1986). For a recent survey of later developments see (Westerståhl 2005).

[17] For Buridan's treatment, see SD 5. 8.

In standard predicate logic, the De Morgan-example can be reconstructed as follows:

$(\forall x)(Mx \rightarrow Ax)$
For every x, if x is a man, then x is an animal

$(\forall x)(\forall y)[(Mx \,\&\, Dxy) \rightarrow (Ax \,\&\, Dxy)]$[18]
For every x and every y, if x is a man and y is the head of x, then x is an animal and y is the head of x.

Using restricted variables, the same example can be reconstructed in the following way:

$(\forall x\, Mx)(\exists y\, Ay)(x. = y.)$[19]
Every (x that is a) man is (identical with) some (y that is an) animal

---

$(\forall x.(\exists y.My)(Dxy.))(\exists u.(\exists v.Av)(Duv.))(x. = u.)$
Every (x that is a) head of some (y that is a) man is (identical with) a (u that is a) head of some (v that is an) animal[20]

One important difference between these two reconstructions is that if we drop the parenthetical phrases in the semi-formal sentences that are simply transcribed into the formulae with restricted variables, then we get perfectly good English sentences, which cannot be done with the semi-formal sentences transcribed into the standard formulae. This quite clearly indicates the close match between the syntax of the natural language sentences and the formulae with restricted variables.

Another important difference is that while the standard formulae are true if there are no men or they have no heads, those with restricted variables in those circumstances would be false. Therefore, according to the formalization with restricted variables, the inference is not formally valid, unless there is a further premise to guarantee that if there are men, then there are men's heads. Actually, this is how it should be. After all, even if it is actually true, it is not a *logical truth* (i.e., a truth based on the meaning of logical connectives) that if there are men, then they have heads. Therefore, the formulation with restricted variables provides an even better analysis of the natural language sentences, in the sense that it better reflects our semantic intuitions as to what is and what is not implied by the sentences in question.

[18] I am providing here the "stronger", but "more intuitive" formalization of this sentence. Cf. (Merrill 1977).

[19] To simplify formulae with restricted variables, the matrix of a restricted variable may be omitted after its first occurrence.

[20] For a similar analysis with the same results, see (Orenstein 2000). For a detailed discussion of the neat syntactical match between restricted quantification and natural language sentences, see (Klima 1988).

Thus, we have to conclude that the "Buridan-inspired" predicate logic with restricted variables, besides covering more than standard predicate logic does as far as non-standard quantifiers are concerned, can handle what standard predicate logic can, indeed, while sticking more faithfully to the syntactic construction of natural languages and reflecting better our semantic intuitions concerning reasoning in natural languages.

(iv) What accounts for the difference between the judgments of the two different formalizations concerning the validity of De Morgan's example is their difference in attributing vs. denying existential import to universal affirmative propositions. The reason why De Morgan's example at first appears to be intuitively valid is that we tend to tacitly assume the non-logical truth that if there are men, then they have heads too. However, a formally valid inference has to yield truth from truth with any terms, which is actually not obvious with De Morgan's example. Consider the following, analogous example: 'Every man is an animal; therefore, every man's hat is an animal's hat'. Suppose there are men, but no man has a hat, which is certainly possible. In that case it is obviously true that every man is an animal, but is it true that every man's hat is an animal's hat? Or take the following, perhaps even more obvious example: 'Every horse is an animal; therefore, every horse's wing is an animal's wing'. Knowing that there are no winged horses, and hence no horse's wings, we would naturally tend to reject the conclusion. To be sure, one may still understand this conclusion conditionally, as saying that *if* something is a horse's wing, *then* it is an animal's wing, but that conditional reading would lose precisely the matter-of-fact character of the original categorical claim.

Indeed, other examples can bolster our intuition that even if universal affirmatives may occasionally have the force of a conditional, hypothetical claim, especially when they are supposed to express a law-like statement; nevertheless, it is simply wrong to assume that they *always* have to be interpreted this way. Consider for example the case of Mary boasting to her friends that every boy kissed her at the party yesterday. If her friends later find out that there were no boys at the party, then they will certainly take her for a liar, rather than accept her claim as being "vacuously" true on account of her universal claim expressing a universally quantified conditional with a false antecedent. Such and similar examples could be multiplied *ad nauseam*. What is important, though, is the fact that we do have the intuitive distinction between the categorical and hypothetical readings of universal affirmatives; therefore, a logic that can acknowledge both of these readings is certainly preferable to one that can only handle one of them. Since predicate logic with restricted variables is a conservative extension of standard predicate logic in the sense that all formulae of the standard logic are formulae of the logic enhanced with restricted variables, the latter is of course capable of representing whatever the former can, but not *vice versa*.

CONCLUSION: BURIDAN'S "NATURAL LOGIC" VS. ITS RECONSTRUCTION

Nevertheless, although this has to be the end of this lecture, this is far from being the end of the story of comparing medieval and modern logic. The foregoing could serve merely to illustrate that heeding medieval logicians' *regimentation* of *natural language*, we may be able to come up with some *more natural formalization* in a *formal language*. But, as I argue in detail in my monograph on John Buridan (Klima 2009), where this lecture comes from, Buridan would still not be quite happy with this formal reconstruction of his logic. And the reason would not be its formalism (after all, Buridan also uses some symbolism time and again), but rather its restricted applicability in other areas, where Buridan's logic still has important lessons to teach us. In particular, even if quantification theory with restricted variables can easily be extended to cover a great deal of Buridan's modal and temporal logic, it cannot quite properly handle Buridan's treatment of intentional contexts generated by words signifying our mental acts. Moreover, Buridan would not be quite happy with restricted variables representing his common terms, since for him there are *simple* common terms, say, *F*, the semantic properties of which are different from a *complex term*, such as 'an x that is an *F*'. In fact, in various contexts, Buridan would sharply distinguish between the logical import of the two. Finally, and even more importantly, Buridan would reject both the Quinean idea of ontological commitment usually associated with quantification theory and the global distinction between object language and meta-language, built into the very construction of this theory. As I argue in my book, this double rejection allows Buridan to work out a third alternative "between" a Quinean and a Meinongian approach to ontological commitment, as well as a viable logical theory for semantically closed natural languages, avoiding Liar-type paradoxes. But this much may be just enough by way of "a shameless plug" to finish this lecture.

REFERENCES

Aristotle, *On Interpretation*

SD = Johannes Buridanus, *Summulae de Dialetica*. Translation used: Klima, G., John Buridan: *Summulae de Dialectica*, an annotated translation with a philosophical introduction; New Haven, Yale University Press, 2001.

Barwise, J. and Cooper, R., 1981, Generalized Quantifiers and Natural Language. *Linguistics and Philosophy 4*(1981). 159—219.

Klima, G., 1988, Essay III.: General Terms in their Referring Funcion. In id., *Ars Artium: Essays in Philosophical Semantics, Medieval and Modern*, Budapest, Institute of Philosophy of the Hungarian Academy of Sciences. 44—84.

Klima, G., 1991, Latin as a Formal Language: Outlines of a Buridanian Semantics. *Cahiers de l'Institut du Moyen-Âge Grec et Latin 61*. 78—106.

Klima, G., 2001, Existence and Reference in Medieval Logic. In A. Hieke and E. Morscher (eds.), *New Essays in Free Logic*. Dordrecht, Kluwer Academic Publishers. 197—226.

Klima, G., 2009, *John Buridan*. Oxford, Oxford University Press.

Merrill, D., 1977, On De Morgan's Argument. *Notre Dame Journal of Formal Logic 18*. 133—139.

Montague, R., 1973, The Proper Treatment of Quantification in Ordinary English. In J. Hintikka, J. Moravcsik and P. Suppes (eds.), *Approaches to Natural Language*. Dordrecht, Reidel. 247—270.

Montague, R., 1974, English as a Formal Language. In R. Thomason (ed.), *Formal Philosophy*. New Haven-London, Yale University Press.

Orenstein, A., 2000, The Logical Form of Categorical Sentences, *Australasian Journal of Philosophy 78*, 517—533.

Priest, G., 1984, Semantic Closure. *Studia Logica 43*. 117—129.

Ruzsa, Imre 1989, *Logical Syntax and Semantics* vol. II. (in Hungarian) Budapest, Akadémiai Kiadó.

Tarski, A., 1944, The Semantic Conception of Truth. *Philosophy and Phenomenological Research 4*. 342—375.

van Benthem, Johan and Alice ter Meulen (eds.), 1985, *Generalized Quantifiers in Natural Language*. Dordrecht: Foris Publications.

van Benthem, Johan, 1986, *Essays in Logical Semantics*. Dordrecht, Reidel.

Westerståhl, D., 2005, Generalized Quantifiers. In *The Stanford Encyclopedia of Philosophy* (Winter 2005 Edition), ed. E. N. Zalta, URL = http://plato.stanford.edu/archives/win2005/entries/generalized-quantifiers/.

EDWARD KANTERIAN

Frege's Definition of Number: No Ontological Agenda?

Abstract. Joan Weiner (2007) has argued that Frege's definitions of numbers constitute linguistic stipulations that carry no ontological commitment: they don't present numbers as pre-existing objects. This paper offers a critical discussion of this view, showing that it is vitiated by serious exegetical errors and that it saddles Frege's project with insuperable substantive difficulties. It is first demonstrated that Weiner misrepresents the Fregean notions of so-called *Foundations*-content, and of sense, reference, and truth. The discussion then focuses on the role of definitions in Frege's work, demonstrating that they cannot be understood as mere linguistic stipulations, since they have an ontological aim. The paper concludes with stressing both the epistemological and the ontological aspects of Frege's project, and their crucial interdependence.

1 THE PROBLEM

It is indisputable that Frege's logicist project, including the development of his logical calculus, had an epistemological aim, namely to prove the *a priori* and analytic status of the arithmetical truths, and thus to prove that they are deducible from the laws of logic. More problematic, and a subject of recent debate, is the question concerning the status of definitions within this project.[1] Frege dismisses previous attempts at the definition of number, and replaces them with new definitions. In addition, in several passages he describes definitions as arbitrary conventions. So it seems as if Frege is not interested in capturing with his definitions the pre-existing meanings of arithmetical symbols, but in stipulating new ones. But how can this revisionary project be brought into harmony with the epistemological aim, i.e. how can arbitrary

[1] Some key contributions to this debate are (Benacerraf 1981), (Weiner 1984), (Picardi 1988), (Kemp 1996). For more details, see the recent overview in Shieh (2008).

definitions contribute to proving the logical status of *the* truths of arithmetic, i.e. of antecedently existing truths?

In the most recent contribution to this debate Joan Weiner (2007) offers a radically new solution: Frege was a more thorough revisionist than the dilemma above presents him. His revisionism affected not only his conception of definition, but also of sense, reference and truth. Prior to his work, numerals did not have a determinate sense and reference, and arithmetical statements were not strictly speaking true. According to Weiner, Frege did not believe that concept-script systematisation is unveiling the true nature of numbers and the true referents of numerals, but only that it introduces stricter semantic and inferential constraints of precision stipulating the sense and reference of numerals and arithmetical statements for the first time. Thus talk about numbers as objects and strict arithmetical truth is only possible as a system-internal discourse, and concept-script systematisation is a normative linguistic precisification serving an epistemological aim, with no ontological and semantical discoveries about pre-systematic arithmetic and its language. In particular, definitions carry no content-preserving and ontological commitment.

2 WEINER'S ARGUMENT

Weiner offers a wealth of substantive and exegetical considerations in favour of her view, focusing most explicitly on the role of definitions within Frege's work, especially in the *Foundations of Arithmetic*. She investigates what requirements a definition (of number, numerals etc.) must satisfy in order to qualify as adequate or faithful to prove the truths of arithmetic from primitive truths (Weiner 2007, 683). One such obvious requirement seems to be the following:

> *The obvious requirement*: A definition of an expression must pick out the object to which the expression already refers or applies (ibid. 680).

Weiner denies there is any evidence in Frege's writings for this requirement. Definitions are not preserving the putative pre-systematic reference of numerals. Still, they must be faithful to pre-systematic arithmetic in some sense, since systematisation is not meant to transform arithmetic into some 'new and foreign science' (ibid. 687). Her explanation is as follows: 'Faithful definitions must be definitions on which those sentences that we take to express truths of arithmetic come out true and on which those series of sentences that we take to express correct inferences turn out to be enthymematic versions of gapless proofs in the logical system' (ibid. 690, 790). In other words, what systematisation preserves is truth-related and inference-related content. For example, regarding truth-related content a definition of '0' and '1' is unacceptable, if it presents as true a sentence

which in pre-systematic arithmetic is taken to express a falsehood, namely '0=1'. Thus faithful definitions must cover for what are taken to be the well known properties of numbers.[2] Regarding inference-related content, faithful definitions must preserve inferences that we take to be valid, for example 'If Venus has zero moons and the Earth one, then given that 0<1, the Earth has more moons than Venus' (ibid. 686). Thus faithful definitions must cover for all applications of number, including those in applied arithmetic.[3]

Frege is therefore not concerned with preservation of reference, and not even simply with preservation of putative truths, rather of what Weiner calls '*Foundations*-content'. This is 'some sort of content connected with inferences' (ibid. 692). *Foundations*-content partly points back to the judgeable content of *Begriffsschrift*, which was defined by Frege as content that has only 'significance for the inferential sequence' (1879, x.). But *Foundations*-content also partly anticipates the later notion of sense, i.e. *Sinn* (Weiner 2007, 689-1), for two reasons. First, the judgeable content of a term, she claims, is not its referent (ibid. 690, fn. 17), just as much as sense is not reference. Second, a term hitherto considered non-empty will not cease to have *Foundations*-content if we discover it is empty, for the discovery will lead to a re-evaluation of our pre-systematic beliefs and inferences, a re-evaluation still involving the term itself (ibid. 690). Equally, a fictional term like 'Hamlet' has *Foundations*-content, since there are speakers who think it enables them to express truths and correct inferences (ibid. 691). Hence, it is not required for a term to have a referent in order to have *Foundations*-content, and this brings *Foundations*-content in the vicinity of *Sinn*.

Thus, what concept-script systematisation achieves is preservation of *Foundations*-content. However, this should not be understood in the trivial sense of 'preservation', as if something outside of the system is identical with something in the system. As it transpires from Weiner's argument, preservation of *Foundations*-content means rather something like 'normative transformation of pre-system content into systematic content'. As quoted above, systematisation involves the process of proving *within the logical system* the truth of the pre-systematic sentences taken to express truths as well as the correctness of pre-system inferences taken to be correct. But proving 'in the logical system' is a highly normative process, guided essentially by *two precision requirements* that distinguish sharply the system from the pre-system: the gapless proof requirement, i.e. all proofs are absolutely gapless, and the sharpness requirement, i.e. genuine concepts must have sharp boundaries.[4] Essentially, this 'preservation' is to be understood as a creative process of precisification of pre-systematic language,

[2] (Weiner 2007, 688.) Cf. (Frege 1879, §70).
[3] (Weiner 2007, 689.) Cf. (Frege 1879, §19).
[4] See (Weiner 2007, 701), (Frege 1884, §62, §74), (Frege 1903, §56).

which is indeterminate and vacillating (ibid. 697), as himself Frege seems to suggest about expressions quite generally (see (Frege 1906a, 302–3)). 'Frege's task is to replace imprecise pre-systematic sentences with precise systematic sentences [of arithmetic]' (Weiner 2007, 710). This suffices to make sense of Frege's epistemological aim, the core of his logicist project.

This view has some intriguing implications and corollaries, the most important of which shall be briefly summarised. More details evidence will be presented during the discussion further below.

(a) *Foundations*-content is close to sense, but not identical with it. To have a determinate sense, an expression must have a definition satisfying the precision requirements. But pre-systematic expressions don't have such a definition, hence they don't have a sense and, by extension, no reference.[5] Sense and reference (*Sinn* and *Bedeutung*) are therefore only system-internal features of expressions.[6] There is no evidence to the contrary in Frege's writings. In particular, Frege never says that terms of pre-systematic language have a reference or require one in order to have a use (ibid. 706f.).[7] The absurdity of this view is merely apparent, for fixing the sense and reference of a term is only the ideal end of a science, once it comes to fruition in a system (ibid. 709f.)[8].

(b) If pre-systematic terms don't have a determinate reference, then given compositionality, pre-systematic sentences don't have a determinate reference either, i.e. a truth-value. We only 'take them to express truths' (ibid. 690f.). This does not mean there is nothing 'right' about them (ibid. 710), but only that their rightness does not satisfy the constraints imposed by systematisation. We must distinguish between different notions of truth, as Frege does, i.e. *pre-systematic truth* and *strict truth* (ibid. 709f.).[9] Pre-systematic truth is one of the aforementioned faithfulness requirements a definition (of number etc.) must satisfy.

(c) Since pre-systematic arithmetical expressions do not have a determinate reference, ordinary arithmetical predicates like 'is a number' do not have a reference either. Hence, the concept of number is not already fixed prior to Frege's definitions (ibid. 696). Quite the opposite: in *Foundations* (§100) Frege stresses the arbitrary, stipulative character of definitions (ibid. 695ff.), and he does so again in the important posthumous

[5] The alternative of having an indeterminate sense (and reference) is excluded for Frege, since there is no such thing for him. See for instance (1903, §56).

[6] This claim has been advanced before. See e.g. (Stekeler-Weithofer 1986, 8.,10.).

[7] Weiner's argument seems to come close here to a Wittgensteinian theory of meaning as use, although Wittgenstein is not mentioned.

[8] See also (Frege 1914, 242).

[9] 'Pre-systematic truth' is not Weiner's term, but my terminological correlate to her label 'regarding a sentence as true' (ibid. 706, 709).

text "Logic in Mathematics" (Frege 1914). Here he claims that a determination of sense is either decompositional, in which case it is a self-evident axiom, or it is a mere stipulation ('constructive definition'). Since Frege does not seem to present his *Foundations* definitions as self-evident (1884, §69), they must be stipulations, stipulations which precisify *Foundations*-content and thus transform arithmetic into a system of science.

(d) A cursory reading of Frege's writings might induce one to assume that he thinks numbers are pre-existing, language-independent objects whose nature his definitions aim to capture. Call this *the ontological thesis*. Weiner rejects this thesis.[10] On her view Frege makes no claim that numbers existed prior to his definitions, or else he would have to say that the definitions are (or articulate) discoveries about pre-existing objects. But they are linguistic stipulations, not ontological discoveries. Frege does not claim that it is part of the nature of numbers to be extensions, but is interested only in the linguistic question 'Are the assertions we make about extensions assertions we can make about numbers?', which he answers by means of a linguistic principle par excellence, the context principle (Weiner 2007, 698f.). As Frege writes: 'I attach no decisive importance to bringing the extensions of concepts into the matter at all' (1884, §107).

Weiner's interpretation is certainly intriguing and original. Nevertheless, it is vitiated by serious exegetical errors, and it saddles Frege's theory of numbers with insuperable substantive difficulties. I will first show that Weiner misrepresents so-called *Foundations*-content, sense and reference, and the notion of truth in Frege's work (sections 3-5). Then I will focus on the role of Fregean definitions, demonstrating that they have, *pace* Weiner, an ontological point, and that they are not mere stipulations. The paper concludes with stressing both the epistemological and the ontological aspects of Frege's project, and their crucial interdependence.

3 '*FOUNDATIONS*-CONTENT'?

We can start with the notion of *Foundations*-content, on which Weiner bases her rejection of the obvious requirement. Is there really a notion of content in the *Foundations* closely related, although not identical to sense? There is no decisive evidence. Frege uses the term loosely. It may mean various things such

[10] Weiner has defended this anti-ontological stance in previous work. See for instance (Weiner 1990, ch. 5.)

as 'sense of a sentence', i.e. a judgement or thought (1884, x., §3, §70, §106), 'sense of a recognition judgement' (ibid. §106, 109), 'judgeable content'(ibid. §62, §74), 'reference' (ibid. §74fn.), and also just conventional meaning and use (ibid. §60). 'Content' in the *Foundations* is simply not a technical term, which fits the prolegomenous character of the book. There is nowhere an argument specifying that an expression could have a content but lack a referent, i.e. that 'Hamlet' has *Foundations*-content. In one place Frege claims the exact opposite in fact: 'the largest proper fraction' has no content and is senseless, because no object falls under it (ibid. §74fn.). He makes a similar point about 'the square root of -1' (ibid. §97). Moreover, the predominant and philosophically significant role of 'content' in the *Foundations* is found in Frege's repeated requests to specify a content of arithmetical judgements in such a way that they turn out to express identities, and thus to secure the objecthood of numbers, given his acceptance of the principle that identity is an essential mark of objecthood (1884, §62f.).[11] Thus we have positive evidence that content in the *Foundations* is closer to reference than to sense.

At the very least, is Foundations-content not related to Begriffsschrift content and insofar only inferential, not referential? But there is no dichotomy here. Judgeable content is so intimately tied to reference that it affects the most basic formation rules of the notation in *Begriffsschrift*. Frege stipulates that any expression following the content stroke must have a judgeable content. That the relation between a judgeable content and its expression is one of being designated, is visible from at least two facts: that the expression of a judgeable content is a designator starting with the definite article, paradigmatically the nominalised form of a proposition ('the circumstance that there are houses', 'the violent death of Archimedes at the capture of Syracuse')[12], and that the expression of a judgeable content can flank the sameness of content sign, i.e. the identity sign. Hence, the expression of a judgeable content is a name of the judgeable content and no formula in concept-script is even syntactic if such a name fails to designate anything.[13]

Weiner argues that since 'Phosphorus = Phosphorus' is derivable from the law of identity, while 'Hesperus = Phosphorus' is not, the two names cannot have the same *Begriffsschrift*-content (Weiner 2007, 690). But this example is uncongenial to Frege's concerns: 'Phosphorus' and 'Hesperus' are names of *un*judgeable content, while in *Begriffsschrift* he is interested only in names of judgeable content—content that can be asserted. Hence, unlike with names of unjudgeable content, there is no room for distinguishing between what a name

[11] See also (Rumfitt 2003, 198.)

[12] (Frege 1879, §2-3.)

[13] Frege refers to the expression of a judgeable content flanking the identity sign explicitly as a name (Frege 1879, §8, passim).

of a judgeable content refers to and its inferential content. Therefore, once it has been established (by a synthetic judgement) that two names name the same judgeable content, 'A' and 'B' are intersubstitutable for the purposes of inference ('≡' functions as an identity, but also as a stipulative sign; see (Frege 1879, §8, §23)), just as much as, vacuously, 'A' and 'A'. The difference between two names of the same judgeable content is neither a referential nor an inferential difference, but only of a different mode of determination of the same inferential content, as Frege explicitly states (Frege 1879, §8). Claiming that *Foundations*-content is somehow connected to *Begriffsschrift* content, while arbitrarily revising Frege's own view of the latter, is question-begging.

4 SENSE AND REFERENCE

It is equally unwarranted to treat sense and reference (*Sinn* and *Bedeutung*) as system-internal features which do not apply to ordinary expressions. There are enough passages to the contrary, for example when Frege writes that 'The moon is the reference [*Bedeutung*] of "the moon"' (1919, 255) or when he explains, in a lecture, that '[i]f we claim that the sentence "Aetna is higher than Vesuvius" is true, then the two proper names do not just have a sense [*Sinn*] […], but also a reference [*Bedeutung*]: the real, external things that are designated'[14]. In "Introduction to Logic" he writes: 'If we say "Jupiter is larger than Mars", what are we talking about? About the heavenly bodies themselves, the references [*Bedeutungen*] of the proper names "Jupiter" and "Mars"' (1906b, 193). In *Grundgesetze* he points out that the notion of reference (*Bedeutung*) cannot be (genuinely) explained, since any such explanation would presuppose knowledge that some terms have a reference (Frege 1893, §30); hence reference must predate the setup of any system. Since to have a reference implies having a sense, it follows that those expressions of natural language that have reference also have a sense. Neither feature is therefore exclusively system-internal for Frege. Hence, if it is insisted that Frege's definitional project is to be described as preserving so-called *Foundations*-content, then this will involve preservation of pre-systematic content that has a referential component, and this will not be content of a wholly different kind from the content described in *Grundgesetze* split into sense and reference (see (Frege 1893, x.)).

[14] Carnap 2004. 150.

5 KINDS OF TRUTH?

Weiner claims that Frege allows for many notions of truth, including pre-systematic and strict truth. But nowhere does Frege make such a claim. He argues, in "Thoughts", that truth is undefinable, on the basis that any attempted definition would have to analyse truth into constituent properties (of the truth bearer), of which in turn it would have to be *true* that they apply in a particular case in order to make the definition applicable (1918a, 60). This circularity suggests not only the indefinability of truth, but also its simplicity, and thus its univocity. Frege intimates in "Thoughts" additional arguments—very plausible ones—against Weiner's claim. Thus he distinguishes sharply between 'taking something to be true' ('*Fürwahrhalten*') and 'proving the true' ('*Beweis des Wahren*'). The former arises through psychological laws, while the latter belongs to the laws of truth, which are not the object of psychology, but only of logic (ibid. 58f.). Since Weiner characterises pre-systematic truth in terms strongly resembling *Fürwahrhalten*, e.g. 'what we take to express truth' or 'what we regard as true', it would follow, absurdly, that Frege takes truths of pre-systematic arithmetic to belong to the realm of the psychological, and pre-systematic arithmetic to psychology. Equally, if pre-systematic truth is vacillating and vague, it would seem to be a predicate coming in degrees. But Frege specifies that truth does not allow for 'more or less' (ibid. 60). Finally, Frege speaks on repeated occasions about the truths of arithmetic or mathematics as such (e.g. (1884, §3, §11, §14, §17, §109)). But he nowhere qualifies them as merely pre-systematic, to be distinguished from the truths arrived at within the system.

Quite generally, the truth of a thought is timeless (e.g. (1884, §77, 1918a. 74)). Hence, either a pre-systematic notion of truth is timeless as well, in which case it is not clear how this squares with its indeterminacy and vagueness, or a pre-systematic assertion does not express a thought at all. Of course, Weiner can retort that this is indeed so: a pre-systematic assertion only expresses *Foundations*-content. But then *Foundations*-content is assertable. And it is negatable, thinkable, judgeable etc. However, *Foundations*-content is not really judgeable, since judging is 'acknowledging the truth of a thought' (ibid. 62). Judgeable is only that to which strict truth can apply. But then *Foundations*-content is not assertable either, for to assert is to make manifest a judgement (ibid.). And if it is not judgeable, *Foundations*-content lacks the essential association with judgeable content Weiner claims it has, and thus it does not have an inferential character either, for to infer is to judge (1879-1891, 3). Pre-systematic arithmetical proofs and arguments could not be counted as valid, if we continue this line of thought. In conclusion, the distinction between

pre-systematic and strict truth, in conjunction with the notion of *Foundations*-content, leads to catastrophic consequences.[15]

6 THE OBJECTIVITY OF DEFINITIONS: THE MATERIAL MODE

Doubts about Weiner's interpretation also arise if we look more carefully at what Frege says about and what he does with his definitions in the *Foundations*. There are passages in the *Foundations* whose wording stress the stipulative character of definition, e.g. when Frege speaks about fixing ('*festsetzen*') the sense or meaning of expressions, or refers to definitions as stipulations ('*Festsetzungen*').[16] But his phraseology indicates an objective side to definition as well. Thus he speaks of the need to ascertain, find out ('*feststellen*'), explain the sense of an equation (ibid. x., p. 73, §62, §106), which is inaccurately translated as 'to fix/ to define the sense' by Austin, or of the need to find or attend to ('*aufsuchen*') a judgeable content which can be transformed into an identity whose sides contain the new numbers (1884, §104).[17]

The objectivity of definitions is also manifest in Frege's tendency to adopt the material mode and define objects, as opposed to mere expressions. This is confirmed by many passages in the *Foundations* (1879, §7, §8, §9, §10, §18, §67). In two places he speaks explicitly of 'the definition of an object' (1879, §67, §74). A case in point is the very passage in §100 which Weiner takes as decisive evidence in favour of her thesis that Fregean definitions are creative linguistic stipulations. She claims that Frege is telling us that the meaning of 'the square root of -1' is not fixed prior to our definitions, but only fixed for the first time by the definitions (Weiner 2007, 695). However, let us look at the full context:

> We should be equally entitled to choose as further square roots of -1 a certain quantum of electricity, a certain surface area, and so on; but then we should naturally have to use *different symbols* to signify these *different roots*. That we are able, *apparently*, to create in this way as many square roots of -1 as we please, is not so astonishing when we reflect that the meaning of the square root of -1 is not something which was already unalterably fixed before we made these choices, but is decided for the first time by and along with them' (1884, §100; my italics).

[15] One additional problem: if we accept the distinction, what are we to do with Weiner's own arguments, which are formulated in pre-systematic philosophical prose, and not in concept-script? Are they not valid either? Are her conclusions not strictly true? Both a 'yes' and a 'no' answer invalidate her theory.

[16] E.g. §7, §65, §67, §68, §75, §104, §109.

[17] See also §106, where Frege reports that he has established that numbers are not collections of things or properties.

Clearly, Frege is discussing here the possible definition of an object. This is why he talks about different *symbols* having to be assigned to each square root of -1, *once* each square root of -1 has been chosen. So the definitional choice is not a linguistic one. In addition, Frege's tone is verging on sarcasm here, as it is in the embedding discussion.[18] He is certainly not telling us that a definition entitles us, in virtue of its creative powers, to introduce as many square roots of -1 as we want, rather he is presenting his opponent's point of view ('apparently').[19] The latter is a sort of reformed formalist, who has accepted Frege's anti-formalist arguments (up to §100), according to which the introduction of signs alone will not bring complex numbers into existence. Therefore, the reformed formalist sets out to supplement his definition of a complex number with the assignment of a random object, to 'fix' the 'meaning' of the complex number. Frege is spelling out the absurd consequences of this approach, viz. the possibility of a plurality of such assignments.[20] Weiner's misunderstanding of this passage is twofold: Frege is not propounding *his* own view of definition of a linguistic *symbol* as creative stipulation, but his imaginary *opponent's* view of definition of an *object* as a random ontic assignment. There is no evidence in *Foundations* that Frege takes definitions to be arbitrary, creative definitions.

Weiner falls prey to a similar misunderstanding when she discusses an eligibility condition of primitive truths. In "On Formal Theories of Arithmetic" (1885), Frege argues that every definition must come to an end, hitting upon indefinable primitives, the original building blocks of science (*Urbausteine*), which are expressed in axioms (1885, 96). Weiner takes this to be a semantic point: the eligibility condition 'is that the *expression* of the primitive truth should include only simple, undefinable *expressions*. For these simples are the ultimate building blocks of the discipline' (Weiner 2007, 682, my italics). But this is not Frege's point. Again, Frege speaks here in the material, not the formal mode, concerned with defining the objects themselves, in this case the objects of geometry: 'It will not be possible *to define an angle* without presupposing knowledge about the straight line. Of course, what a definition is based on might itself have been defined previously' (1885, 104, my italics). The primitives terminating such a chain of definitions are not expressions, but undefinable objects 'whose

[18] In the same context he remarks: 'Let the Moon multiplied by itself be -1. This gives us a square root of -1 in the shape of the Moon' (1884, §100). This was hardly written with a straight face, given Frege's usual predilection for sarcasm and irony. See also his related attack against Kossack (1884, §103): 'We are given no answer at all to the question, what does 1 + i really mean? Is it the idea of an apple and a pear, or the idea of toothache and gout? Not both at once, at any rate, because then 1 + i would not be always identical with 1 + i.'

[19] Dummett is another author who misunderstands this passage. See (Dummett 1991, 178f).

[20] And he shows a few pages on that the formalist origin of this kind of reasoning leads to misconstruing the subject matter of arithmetic as synthetic and even as synthetic a posteriori (1884, §103).

properties are expressed in the axioms' (ibid.).[21] It is in the material mode that Frege goes on to explain that the building blocks of arithmetic must be of purely logical nature (or else we cannot account for the universality of arithmetic), and to describe the terminological replacement of 'set' with 'concept' as not being a mere renaming, but of importance for the actual state of affairs (1885, 104f.).

7 THE OBJECTIVITY OF DEFINITIONS: FREGE'S PLATONIC REALISM

Another challenge to Weiner's interpretation is Frege's Platonism, which he maintains with respect to arithmetical truth, arithmetical objects and logical objects, and which is manifest in the importance existence proofs play for his definitions. Frege explains the definition of 1 by reference to the sempiternality and apriority of the truth of the propositions it helps to derive, e.g. '1 is the immediate successor of 0'; no physical occurrence, including 'subjective' ones concerning the constitution of our brains, could ever affect the truth of this theorem (1884, §77). A mere linguistic constraint of precisification cannot explain this sempiternality; the truth-value of 'Tom is alive' will continue to depend on contingent, empirical facts even if we cut the boundaries of the embedded expressions sharp.[22] What explains the sempiternality is the specific objectivity of the content of arithmetical propositions, the nature of arithmetical objects. This nature is grounded in the nature of logical objects, given that number-statements are ultimately about relations between logical objects (correlations of (extensions of) second-order concepts). Far from explaining logical objects as the result of linguistic constraints and stipulations, he ascribes to them ontological objectivity: they are simply there, ready to be discovered by us. Thus judgeable contents, the paradigmatic logical objects of the early work, are for Frege as objective as any mind- and language-independent object, like the sun (1879-1891, 7), although not physical.[23] Definitions could not play any creative role at this ontic level, and they certainly don't play it in concept-script, where a definition is merely an abbreviation: it stipulates that a simple sign is to have the same judgeable content as a more complex one (1879, §24). The existence

[21] We can call the expressions designating such primitives also 'primitives', but only metonymically.

[22] Frege's eternalist theory truth, maintained elsewhere, does not matter here, since that theory is not concerned with the justification of the truth-value of a proposition, but with the question about the proper bearer of truth.

[23] Elsewhere he says something similar about the relation between an object and its concept: 'To bring an object under a concept is merely to recognise a relation that already existed beforehand' (1984. 198). Also: 'Our relation to logical truths and mathematical structures is inessential to their nature and existence' (1984. 371). The beautiful passage in the preface to *Grundgesetze*, according to which the laws of logic 'are boundary stones set in eternal foundations, which our thought can overflow, but never displace' (1893, xvi.), is also relevant here.

of judgeable contents and of names of judgeable contents thus precedes any definition in concept-script.

Frege claims the same kind of objectivity for numbers in the *Foundations*, i.e. independence of the mind, language, stipulations. Corresponding passages are legion (e.g. §26, §60, 62), and it is hard to see how mere precision requirements imposed by system construction could make sense of them. Take the remarkable passage in which he compares the objectivity of number with that of the North Sea, pointing out the independence of both from our arbitrary stipulations (1884, §26).[24] If we were to slightly change the meaning of 'North Sea' today, whatever true content (thought) has been expressed until now by 'The North Sea is 10,000 miles in extent' would not become false. The North Sea is out there, objective, ready to be discovered by us. Presumably, then, there are correct and incorrect definitions of the North Sea, if the North Sea is independent of the definition of 'the North Sea': a correct definition will pick out precisely the North Sea. *Pace* Weiner, what this suggests is that there is an element of discovery in at least a subclass of definitions: the correct ones pick out a pre-existing object in a precise and determinate manner. This is entirely compatible with what Frege says elsewhere about numbers, namely that we discover them in the concepts (1884, §48, §58), and about the similarity between the mathematician and the geographer: neither can create 'things at will; [both] can only discover what is there and give it a name' (1884, §96).[25] It is unclear how Weiner's interpretation can cope with such passages. They are not mentioned in her article.

As is visible from his rejection of empiricism in logic and mathematics, the objectivity Frege claims for logical and arithmetical entities is actually stronger than that of physical entities. Numbers are abstract objects, ready to be recognised by us, but without physical properties, including spatiality and temporality. Our recognition of abstracta is itself situated in time (and presumably space), but abstracta are not, as he explains using the example of the equator. The equator has not been created in the sense that nothing positive could be said about it prior to its alleged creation (1884, §26). This example is not wholly fortunate, since the equator is a dependent abstractum (prior to the creation of the Earth there was indeed nothing positive to say about the equator), but Frege's point is on firm grounds with respect to self-subsistent abstracta like numbers: there is something positive to say about them at all times (with the appropriate linguistic usage in place). This renders the temporalisation of the truth of statements concerning numbers (e.g. '"Numbers are extensions" is not

[24] Note, and ignore, that Frege's discussion of the arbitrariness of stipulations is confusingly embedded in a discussion against psychologism.

[25] Frege also compares the mathematician with the botanist who determines something objective when he determines the number and colour of a plant's petals. See also (1893, xiii.)

a true statement prior to Frege's definitions formulated in 1879'), as entailed by Weiner's argument, both objectionable and unFregean.

Frege's drive towards a Platonist ontology is also manifest in another respect. In the *Foundations* Frege does not merely provide us with a definition of, say, 0, and content himself with the fact that it satisfies the sharpness requirement. Instead, he gives us various *existence proofs*, e.g. that if 0 is the number of an empty concept *F* and an empty concept *G*, then there must be a relation φ bringing *F* and *G* under one-one correlation (1884, §75), or that there is something which immediately succeeds 0 (1884, §77). Equally, he justifies his diagnosis that the formalists have only introduced empty signs instead of new number-words on their failure to prove the existence of the new numbers (1884, §92ff.). The great importance of existence proofs comes out in the sharp contrast Frege draws, on several occasions, between defining a concept by means of the properties an object must have to fall under the concept and proving that something does fall under the concept (see 1884, §74, 1893, xiv.). This shows that definitions, on his account, can never bring objects into existence, but only specify concepts hitherto lacking a designator.

8 NUMBERS AS EXTENSIONS

Frege's project thus clearly has an ontological aim, to discover what is already there. Weiner's denial of this point even with respect to the *Foundations* is not convincing. She claims that there is no argument in the *Foundations* that numbers are really extensions, and points at §69 and §107 for evidence of this. Concerning §69, she seems to suggest that Frege avoids addressing the ontological question 'Are numbers extensions?', and asks instead a linguistic question, motivated by the context principle, namely whether the assertions which we make about extensions are assertions we can make about numbers. But Frege does not believe that in answering the linguistic question he is eschewing the ontological question, or dismissing it as out of place. On the contrary, his interest in language has an explicit ontological agenda: 'There is no intention of saying anything about the symbols; no one wants to know anything about them, except insofar as some property of theirs directly mirrors some property in what they symbolise' (1884, §24). The context principle, as formulated in the *Foundations*, is not employed as an anti-ontological tool, but is called to dispel the psychologistic prejudice that an expression can stand for an entity only when, taken in isolation, we associate a mental idea with it (1884, §59). Thus the context principle actually serves the ontological agenda: numbers are self-subsistent entities because their expressions can be construed as singular terms, not in isolation, but at least in the sentential contexts of their most paradigmatic arithmetical use, which is also ontologically significant (equations understood as identities).

Frege's remark that he does not attach decisive importance to bringing in extensions (1884, §107) is also not evidence against his ontologism. The opposite is true. Frege brings in extensions in the course of offering his 'explicit' definition of number. This definition is given expressly in response to the 'Julius Caesar' problem, namely that all recognition statements of the form '$N_\tau F(\tau) = a$' must have a sense, i.e. that we must decide for any object *a* whether the statement is true or not (1884, §66-8, §107). But of course, recognition statements are identity statements, and their truth is a criterion for the objecthood of the content of the signs flanking the identity sign (1884, §62, §107). Hence, what Frege is most concerned with here is, again, an ontological issue: to specify, or at least sketch, a logicist criterion for the objecthood of numbers. He brings in extensions of concepts for this, which are logical objects on his account. The wariness he exhibits in §107 is therefore certainly not about the need to import *some* suitable objects to underpin his definition: such an import is essential, not indifferent to his definition of number. We can see this from the defence of his own suggestion that one could write 'concept' instead of 'extension of the concept' in his definition of number (1884, §68fn): by substituting 'concept' for 'extension of the concept' in the definition 'the Number which belongs to the concept *F* is the extension of the concept "equinumerous to the concept *F*"' the word 'concept' would be preceded by the *definite article*, and the whole phrase ('the concept') would be thus still a singular term, determining numbers as objects.[26] The wariness is rather about the fact that the phrase 'extension of the concept' is itself left undefined ('presupposed') in the *Foundations*, and hence that the 'Julius Caesar' problem remains open; for to decide whether $N_\tau F(\tau) = a$ we will have to be able to decide sharply whether *a* is a certain extension. So Frege's wariness has an ontological motivation, that of securing a sharp objecthood criterion for numbers, which is not achieved just by employing the notion 'extension of the concept'. This interpretation is confirmed once we look at Frege's mature solution to the problem, as offered in *Grundgesetze*, where he brings in extensions as value-ranges, not as *defined*, but as *primitive* objects.[27] This is obviously an ontological move, moreover one entirely untouched by the stipulative role of definitions. Weiner's insistence on the stipulative role of definitions, as allegedly ruling out an ontological agenda, is out of focus.

[26] See Frege (1892, 48.) See also the illuminating discussion in (Burge 1984, 274-84.)
[27] See Dummett (1991, 159.)

9 CONCLUSION

I hope to have shown two things in this paper. First, Weiner's interpretation of the notions of '*Foundations*-content', sense, reference and truth is extremely problematic. Second, the obvious requirement and the ontological thesis are very likely correct. Frege is a Platonic realist. He aims to define and analyse pre-existing arithmetical objects and concepts. A full defense of this view would have to look in more detail at the role of definitions in both *Foundations* and *Grundgesetze*, and the relation between the two, which cannot be done here.[28] In any case, I think it is more than probable, given the discussion above, that we have little chance to understand Frege's project of defining number, if we neglect his ontological agenda.

One final remark is called for. While Weiner is wrong in underplaying Frege's ontological agenda, she is surely on more firm ground in stressing the epistemological aim of his project. However, this also needs qualification. Weiner sees Frege realising the epistemological aim by means of a semantic undertaking, the sharpening of pre-systematic arithmetical language. But it is unclear how such a sharpening, by itself, would ever satisfy Frege's Cartesian craving for absolute certainty. Consider his repeated insistence that the *Foundations* have only established a probable thesis (1884, §87, §90), while his demand for total proof aims 'to place the truth of a proposition beyond all doubt' (1884, §2), to give it 'absolute certainty that it contains no mistake and no gap' (1884, §91), to raise the probability that arithmetical truths are analytic and *a priori* to a certainty (1884, §109) etc.[29] Clearly, vagueness of concepts is not the only source of doubt and error; a thinker might have sharpened all his concepts and still not be able to reach more than probable knowledge. '*X* is a sharp concept, but it is uncertain whether *y* falls under *X*' is not incoherent. Adding the gapless proof requirement does not yield the desired certainty either, as we still need to access the unshakeable ground on which the derived propositions rest, the axioms expressing primitive truths (*Urwahrheiten*, *Urgesetze*, §2-4).[30] Frege has his eyes set on more than just increased conceptual and proof-technical rigour, to be achieved by mere stipulations. Instead, he formulates a programme of genuinely reductive analysis: an arithmetical truth has found its epistemological classification if we can trace its proof back to the primitive truths (1884, §4), whose number we have reduced to a minimum (1884, §2). Since primitive truths are truths evident without further proof, they must involve an

[28] See my forthcoming article investigating this.

[29] See also his talk about the 'unconditional assurance against a proof or a gap' (1884, §91fn.) and 'the secure ground under our feet (1903, §62).

[30] Frege's simile is that of the 'Unerschütterlichkeit eines Felsblockes', which is best translated as 'unshakeability of a boulder'. This places Frege's metaphor in the gravitational orbit of Descartes' *fundamentum inconcussum*.

indubitable source of knowledge. Hence, Frege's epistemological project has a foundationalist and rationalist agenda. Moreover, there is no tension between Frege's foundationalism and his ontological agenda. On the contrary, the former presupposes the latter. The truths about logical objects are self-evident because of their nature: 'In arithmetic we are not concerned with objects which we come to know as something alien from without through the medium of the senses, but with objects given directly to our reason and, as its nearest kin, utterly transparent to it. And yet, or rather for that very reason, these objects are not subjective fantasies. There is nothing more objective than the laws of arithmetic' (1884, §105). In fact, Platonism is the basis of all knowledge: 'If there were nothing firm, eternal in the continual flux of all things, the world would cease to be knowable, and everything would be plunged in confusion' (1884, vii)[31]. At the ultimate level, epistemological questions are intimately bound with ontological ones.

REFERENCES

Note: *Begriffsschrift*, *Foundations* and *Grundgesetze* quotations always refer to the sections of the books. Frege's other published writings are cited by the original page. Posthumous writings are cited by the English translation in Gabriel et al. (eds.), 1979.

Benacerraf, P., 1981, Frege: The last logicist. *Midwest Studies in Philosophy 6*. 17-35.
Burge, T., 1984, Frege on extensions of concepts. In Burge 2005.
Burge, T., 1990, Frege on sense and linguistic meaning. In Burge 2005.
Burge, T., 2005, Truth, Thought, Reason: Essays on Frege. Oxford, Oxford University Press.
Carnap, R., 1910-1914/2004, *Frege's Lectures on Logic: Carnap's Student Notes*.
Dummett, M., 1991, *Frege: Philosophy of Mathematics*. Cambridge/Mass., Harvard University Press.
Frege, G., 1879, *Begriffsschrift*. Halle, Louis Nebert.
Frege, G., 1879-1891, Logik. In Frege 1983.
Frege, G., 1884, *Grundlagen der Arithmetik*. Breslau, Wilhelm Koebner.
Frege, G., 1885, Über formale Theorien der Arithmetik. In Frege 1990.
Frege, G., 1892, Über Begriff und Gegenstand. In Frege 1990.
Frege, G., 1893, *Grundgesetze der Arithmetik I*. Jena. Verlag Hermann Pohle.
Frege, G., 1899, Letter to Hilbert, 27.12.1899. In Gabriel et al. (eds.), 1980.
Frege, G., 1903, *Grundgesetze der Arithmetik II*. Jena. Verlag Hermann Pohle.
Frege, G., 1906a, Über die Grundlagen der Geometrie I. In Frege 1990.
Frege, G., 1906b, Einleitung in die Logik. In Frege 1983.
Frege, G., 1914, Logik in der Mathematik. In Frege 1983.
Frege, G., 1918a, Der Gedanke. In Frege 1990.

[31] English translation amended by the author.

Frege, G., 1918b, Die Verneinung. In Frege 1990.
Frege, G., 1919, Aufzeichnungen für Ludwig Darmstaedter. In Frege 1983.
Frege, G., 1924-1925, Sources of knowledge of mathematics and the mathematical natural sciences. In Frege 1983.
Frege, G. 1983, *Nachgelassene Schriften*. Hamburg, Felix Meiner.
Frege, G. 1984, *Collected Papers on Mathematics, Logic, and Philosophy*. Oxford, Basil Blackwell.
Frege, G. 1990, *Kleine Schriften*. Hildesheim, Georg Olms Verlag.
Gabriel, G., Hermes, H., Kambartel, F., Thiel, C., and Veraart, A. (eds.), 1980, *Philosophical and Mathematical Correspondence*. Chicago, University of Chicago Press.
Geach, P. and M. Black (eds.), 1980, *Translations from the Writings of Gottlob Frege*. Oxford, Basil Blackwell.
Hermes, H., F. Kambartel, F. Kaulbach (eds.), 1979, *Posthumous Writings*. Chicago, University of Chicago Press.
Kemp, G. 1996, Frege's Sharpness Requirement. *Philosophical Quarterly 46*. 168–84.
Picardi, E. 1988, Frege on definition and logical proof. In Cellucci C., Sambin, G. (eds.), *Temi e Prospettive della Logica e della Filosofia della Scienza Contemporanee*, vol. i. Bologna, Cooperativa Libraria Universitaria Editrice Bologna.
Rumfitt, I. 2003, Singular terms and arithmetical logicism. *Philosophical Books 44*. 193-219.
Shieh, S. 2008, Frege on definitions. *Philosophical Compass* 3/5. 992-1012.
Stekeler-Weithofer, P. 1986, *Grundprobleme der Logik: Elemente einer Kritik der formalen Vernunft*. Berlin/*New York. Gruyter*.
Weiner, J. 1984, The philosopher behind the last logicist. In Wright (ed.), 1984.
Weiner, J. 1990, *Frege in Perspective*. Ithaca NY, Cornell University Press..
Weiner, J. 2007, What's in a numeral? Frege's Answer. *Mind 116*. 677-716.
Wright, C. (ed.), 1984, Frege: Tradition and Influence. Oxford, Basil Blackwell.

Nenad Miscevic

The Indispensability of Logic

Abstract. The paper discusses the currently prominent strategy of justifying our elementary logical-inferential practices by their unavoidability and global indispensability for all our cognitive efforts. It starts by agreeing with prominent apriorists about their attempt to justify such beliefs based on constitutiveness (Boghossian) or based on the Global Indispensability Argument (C. Wright), and then proceeds to argue that unavoidable and indispensable tools provide entitlement/justification for projects if those projects are themselves meaningful. However, we are justified to think that our most general cognitive project is meaningful, and justified partly on the basis of its success up until now; and this basis is *a posteriori*. Therefore, the whole reflective justification from compellingness and unavoidability is *a posteriori*. This suggests that the justification of our intuition-based armchair beliefs and practices in general is plural and structured, with *a priori* and *a posteriori* elements combined in a complex way. It seems therefore that *a priori*/ *a posteriori* distinction is useful and to the point. What is needed is refinement and respect for structure, not rejection of the distinction.

1 INTRODUCTION

How is simple and naïve logical reasoning justified? If, in the case of naïve cognizer, we distinguish the immediate entitlement, normally not consciously available to her, and in this sense 'external' to her, from reflective justification reserved for more sophisticated cognizers, our question branches into two. First, where does the entitlement of a naïve thinker who spontaneously and unthinkingly uses Modus Ponens or Conjunction Elimination come from? And when the naïve thinker becomes more sophisticated, where does her reflective justification come from? One powerful family of arguments for ultimate justification of our logical practice(s) concerns the following facts: simple rules of logic are compelling and unavoidable for humans, they enable the very having of

beliefs and constitute the rationality of reasoning. Because of this, they are both unavoidable and indispensable for our thinking, and for any sort of cognitive projects we might engage in. This secures the entitlement for the naïve reasoner, and justification for the sophisticated, reflective thinker. The apriorists add that these two, entitlement and justification, are a priori. This line of argument and the resulting family of arguments is extremely popular in contemporary debates, represented by leading thinkers on the matter, such as C.Wright, P.Boghossian, and to some extent P. Horwich and C. Peacocke (see References).

I shall assume here that the above line of argument is plausible indeed, but will argue that it is as yet incomplete and demands a further step that leads it away from the conclusion these thinkers prefer. I shall point out that *unavoidable and indispensable tools provide entitlement/justification for projects only if projects are themselves meaningful, and provide reflective justification only if we are also justified in finding them meaningful. However, we are justified to think that our most general cognitive project is meaningful partly on the basis of its success up until now; and this basis is available only a posteriori; we are justified in trusting that these projects are meaningful a posteriori, because they seem to have worked decently well up until the present. Therefore, the justification from compellingness and unavoidability is a posteriori, both in its role of warrant, usually spontaneously had by the naïve thinker, and in its role of reflective justification, found out and deployed by philosopher(s).*

The line of thought to be proposed in this paper has important similarities and dissimilarities with the traditional Quinean argument for indispensability. The similarity is in the basic appeal to a posteriori considerations of indispensability and success. The difference is structural: first, it is only one line among several, and it does not exclude some prima facie justification by obviousness and compellingness which is a priori; second, it concerns either the entitlement that is normally external and not present to the cognizer's awareness, or, if she is a reflective thinker, then also her high-level reflective justification.

In this paper I will present this line of argument in more detail, using as my foil the work of Boghossian and C. Wright. I apologize for not going into a longer debate with them, for lack of space. Let me start.

Our topic is the ordinary, naïve use of logic in reasoning. So, take a naïve reasoner N who passes from accepting a conjunctive statement to accepting one of its conjuncts (and acting upon her belief in it), thereby performing (what we would describe as) a step of elimination of conjunction. This can be brought to light by explicitly asking her a question about *p*-and-*q* but not-*p* situation, e.g. whether it is possible that the whole (conjunctive) statement holds, i.e. that the complex situation obtains, without things being as the relevant conjunct describes them. The "Of course, not" answer would confirm the impression that she does have a mastery of the rule governing conjunction. Call this knowledge instance-knowledge. It seems that knowledge manifested in such spontaneous inferences is knowledge how. However, when a naïve thinker begins to reflect,

the inferential step seems obvious, at least in its concrete implementation: the impossibility of the combined concrete situation is obvious, vividly presented to her. Thus she finds the situation in which it is the case that *p*-and-*q* but not-*p* (for some instances of "*p*" and "*q*") inconceivable.

Let us follow common sense in assuming that N is making no mistake in reasoning the way she does. Let us refine this assumption by ascribing to N an immediate entitlement that she would find difficult to formulate; part of this entitlement stems from the fact that Conjunction elimination is in fact a valid rule, and part of it, we shall argue (along with our apriorists) comes from the indispensability of such steps for N's reasoning, cognitive project(s) and action. N could learn logic and epistemology and become reflectively aware of the issues involved; call N at that stage N*. Or, an epistemologist might start reflecting upon N's (or her own) performance. They would then pass to a level of reflective internal justification.

In principle we thus have two directions of reflection: 1st person type, exemplified by our sophisticated N* character, and 3rd person type, exemplified by our epistemologist. It would be interesting to explore similarities and differences between the two, but I must be short. Let me just state my view: 1st person reflective justifiedness is necessary for being completely justified. I am using as my framework a two-level picture of justification that I have picked up from the work of Russell and Sosa (see their works in References, and also the paper by Tom Baldwin listed there); but I hope that the results of the discussion can be easily applied, *mutatis mutandis*, to other frameworks; for instance, if you believe only in conscious, reflective justification, the result will be relevant for your framework as well. On the more externalist side, even if you disagree with my line on the importance of 1st person reflective knowledge, you might think that completely blind spontaneous thinking is not sufficient for justification. Only the most simple reliabilists think it is. Others think that some sort of availability of the 3rd person type justification is needed. (Peacocke and Boghossian might be good examples). For all of them, the issue of indispensability is crucial. So, what is exactly the role of indispensability, and what is its character; is the justification it bestows a priori or a posteriori?

2 THE INDISPENSABILITY ARGUMENT(S)

Here is then the preliminary sketch of my main argument:

(0) Logical practice and beliefs stand in need of entitlement and justification.
(1) Simple rules of logic are indispensable for humans, therefore

(2) they are both unavoidable and indispensable for our thinking, and for any sort of cognitive projects we might engage in.
(3) Unavoidable and indispensable tools provide entitlement/justification for projects iff the projects are themselves meaningful.
(4) Our most general cognitive project has been at least minimally successful, and therefore it is meaningful and we are justified in believing that it is, and the naïve thinker is entitled to her logical reasoning.
(5) This justification and entitlement are to a large extent a posteriori.

We now look briefly at each step. First, premise (0): the need for entitlement and justification. By entitlement for A-ing we mean that the A-er is permitted to A and is blameless in doing so. We also allow for first-level justification: obviousness and immediate compellingness does provide some good reason, but not all obvious and compelling beliefs and procedures are justified. Nash once said that he is receiving messages from extraterrestrials which come to him with the same degree of persuasiveness (and he probably meant obviousness and immediate compellingness) as his mathematical theorems. Since this phrase is printed on the cover of the paperback edition of the "Beautiful Mind", let me call the problem The Beautiful Mind Problem. It introduces the need for second-level, reflective justification: why do I trust my use of Conjunction elimination, if geniuses like Nash found their logical reasoning as persuasive as messages from extraterrestrials? It is a more moderate problem than the related extreme Cartesian Problem of Madness which C. Wright uses as part of the motivation for his proposal for the appeal to indispensability.

In a discussion at Tim Williamson's lecture in Dubrovnik (in the summer of 2009) he formulated an interesting point against The Beautiful Mind Problem: if you concede this much to the skeptic and you let the problem in, it will never get out! However, The Beautiful Mind Problem is not as global as the Madness problem, although the skeptic might try a slippery slope ascent from the former to the latter. Instead, the former merely explores and exploits our occasional doubts that are neither pathological nor, to my mind exaggerated. And we don't have to convince the extreme skeptic, just come up with reasons we find good enough for trusting our reasoning faculties and strategies. So, we can uphold

(0) Logical practice and beliefs stand in need of entitlement and justification.

We now pass to the next two premises, that contains the main point of the first part of our argument, stages (0)–(3), namely the one about indispensability. We shall discuss (1) and (2) together.

(1) Simple rules of logic are indispensable for humans.
(2) They are indispensable for our thinking, and for any sort of cognitive projects we might engage in.

So let me comment on (1) and (2) and summarize the extant arguments for them.

Let me start from what we might dub The Constitutiveness Argument, or Constitutiveness variant of the Indispensability Argument, as stated by Boghossian in his "Knowledge of logic" and developed in the recent book (2008). Then we shall pass to C. Wright's Global Indispensability argument (2004) (the name for the argument is ours!). Boghossian's proposal develops around the idea that logic can be justified in a rule-circular manner, due to its indispensability for thinking almost any contents whatsoever. Without dispositions to reason in accordance with logic, we could not even have the general belief whose justification is supposed to be in question, i.e. the belief about inferential potentials of a given logical constant. In a nutshell, the argument can be reconstructed as follows:

> Certain of our inferential dispositions fix what we mean by our logical words (in the language of thought), therefore
> without those dispositions there is nothing about whose justification we can even intelligibly raise a question.
> Moreover, without those dispositions we could not even have the general belief whose justification is supposed to be in question. Therefore
> We are entitled to act on those inferential dispositions prior to, and independently of, having supplied an explicit justification for the general claim that they are truth-preserving.

Here is the brief background story. The chapter "Epistemic Analyticity: A Defense", of Boghossian's (2008) book, starts with a piece of self-criticism: "at the time of writing that paper (i.e. "Analyticity Reconsidered"—NM), I did not delineate sufficiently clearly the difference between inferential and constitutive construals of the relation between meaning and entitlement" (Boghossian 2008, 225). The distinction is crucial for the author's more recent work. Boghossian notes that the Implicit Definition Template involves a lot of inferring that already uses logic. So this premise-and-derivation model can neither entitle the reasoner to the crucial premise (3), i.e. that S(f) is true, nor explain what her entitlement to reason according to certain deductive rules consists in, since the entitlement presupposes such reasoning. We therefore need a contrasting model, and its crucial point is that the mere fact that the thinker grasps S's meaning entails that the thinker is justified in holding S to be true. (The epistemological consequences of the proposal are then developed in the next chapter, "How are objective epistemic reasons possible?".) Take conditionals. If I don't follow

Modus Ponendo Ponens (MPP), I can't have if-thoughts at all. So, if I do follow it, with *p* and 'if *p* then *q*' as my premises, I cannot be blamed, so, I am entitled to follow it.

> If inferring from those premises to that conclusion is required if I am to have the ingredient propositions, then, as a matter of metaphysical necessity, I cannot so much as consider the question whether the inference is justified without being disposed to reason in that way. Under those circumstances, then, it looks as though inferring according to MPP cannot be held against me, even if the inference is, as I shall put it, blind–unsupported by any positive warrant (Boghossian 2008, 230).

The chapter concludes by stressing that according to the "Constitutive model" the most fundamental relation between grasp of meaning and entitlement occurs when a thinker is entitled to reason in accordance with a certain rule simply by virtue of the fact that this rule is constitutive of a concept of his. The author expresses his hope that the model can be extended from reasoning to beliefs, if they are similarly constitutive of the possession of a concept (which has to be non-defective—we shall come to this in a moment). He proposes that this will solve the issue, famously raised by Aristotle (in *Metaphysics* Γ), about our entitlement to accept the principle of non-contradiction.

Boghossian has been developing the first line, on meaning-constitutiveness as the main *a priori* justifier, combined in his "Knowledge of Logic" with occasional remarks on compellingness, i.e. on the alleged fact that "it is not open to us to regard our fundamental logical beliefs as unjustifiable." (Boghossian and Peacocke 2000, 253) For instance, in the same paper Boghossian argues for the warrantedness of logical rules mainly based on negative compulsion, i.e. from deeply felt unacceptability and inconceivability. He does not offer any causal or psychological explanation of compulsion, which is, after all, a felt item. Here is the relevant quote:

> [W]e cannot accept the claim that we have no warrant whatsoever for the core logical principles. We cannot conceive what such a warrant could consist in […] if not in some sort of inference using those very core logical principles.

And further down:

> It is not open to us to regard our fundamental logical beliefs as unjustifiable. (Boghossian 2000, 253.)

Another variant of the appeal to indispensability is the Global Indispensability Argument due to Crispin Wright. He revives the Wittgensteinian conception of hinges, generalizes it and enriches it with his own idea of "cornerstones":

> Call a proposition a cornerstone for a given region of thought just in case it would follow from a lack of warrant for it that one could not rationally claim warrant for *any* belief in the region. The best—most challenging, most interesting—sceptical paradoxes work in two steps: by (i) making a case that a certain proposition (or restricted type of proposition) that we characteristically accept is indeed such a cornerstone for a much wider class of beliefs, and then (ii) arguing that we have no warrant for it. (Wright 2004, 167-8.)

If a cognitive project is "rationally non-optional", i.e. indispensable in rational inquiry and in deliberation, then we may rationally take for granted the original presuppositions of such a project without specific evidence in their favor. The absence of defeating information is sufficient. So, elementary logic is both unavoidable and indispensable.

None of the arguments is final, as no philosophical argument is. But all of them converge on unavoidability and indispensability. If you believe in constitutive conceptual connections, then the Consitutivity Argument might appeal to you. And if you find Wittgenstein most congenial, on some of many readings of his text, then the Global Indispensability Argument will probably convince you. So, there are good reasons to accept the claim above.

3 ONLY A GOOD PROJECT JUSTIFIES ITS MEANS

We now pass to the second part of the argument, and move on to the particular twist we want to give to it, against its apriorist reading. Let us start with

(3) Unavoidable and indispensable tools provide entitlement/justification for projects iff the projects are themselves meaningful.

The premise encapsulates the commonsense wisdom that no justification can come from bad, impossible and/ or idiotic projects! E.g., imagine a beginner who reasons: "If we want to square the circle, we need theorem Θ; therefore Θ." He would be very quickly taught that the project is impossible, so the theorem needed for it cannot be justified by the need. Next, consider clearly impossible lifelong projects: I want to achieve, by exercise, the height of 12 ft, so I am rational in doing the exercise. And finally, morally bad projects also demand lots of means: if one wants to build a torture chamber, one needs electricity at the least. So, one would desire some electricity for this purpose. Does indispensability of electricity morally justify the instrumental desire? Not really.

Corine Besson has objected (at a talk in Geneva, winter 2010) that even bad projects yield some instrumental justification for their means (thanks, Corine).

It seems to me that such a justification is conditional: if the project P is a good one, *then* means M is justified. It cannot be detached, and used independently, until P is independently justified.

Let us now consider the application of this reasoning to cognitive projects. Here is a story about *Mr. Magoo*.

> Mr. Magoo has a very defective cognitive apparatus. His inductive propensities are idiotic, to use politically incorrect vocabulary, his senses most often deceive him, and his "heuristics" are ridiculous. (He lives in a super-hospitable environment, but hardly manages to survive.) His idiotic inductive propensities and ridiculous "heuristics", plus his misplaced uncritical trust in his senses are indispensable for his ever forming any belief. Therefore, he is warranted in taking them as unquestioned and unquestionable starting points.

Of course, he is not warranted, most people would say, he is just being stupid. We can derive an argument from this kind of reaction: if you don't find Mr. Magoo's reasoning convincing this suggests that the justification of means for a cognitive project depends on the meaningfulness of the project itself. If the project has no chances to succeed, if *it* is hopelessly flawed (in the given context), then *it* is not justified. If it is not justified, it cannot lend its justifiedness to the means, since it does not have any.

This is valid both for entitlement and would be valid for reflective justification, if Mr. Magoo were capable of producing one. He is not entitled to his propensities and heuristics, nor to his trust in his senses. And if he were able, per impossibile, to produce a piece of reflective attempt at self justification, that would fail as well.

Let us apply this "Mr. Magoo Argument" to Crispin Wright's idea of cornerstones for a project. It seems that the acceptance of hinges and cornerstones is justified by the quality of cognitive projects they enable, and is sensitive to the chances of their success. But these chances are revealed by trying. Therefore, our best access to our own warrant involves information about the success of the relevant cognitive project. The warrant for logic is thus sensitive at least to the chances of success of our "total inquiry", and our awareness of it depends on the information about the success. Meaningfulness is not independent of chances of success, in virtue of "ought implies can" principle.

Would this make the final justification merely pragmatic? This is a question that I have often heard at talks. When meant as a criticism, I think it rests upon confusion. The justification is instrumental, but the goal appealed to is truth, reliability or some such epistemic value, so being instrumental does not make it into a pragmatic justification in a non-epistemic, merely practical sense. "Success" here is epistemic success, not practical-pragmatic. Alternatively, if

"pragmatic" just means instrumental, then the question does not entail criticism: there is nothing wrong with a piece of belief (or cognitive habit) being justified by its contribution to the achievement of epistemic value(s). So we come to the two final steps:

(4) Our most general cognitive project has been at least minimally successful; therefore, it is meaningful and we are justified in believing that it is, and the naïve thinker is entitled to her logical reasoning.
(5) This justification and entitlement are to a large extent a posteriori.

Now, the success of our "total inquiry" is to a large extent an empirical matter. Therefore, our awareness of it depends, to a great extent, on empirical information. Such information is a posteriori. Consider reflective justification: how does the cognizer arrive at justified beliefs about herself being warranted? Well, partly by relying on relative success of her total project. And this reflective justification might therefore be seriously a posteriori, in a way that precludes purely a priori justification of logic.

But you need deductive logic from the start of the inquiry, and you need inductive-logical assumptions for evaluating your empirical evidence; so you can't get rid of the a priori, the objector might argue. Not really: some assumptions may be pragmatically antecedent to a cognitive project, e.g. those that will later be classified as logical, but they are, firstly, reflectively justified by the overall success-chances of the project, and secondly, revisable in the light of some advanced stage of the project. We can thus conclude: *both first-order entitlement and reflective justification from indispensability are a posteriori*.

But aren't we back to the old Quine-Putnam indispensability? A comparison is needed. First, the old indispensability has been presented as replacing obviousness and compellingness, and second, it was implicitly presented as first-order property of our beliefs. This has made it vulnerable to quite dangerous objections: that it bypasses well-entrenched and extremely stable traditional justifiers, that it does not correspond to the first-order intuitions of practitioners themselves (for instance, mathematicians who just accept obvious-looking moves and their results). Our proposal avoids these pitfalls: it is compatible with a prima facie role for obviousness, and it situates the appeal to success by the thinker herself into the lofty spheres of reflective self-understanding where it belongs.

4 CONCLUSION: FROM INDISPENSABILITY TO APOSTERIORITY

Much more needs to be said, but let us summarize how far we have come: unavoidable and indispensable tools provide entitlement/justification for projects if projects are themselves meaningful.[1] However, we are justified to think that our most general cognitive project is meaningful, and justified partly on the basis of its success to date; and this basis is a posteriori. There is more than just a touch of aposteriority present in the considerations of meaningfulness, and much more in our coming to know about our warrant: reflective justification based on compellingness and unavoidability is wholly a posteriori. This suggests that the justification of our intuitional armchair beliefs and practices in general is plural and structured, with a priori and a posteriori elements combined in a complex way. It seems therefore that the a priori/ a posteriori distinction is useful and to the point.

Some philosophers suggest that we instead drop the a priori/ a posteriori contrast. (A. Goldman claims that warrant is just a complex and multi-dimensional affair (1999, 48), so the contrast is misplaced. T. Williamson argues for the same conclusion in his recent (2007) book. It is a bad idea: we need to distinguish and recognize structure, rather than obscure it. For instance, the immediate justification of logical moves is certainly a priori. But the reflective justification is not. What is needed is refinement and respect for structure, not rejection of the distinction. But, if there is structure, what is the final verdict? Is justification of simple logical moves and beliefs a priori or a posteriori? A traditional principle insisted on purity of the a priori: if justification (or entitlement) contains a posteriori elements, then it is ultimately a posteriori (e.g. if it is mixed and contains one a posteriori element, it is ultimately a posteriori).So, if you accept the principle, you might talk about structured a posteriority. If not, you would merely talk about structured justification and entitlement.

To conclude and reiterate: many prominent apriorists have given up on focusing their defense of knowledge of logic on traditional internalist grounds of obviousness, self-evident character and the like. Instead, they have revived the Indispensability Argument in a sophisticated setting, appealing to entitlement and reflexive justification, hoping to cleanse the argument from its Quinean a

[1] What about primitive compulsion as an alternative? It would solve the justification problem by "ought implies can" and offers some guidance, but opens the problem of being primitively compelled to hold true sentences (and to hold correct rules) one only partially understands. Worse, it offers a competing explanation: it is not stipulation per se but its irresistible force. (In addition, it is hard to see how a set of sentences does not yield analytic propositions by merely being held true, but does yield by by being irresistibly held true .) There is a possibility: retreat to anthropocentrism: well, these are *our* concepts. Irresistibility is the mark of conceptual character, and irresistibility-for-us is the mark of ownership, of being "our own" concepts. But this move is not opened to Boghossian, since it would open the door to epistemic relativism, a view he is justly combating.

posteriorist heritage. I have argued that in this very setting the Indispensability Argument brings a posteriority back, since the issue of both entitlement and of reflective justification of logical and elementary mathematical beliefs and inferential propensities is to be decided to a large extent on the basis of the global successfulness of our cognitive effort, which is largely an a posteriori matter. To a significant degree, logic and elementary mathematical understanding are reflectively justified in an a posteriori manner. This should prompt us to opt for a more structured view of justification, containing essentially a posteriori elements, but in a different form than in the Quinean tradition.*

REFERENCES

Baldwin, T., 2003, From knowledge by acquaintance to knowledge by causation. In Griffin, N. (ed.), *Cambridge Companion to Bertrand Russell*. Cambridge, Cambridge University Press.

Boghossian, P., 1996, Analyticity. In Hale, B. and Wright, C. (eds.), *A Companion to the Philosophy of Language*. Oxford, Blackwell.

Boghossian, P., and Peacocke C. (eds.), 2000, *New Essays on the A Priori*. Oxford, Clarendon Press.

Boghossian, P., 2000, Knowledge of logic. In Boghossian and Peacocke (2000).

Boghossian, P., 2001, How are objective epistemic reasons possible?. *Philosophical Studies 106*. 1–40. Reprinted in Boghossian, 2008.

Boghossian, P., 2003, Blind reasoning. *Proceeding of the Aristotelian Society* Supplementary Volume 77. 225-48.

Boghossian, P., 2008, *Content and Justification*. Oxford, Oxford University Press.

Goldman, A., 1999, A priori warrant and naturalistic epistemology. *Philosophical Perspectives 15*.

Horwich, P., 2000, Stipulation, meaning, and apriority. In Boghossian and Peacocke (2000).

Boghossian, P. and Peacocke C. (Eds.), 2000, *New Essays on the A priori*. Oxford, Oxford University Press.

Peacocke, Ch., 2004, *The Realm Of Reason*. Oxford, Oxford University Press.

Sosa, E., 1991, *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge, Cambridge University Press.

–, 2007, *A Virtue Epistemology—Apt Belief and Reflective Knowledge*: Volume One. Oxford, Oxford University Press.

Russell, B., 1926, *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*.[2] Chicago and London, Allen and Unwin.

–, 1927, *An Outline of Philosophy*. London, Allen and Unwin.

–, 1940/1968, *An Inquiry into Meaning and Truth*. London, Allen and Unwin (paperback edition).

–, 1948, *Human Knowledge: Its Scope and Limits*. London, Allen and Unwin.

Williamson, T., 2007, *Philosophy of Philosophy*. Oxford, Blackwell.

Wright, C., 2004, On Epistemic Entitlement: I. Warrant for Nothing (and Foundations for Free)?. *Proceeding of the Aristotelian Society*,167-212.

* I wish to thank the organizers of ELTE conference, and the participants in the discussion.

Edi Pavlović

Fitch's paradox and Labeled Natural Deduction System

Abstract. This paper introduces a relatively novel system of representing modal logic in a form of natural deduction. It then expands it to accommodate the epistemic operator and applies it to generate a more precise formulation of Fitch's paradox of knowability. Finally, an illustration of the paradox's pertinence to contemporary philosophical debate is laid out.

1 INTRODUCTION

The purpose of this paper is to provide the means for presenting Fitch's paradox, a philosophical argument requiring multiple modalities, within a purely formal deduction system. A labeled natural deduction system for modal logic offered by David Basin, Sean Matthews and Luca Vigano, provides the basis which is then expanded to accommodate an epistemic operator "know." An advantage of this system: anyone familiar with first-order natural deduction is provided with the means to formulate a useful and fruitful philosophical argument in a more precise manner at no added complexity.

The remainder of this section will lay out some desirable properties of any natural deduction system that we will naturally strive to meet in this paper. The second section introduces the labeled natural deduction system of Basin *et al.* and expands on it to allow us to formulate Fitch's paradox. Note that, while the authors use the "Gentzen-style" form of representing natural deduction, due to the relative length of the argument and the number of assumptions needed, for ease of presentation, the form used here is the "Suppes–Lemmon style." The third section presents Fitch's paradox first in an informal, and then in a formal manner. Finally, the fourth section provides the summary.

1.1 Natural deduction systems

Although natural deduction was first developed 1934, it is partially based on a proposal put forth in 1926 by Jan Lukasiewicz, who called for a system that can yield the same theorems as the axiomatic systems of the time, but which would follow more closely the actual practice of constructing a proof[1]. This system, reflecting the "natural" way humans reason, would follow where "arbitrary assumptions" lead and how long they stay in effect. This desirable property is something to keep in mind while presenting a natural deduction system.

In their widely used logic handbook *Language, Proof and Logic* Jon Barwise and John Etchemendy state that: "… [natural deduction] systems are intended to be models of the valid principles of reasoning used in informal proofs."[2] This is precisely the purpose of the natural deduction system they present.

2 LABELED NATURAL DEDUCTION SYSTEM FOR MODAL LOGIC

This section introduces a labeled natural deduction system developed by Basin *et al.* Following the ideas of Dov Gabbay, this system provides a framework for capturing a large number of non-classical logics[3]. The focus here will be on modal logic. The peculiarity of the system is that it introduces a set of labels W, $W = \{x_0, x_1, \ldots, x_n, \ldots, y_0, y_1, \ldots, y_n, \ldots\}$. These labels can be thought of as representing worlds in aKripke model. The language of the labeled natural deduction system (henceforth: LNDS) differs from the standard, and widely familiar, (propositional) modal logic language precisely with regard to W.

2.1 The language of LNDS

The language of LNDS comprises two types of formulas, labeled and relational well-formed formulas. The latter concern the relations of labels, and correspond to the properties of the accessibility relation R in a Kripke model, whereas the former are merely modal propositional wel formed formulas expanded with a label; we will define these first, and proceed from there. The (inductive) definition of a modal *wff* should be familiar:

[1] (Pelletier, F., 2000).
[2] (Barwise, J., Etchemendy, J., 2003).
[3] (Basin *& al.* 1998).

Definition 2.1: Modal *wff*
1 Propositional letters *P*, *Q* and *R* are well formed formulae (*wff*).
2 If *P* is a *wff*, then $\neg P$ is a *wff*.
3 If *P* is a *wff* and *Q* is a *wff*, then $(P \wedge Q)$ is a *wff*.
4 If *P* is a *wff* and *Q* is a *wff*, then $(P \vee Q)$ is a *wff*.
5 If *P* is a *wff* and *Q* is a *wff*, then $(P \rightarrow Q)$ is a *wf.f*
6 If *P* is a *wff Q* is a *wff*, then $(P \leftrightarrow Q)$ is a *wff*.
7 If *P* is a *wff*, then $\Box P$ is a *wff*.
8 If *P* is a *wff*, then $\Diamond P$ is a *wff*.
9 Nothing else is a *wff*.
Now we have all the ingredients necessary to define a *labeled well-formed formula*:

Definition 2.2: Labeled well-formed formula (*lwff*)
Let *P* be a modal *wff* (Def. 2.1), let *W* be a set of labels $W = \{x_0, x_1, \ldots, x_n, \ldots, y_0, y_1, \ldots, y_n, \ldots\}$, and let *x* be a member of such a set, $x \in W$. Then *x:P* is a *lwff* which can be understood as meaning "*P* is the case in (a possible world) *x*."
As noted earlier, a *relational wff* is concerned with a relation of two labels:

Definition 2.3: Relational well formed formula (*rwff*)
Let *x* and *y* be members of a set of labels *W* (as above). Then *xRy* is a *rwff* which can be understood as "*y* is accessible to *x*."

2.2 Rules of inference

In this section we will explore the rules of inference in LNDS. The rules of inference for truth-functional connectives should be readily recognizable to anyone familiar with natural deduction—the only novelty here being that each line is expanded with (one and the same) label. The rule for negation introduction deviates from this pattern, and will be discussed separately. Afterwards, rules of inference for modal operators will be covered, along with examples to illustrate them. The mode of presentation is "Suppes–Lemmon" style—the central column contains the enumerated steps of the proof, assumptions or derived formulas. The column on the right contains the "justification" of a step—a rule of inference used, or "*P*" if it is an assumption ("*P**" denotes an additional assumption that needs to be discharged, and the column on the left contains a set of undischarged assumptions the step "relies" on (assumptions "rely" on themselves). In a general form laid out here, the letters *m*, *n*, *i*, *j*, … signify numbers, Greek letters Γ and Δ signify sets of assumptions, and letters *A*, *B*, … signify *wff*s.

2.2.1 Truth-functional connectives

Below are the rules for conditional and conjunction, which behave in a familiar way, the only difference being that each *wff* is expanded into a *labeled wff*, using the same label in each instance.

$\rightarrow$ *Intro*			
$\{m\}$	m	$x{:}A$	P^*
		...	
$\Gamma \cup \{m\}$	i	$x{:}B$	
Γ	j	$x{:}\ A \rightarrow B$	$\rightarrow I{:}\ m,i$
$\rightarrow$ *Elim*			
Γ	m	$x{:}\ A \rightarrow B$	
Δ	i	$x{:}A$	
$\Gamma \cup \Delta$	j	$x{:}B$	$\rightarrow E{:}\ m,i$
$\wedge$ *Intro*			
Γ	m	$x{:}A$	
Δ	i	$x{:}B$	
$\Gamma \cup \Delta$	j	$x{:}\ A \wedge B$	
		or	$\wedge I{:}\ m,i$
$\Gamma \cup \Delta$	j	$x{:}\ A \wedge B$	$\wedge I{:}\ m,i$
$\wedge$ *Elim*			
Γ	m	$x{:}\ A \wedge B$	
Γ	i	$x{:}A$	$\wedge E{:}\ m$
	or		
Γ	i	$x{:}B$	$\wedge E{:}\ m$
$\neg$ *Intro*			
$\{m\}$	m	$x{:}A$	P^*
		...	
$\Gamma \cup \{m\}$	i	$y{:}\bot$	
Γ	j	$x{:}\ \neg A$	$\neg I{:}\ m,i$

Or, alternatively

$\neg$ *Intro*			
$\{m\}$	m	$x{:}\neg A$	P^*
		...	
$\Gamma \cup \{m\}$	i	$y{:}\bot$	
Γ	j	$x{:}A$	$\neg I{:}\ m,i$

Note that not all the lines here contain the same label (the label used in the line (*i*) is "*y*"). An impossible result in one world (i.e. under one label) can "transfer" to another world—a fact that Basin *et al.* call a "global falsum." This is elaborated in Section 2.4.

$\bot$ *Intro*			
Γ	m	$x{:}A$	
Δ	i	$x{:}\neg A$	
$\Gamma \cup \Delta$	j	$y{:}\bot$	$\bot I{:}\ m,i$

$\bot$ *Elim*			
Γ	n	$x{:}\bot$	
Γ	i	$x{:}A$	$\bot E{:}\ n$

2.2.2 Modal operators

The novelty of this approach consists in the introduction of natural deduction rules for the modal operators "$\Box$" ("necessarily") and "$\Diamond$" ("possibly"). Note that these rules of inference are analogous to the rules for universal and existential quantifiers, respectively (with an "arbitrary label" replacing an "arbitrary name").

$\Box$ *Intro*			
$\{m\}$	m	xRy	P^*
		...	
Γ	i	$y{:}A$	
$\Gamma - \{m\}$	j	$x{:}\ \Box A$	$\Box I{:}\ m,i$

Note: y is a new label, such that $x \neq y$, and not appearing in any of the suppositions in Γ, except perhaps $\{m\}$.

$\Box$ *Elim*			
Γ	m	$x{:}\ \Box A$	
Δ	i	xRy	
$\Gamma \cup \Delta$	j	$y{:}A$	$\Box E{:}\ m,i$

Example 2.1: See the Appendix.

◊ *Intro*			
Γ	*m*	*x*:*A*	
Δ	*i*	*xRy*	
Γ ∪ Δ	*j*	*x*: ◊ *A*	◊ *I*: *m*,*n*
◊ *Elim*			
Γ	*m*	*x*:*A*	*P**
{*i*}	*i*	*xRy*	*P**
{*j*}	*j*	*x*: ◊ *A*	◊ *I*: *m*,*n*
		. . .	
Δ ∪ {*i*} ∪ {j}	*k*	*z*:*B*	
Γ ∪ Δ	*l*	*z*:*B*	◊ *E*: *m*,*i*,*j*,*k*

Note: *y* is a new label, such that $y \neq x$ and $y \neq z$, which does not appear in any of the suppositions from Γ and Δ.

Example 2.2: See the Appendix.

2.3 Familiar axioms

As noted, it is expected of a natural deduction system that it provide the same results as an axiomatic theory. Therefore, what follows are proofs of two well-known modal axioms—the rule of necessitation, which states that every theorem is necessary, and axiom K, which demonstrates how the necessity operator "□" is distributed over conditionals.

Proof 2.1: Rule of Necessitation (*RN*)
Let *x*:*A* be a theorem, and *x* an arbitrary label. Proof for *x*: □ *A* will proceed as follows:

{*m*}	*m*	*xRy*	*P**
		the proof of a theorem where each occurrence of the label x is substituted for the label y	
{}	*n*	*y*:*A*	*from the preceding proof*
{}	*j*	*x*: □ *A*	□ *I*: *m*,*n*

Proof 2.2: Axiom K

{1}	1	x: $\Box (A \rightarrow B)$	*P**
{2}	2	x: $\Box A$	*P**
{3}	3	xRy	*P**
{1,3}	4	y: $A \rightarrow B$	$\Box$ *E*: 1,3
{2,3}	5	y:A	$\Box$ *E*: 2,3
{1,2,3}	6	y:B	$\rightarrow$ *E*: 4,5
{1,2}	7	x: $\Box B$	$\Box$ *I*: 3,6
{1}	8	x: $\Box A \rightarrow \Box B$	$\rightarrow$ *I*: 2,7
{}	9	x: $\Box (A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$	$\rightarrow$ *I*: 1,8

2.4 Global falsum

One consequence of the negation introduction rule is the rule called "global falsum": $\Gamma \vdash_{x:\perp} \Rightarrow \Gamma \vdash_{x:\perp}$.[4]

Proof 2.3: Global falsum (*gf*)
Suppose that (1) $\Gamma \vdash x{:}\perp$. Then Γ, $y{:}P \vdash x{:}\perp$ (adding a premise does not alter the validity of a valid argument). It follows by $\neg I$ that (2) $\Gamma \vdash y{:} \neg P$. But in the same way, from (1) we can derive Γ , $y{:} \neg P \vdash x{:}\perp$, and another application of $\neg I$ gives (3) $\Gamma \vdash y{:}P$. Applying $\perp I$ to (2) and (3) yields $\Gamma \vdash y{:}\perp$.
Since x and y represent arbitrary labels, it is obvious that falsum can "travel" freely between labels. The reason for the inclusion of the rule *global falsum* is that it allows a desirable result—interchangeability of W and $\neg \Diamond \neg$.

Global falsum

Γ	1	$x{:}\perp$	*gf*:*m*
Γ	2	$x{:}\perp$	

The following two proofs demonstrate how this inference rule allows for the derivation of that desirable result.

[4] (Basin *& al.* 1998).

Proof 2.4a

{1}	1	$x: \Box A$	*P*
{2}	2	$x: \Diamond \neg A$	*P**
{3}	3	$y: \neg A$	*P**
{4}	4	xRy	*P**
{1,4}	5	$y:A$	$\Box E$: 1,4
{1,3,4}	6	$y:\bot$	$\bot I$: 3,5
{1,3,4}	7	$x:\bot$	*gf*: 5
{1,2}	8	$x:\bot$	$\Diamond E$: 2,3,4,6
{1}	9	$x: \neg \Diamond \neg A$	$\neg I$: 2,7

Proof 2.4b

{1}	1	$x: \neg \Diamond \neg A$	*P*
{2}	2	xRy	*P**
{3}	3	$y: \neg A$	*P**
{2,3}	4	$x: \Diamond \neg A$	$\Diamond I$: 2,3
{1,2,3}	5	$x:\bot$	$\bot I$: 1,4
{1,2}	6	$y:A$	$\neg I$: 3,5
{1}	7	$x: \Box A$	$\Box I$: 2,6

2.5 Relational rules

Relational rules have the general form $t_1Rs_1 \ldots t_mRs_m \vdash t_0Rs_0$, where $t_0,t_1,\ldots,t_m,s_0,s_1,\ldots,s_m$ are members of the set of labels *W*. Relational rules mirror properties of the accessibility relation, and allow us to derive the corresponding axioms. The only rule necessary for the construction of Fitch's paradox is the relational rule of reflexivity, and it is therefore the only one presented here.

Reflexivity

	M		
{}	*I*	xRx	*Rrefl*:

Proof 2.5: axiom T

{1}	1	xRx	*Rrefl*:
{2}	2	x: $\Box A$	P^*
{2}	3	x:A	$\Box E$: 1,2
{}	4	x: $\Box A \rightarrow A$	$\rightarrow I$: 2,3

It is clear how this is in keeping with the historical requirement posed for natural deduction—to yield the same results as an axiomatic theory.

3 FITCH'S PARADOX

Fitch's paradox, also known as the paradox of knowability, first appeared in Fitch's 1963 article "*A Logical Analysis of Some Value Concepts.*" There, the paradox appears in Theorem 5, which states:

> If there is some true proposition which nobody knows (or has known or will know) to be true, then there is a true proposition which nobody can know to be true.[5]

However, an equivalent claim, which states that if all truths are knowable, then all truths are known, is usually considered when discussing the paradox:

$$\forall p\ (p \rightarrow \Diamond\ Kp) \vdash \forall p\ (p \rightarrow Kp)$$

3.1 Informal proof of the paradox

The strength of the paradox derives from the fact that it rests on mostly unproblematic principles. They are:

$$(KIT)\text{: } Kp \vdash p,$$

which states that knowledge is factive, i.e. knowledge implies truth.

$$(K\ Dist)\text{: } K\ (p \wedge q) \vdash Kp \wedge Kq,$$

which states that knowledge is distributed over conjunction, i.e. knowledge of a conjunction implies knowledge of the conjuncts.

[5] (Fitch, F., 1963).

Furthermore, we must rely upon the rule of necessitation (*RN*)—all theorems are necessary. The fourth and final principle states that if p is necessarily false, it is impossible:

$$\text{(P4): } \Box \neg p \vdash \neg \Diamond p$$

Proof 3.1: Fitch's paradox[6]
Now, suppose that every truth is knowable: $\forall p\ (p \rightarrow \Diamond\ K\ p)$. Suppose also that we are not omniscient, that there is a truth which is not known: $\exists p\ (p \wedge \neg\ K\ p)$. Let p be such a truth:

$$(1)\ p \wedge \neg\ K\ p$$

Now, since every truth is knowable, so is (1):

$$(2)\ (p \wedge \neg\ K\ p) \rightarrow \Diamond\ K\ (p \wedge \neg\ K\ p)$$

Therefore, by *modus ponens*,

$$(3)\ \Diamond\ K\ (p \wedge \neg\ K\ p)$$

This, however, can be proven to be false. Let us suppose (for *reductio ad absurdum*):

$$(4)\ K\ (p \wedge \neg\ K\ p)$$

It follows by *K Dist* that both conjuncts are known:

$$(5)\ K\ p \wedge K \neg\ K\ p$$

And, applying *KIT* to the second conjunct, we get a contradiction:

$$(6)\ K\ p \wedge K\ p$$

That allows us to negate (4):

$$(7) \neg\ K\ (p \wedge \neg\ K\ p)$$

And, since (7) is a theorem, we can apply *RN* to get:

$$(8) \Box \neg\ K\ (p \wedge \neg\ K\ p)$$

Applying the fourth principle, *P4*, we get the opposite of (3):

$$(9) \neg \Diamond\ K\ (p \wedge \neg\ K\ p)$$

[6] (Brogaard, B., Salerno, S., 2008).

Obviouslym this means there is no unknown truth

$$(10) \neg \exists p\, (p \wedge \neg K\, p)$$

Or, in other words, that all truths are known:

$$(11)\ \forall p\, (p \rightarrow K\, p)$$

So, supposing that all truths are knowable leads us, very convincingly, to the conclusion that all truths are, in fact, known.

3.2 Formal proof of the paradox

Obviously, for the proof of the paradox to be constructed, we need to have rules for the operator *K*. These rules will mirror the inference rules for the necessity operator (using a separate accessibility relation, R_E) and will allow us to derive all the principles needed in the informal proof.

K Intro			
$\{m\}$	m	xR_Ey	P^*
		...	
$\Gamma \cup \{m\}$	i	$y{:}p$	
Γ	j	$x{:}\ K\, p$	$K\, I{:}\ m,i$
K Elim			
Γ	m	$x{:}\ K\, p$	
Δ	i	xR_Ey	
$\Gamma \cup \Delta$	j	$y{:}p$	$K\, E{:}\ m,i$

Additionally, the relational rule of reflexivity for R_E will be introduced. It insures the *KIT* principle in keeping with the proof of axiom T in *Proof 2.5*.

Proof 3.2: *K Dist*

{1}	1	$x: K(p \wedge q)$	*P*
{2}	2	$xR_E y$	*P**
{1,2}	3	$y: p \wedge q$	*KE*: 1,2
{1,2}	4	$y: p$	$\wedge E$: 3
{1}	5	$x: K p$	*K I*: 2,4
{6}	6	$xR_E z$	*P**
{1,6}	7	$z: p \wedge q$	*KE*: 1,6
{1,6}	8	$z: q$	$\wedge E$: 7
{1}	9	$x: K q$	*KI*: 6,8
{1}	10	$x: K p \wedge K q$	$\wedge I$: 5,9

Proof 3.3: *KIT*

{1}	1	$x: K p$	*P*
{}	2	$xR_E x$	R_E *refl*:
{1}	3	$x: p$	*KE*: 1,2

Obviously, the remaining principles have already been proven—*RN* in the Proof 2.1, and *P4* in the Proof 2.4b, substituting $\neg p$ for *A*.

Proof 3.4: Fitch's paradox in LNDS
The formal version of the proof starts out in the same way—assuming that p is an unknown truth (2), but that every truth is knowable, and thus $p \wedge \neg K p$ as well as (1). Again, this leads to the claim that it is possible to know that something is an unknown truth (3). We now set out to prove (in line 23) that this not the case. Note that for the sake of legibility, the words *label* and *world* are used interchangeably.

{1}	1	$x: (p \wedge \neg K p) \rightarrow \Diamond K (p \wedge \neg K p)$	*P*
{2}	2	$x: p \wedge \neg K p$	*P**
{1,2}	3	$x: \Diamond K (p \wedge \neg K p)$	$\rightarrow E$: 1,2

We assume that there is an accessible world y in which it is known that p is an unknown truth:

{4}	4	$xR_A y$	*P**
{5}	5	$y: K (p \wedge \neg K p)$	*P**

However, in that case $p \wedge \neg K p$ holds in y, and therefore, $\neg Kp$ holds in y. These steps correspond to an application of principles *K Dist* and *KIT*.

{}	6	$yR_E y$	R_E*refl*:
{5}	7	y: $p \wedge \neg K p$	*KE*: 5,6
{5}	8	y: $\neg K p$	$\wedge E$: 7

At the same time, if $p \wedge \neg K p$ is known in y, then $K p$ holds in y.

{9}	9	$yR_E z$	*P**
{5,9}	10	z: $p \wedge \neg K p$	*KE*: 5,9
{5,9}	11	z: p	$\wedge E$: 10
{5}	12	z: $K p$	*KI*: 9,11

Lines 8 and 12 are contradictory—they imply that something is an unknown truth can not be known.

{5}	13	y:$\perp$	$\perp I$: 8,12
{}	14	y: $\neg K (p \wedge \neg K p)$	$\neg I$: 5,13

Since y is an arbitrary world, it is necessarily unknowable that something is an unknown truth. This step corresponds to the line (8) of the informal proof.

{}	15	x: $\Box \neg K (p \wedge \neg K p)$	$\Box I$: 4,14

Now we need to perform the transformation from line (9) of the informal proof. To do so, we will assume that the opposite holds:

{16}	16	x: $\Diamond K (p \wedge \neg K p)$	*P**

Of course, if K (p $\wedge \neg K p$) is possible, then there is a world in which it is true:

{17}	17	y: $K (p \wedge \neg K p)$	*P**
{18}	18	$xR_A y$	*P**

But, even in that world, $\neg K (p \wedge \neg K p)$ is true (since it is, according to line 15, necessary). This leads to a contradiction:

{18}	19	y: $\neg K (p \wedge \neg K p)$	$\Box E$: 15,18
{17,18}	20	y:$\perp$	$\perp I$: 17,19

That contradiction transfers back to the original world:

{17,18}	21	$x{:}\bot$	*gf*: 20

And so the assumption in line *16* proves, be false, as we had hoped.

{16}	22	$x{:}\bot$	$\Diamond E$: 16,17,18,21
{}	23	$x{:} \neg \Diamond K (p \wedge \neg K p)$	$\neg I$: 16,22

Finally, we have shown that there are no unknown truths:

{1,2}	24	$x{:}\bot$	$\bot I$: 3,23
{1}	25	$x{:} \neg (p \wedge \neg K p)$	$\neg I$: 2,24

Of course, since p is an arbitrary proposition, it can be shown that all truths are, in fact, known. This transformation is trivial:

{26}	26	$x{:}p$	*P**
{27}	27	$x{:} \neg K p$	*P**
{26,27}	28	$x{:} p \wedge \neg K p$	$\wedge I$: 26,27
{1,16,27}	29	$x{:}\bot$	$\bot I$: 25,28
{1,26}	30	$x{:} K p$	$\neg I$: 27,29
{1}	31	$x{:} p \rightarrow K p$	$\rightarrow I$: 26,30

We have thus arrived at the conclusion that if all truths *can* be known, than all truths *are* known—Fitch's paradox of knowability.

3.3 Philosophical implications of the paradox—an illustration

The purpose of this section is to demonstrate that Fitch's paradox is not just of logical significance—it also makes a genuine and insightful philosophical contribution.

Timothy Williamson uses Fitch's paradox in his book *Knowledge and its limits*[7] to demonstrate, predictably, what the limits of our knowledge are. Let us briefly examine how. Williamson labels the thesis that all truths are known as *strong verificationism* (*SVER*):

$$SVER{:}\ \forall p\ (p \rightarrow K p)$$

[7] (Williamson, T., 2000).

This is, as Williamson puts it, an "insane sounding thesis" (p. 271). The more plausible sounding thesis that all truths are knowable Williamson calls *weak verificationism* (*WVER*):

$$WVER\colon \forall p\,(p \rightarrow \Diamond\, K\, p)$$

Obviously, the stronger thesis implies the weaker one, since all that is known can be known. But in order to demonstrate some limits to our knowledge, Williamson uses Fitch's paradox to demonstrate that the converse also holds—*WVER* implies *SWER*. They are therefore equivalent, and any objection to the insane-sounding thesis will apply to the more plausible formulation as well.

4. SUMMARY

We put forth two desirable qualities of natural deduction systems at the beginning of this paper. The first—that it provide the same theorems as an axiomatic theory by way of making and following arbitrary assumptions—has clearly been met: we have derived all the axioms needed to prove one famous theorem. Regarding the second property, we have used the system to model the principles of reasoning present in the informal proof. So this was, in a manner of speaking, a textbook example of what natural deduction is supposed to do. Moreover, the formal principles come with no added complexity for someone familiar with first-order natural deduction, yet they are able to contribute to a fruitful philosophical debate.

APPENDIX

Example 2.1

{1}	1	$x\colon \Box\,(A \wedge B)$	*P*
{2}	2	xRy	*P**
{1,2}	3	$y\colon A \wedge B$	$\Box\, E$: 1,2
{1,2}	4	$y\colon A$	$\wedge\, E$: 3
{1}	5	$x\colon \Box\, A$	$\Box\, I$: 2,4
{6}	6	xRz	*P**
{1,6}	7	$z\colon A \wedge B$	$\Box\, E$: 1,6
{1,6}	8	$z\colon B$	$\wedge\, E$: 7
{1}	9	$x\colon\ B$	$\Box\, I$: 6,8
{1}	10	$x\colon \Box\, A \wedge \Box\, B$	$\wedge\, I$: 5,9

Example 2.2

{1}	1	$x: \Diamond (A \wedge B)$	P
{2}	2	$y: A \wedge B$	P^*
{3}	3	xRy	P^*
{2}	4	$y:A$	$\wedge E$: 2
{2,3}	5	$x: \Diamond A$	$\Diamond I$: 3,4
{1}	6	$x: \Diamond A$	$\Diamond E$: 1,2,3,5
{7}	7	$z: A \wedge B$	P^*
{8}	8	xRz	P^*
{7}	9	$z:B$	$\wedge E$: 7
{7,8}	10	$x: \Diamond B$	$\Diamond I$: 8,9
{1}	11	$x: \Diamond B$	$\Diamond E$: 1,7,8,10
{1}	12	$x: \Diamond A \wedge \Diamond B$	$\wedge I$: 6,11

REFERENCES

Barwise, J., Etchemendy, J., 2003, *Language, Proof and Logic*. Stanford, CSLI.

Basin, D., Matthews, S., Vigano, L, 1998, Natural deduction for non-classical logics. In *Studia Logica* 60, 119-160.

Brogaard, B., Salerno, S., 2008, Fitch's paradox of knowability. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. Url=<(http://plato.stanford.edu>.

Fitch, F., 1963, A logical analysis of some value concepts. *The Journal of Symbolic Logic* 28(2), 135-142.

Pelletier, F., 2000, A history of natural deduction and elementary logic textbooks. In Woods, J., Brown, B. (ed.), *Logical Consequence: Rival Approaches* vol 1. 105-138. Oxford, Hermes.

Williamson, T., 2000, *Knowledge and its Limits*. Oxford, Oxford UP.

Jiří Raclavský

Partiality and Tichý's Transparent Intensional Logic

Abstract. The paper focuses on treating partiality within Tichý's logical system. Tichý's logic is two-valued and type-theoretic. His simple theory of types (and the deduction system for it) accepts both total and partial functions. Tichý's late framework is explicitly ramified. So-called constructions (roughly: algorithms) construct, e.g., values of functions at arguments; in some cases, however, they do not construct anything at all. This special partiality phenomenon is discussed in the second part of the paper.

1 INTRODUCTION

It will be convenient to begin with a sketch of the history of Pavel Tichý's transparent intensional logic (briefly *TIL*). One can find roots of TIL already in 1960s when Tichý construed intensions (functions from possible worlds) as classes of algorithms-procedures which were considered as meanings of (empirical) expressions; see Intensions in Terms of Turing Machines (Tichý 1969) reprinted in (Tichý 2004; hereafter *CP*). In the very beginning of the 1970's, Tichý modified Church's typed lambda calculus—accepting not only individuals and truth-values but also possible worlds; see An Approach to Intensional Analysis (1971) in CP. The system differs significantly from that developed by Montague (and Montagovians); the lack of space prevents me to give here a comparison (*cf.* Tichý's own remarks in CP 132-137 and the paper Two kinds of intensional logic, CP 307-325).

A new era of TIL (now explicitly designated by this name) is embodied in the large monograph (Tichý 1976) which remained unpublished. In this book, λ-terms and constructions recorded by them are explicitly distinguished (we will return to this issue later). Secondly, partial functions are admitted. Thirdly, natural deduction for the system is exposed. Selected parts of the book were

published as papers in the second half of 1970s; an exception is (Tichý 1982), which is a condensed paper on deduction.

In 1978 (*cf.* CP 269-270), Tichý added the temporal parameter; intensions are thus considered as functions from ⟨possible worlds, time-moment⟩ couples. Interesting logical analyses of temporal discourse (tenses, etc.) and episodic verbs were published by Tichý in 1980 (see CP). A more important modification of TIL is suggested in (Tichý 1988); Tichý exposed there a remarkable type theory which combines, in fact, simple and ramified type theory.

2 ADOPTING PARTIALITY

Let us begin with the recognition that there are both total and partial functions. Since many phenomena are to be modelled by partial functions (e.g., the chronology of American presidents in the actual world, or the individual concept "the king of France"), it is natural to accept them.[1]

Sometimes it is held that three-valued logic (3V-logic) captures partiality and so it is identical with two-valued logic which adopts partiality (2VP-logic). This is, however, a questionable matter. For 3V-logic—recognizing T (true), F (false), and U (unknown, undecided, …)—is a logic with total functions only. On the other hand, 2V-logic recognizes T and F and accepts also a lack of a value for some function(s). Thus domains of truth-values of 3V-logic and 2VP-logic do differ.[2] For instance, there are 27 unary 3V-truth-functions but there are just 9 total and partial unary 2VP-truth-functions:

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9
T	T	T	T	F	F	F			
F	T	F		T	F		T	F	

Clearly, the function f_4 is classical negation (often denoted by '¬'); it is, however, entirely missing in 3V-logic (this is why we should say that 3V-logic is not a classical logic). Of course, the 3V-function T→F, F→T, U→U looks like a counterpart of ¬. For an obvious reason, however, plenty of 3V-functions cannot be counterparts of any 2VP-functions. (I will return to the problem of representation at the end of the paper.)

Once partial functions are admitted, strange phenomena appear. For instance, Schönfinkel's reduction does not work because one multi-argument

[1] The reader knows that Imre Ruzsa stressed the importance of partiality (*cf.*, e.g., 1.2 in Ruzsa 1991).

[2] Of course, the acceptance of U or "gap" (as we may call it) is governed by the same intuition.

(m-ary; $m > 1$) partial function corresponds to more than one 1-argument function (Tichý 1982, 59-60); thus multi-argument functions are irreducible entities. Before we proceed further, let me introduce some notions.

3 TICHÝ'S SIMPLE THEORY OF TYPES

Tichý's simple theory of types—e.g., (Tichý 1982, 60)—treats both total and partial functions. It is quite general, since it has an unspecified basis B:

Let B consist of mutually non-overlapping collections of objects.

a) Any member of B is a *type over B*.

b) If $\zeta, \xi_1, \ldots, \xi_m$ are (not necessarily distinct) types over B, then $(\zeta\xi_1\ldots\xi_m)$, which is a collection of all total and partial functions from $\xi_1, \ldots, \xi_m$ into ζ, is a *type over B*.

The (specific) basis of TIL comprises ι (individuals), ο (truth-values T and F), ω (possible worlds), and τ (time-moments/real numbers). *Intensions* are functions from ω to (total or partial) chronologies of ξ-objects (a chronology is a function of type $(\xi\tau)$). Briefly speaking, intensions are functions from ⟨possible world, time-moment⟩ couples. '$((\xi\tau)\omega)$' will be abbreviated to '$\xi_{\tau\omega}$'. *Propositions* are of type $\text{ο}_{\tau\omega}$; *properties of individuals* are of type $(\text{οι})_{\tau\omega}$; individual offices (Tichý's term) are of type $\iota_{\tau\omega}$; etc. Objects which are not intensions may be called extensions. For instance, classical unary ($\neg$) or binary ($\wedge, \vee, \rightarrow, \leftrightarrow$) truth-functions are of types (οο) and (ooo), respectively; classical quantifiers ($\forall^\xi, \exists^\xi$) are of type $(\text{o}(\text{o}\xi))$; $=^\xi$ is of type $(\text{o}\xi\xi)$ ('$^\xi$' will be suppressed).[3]

4 CONSTRUCTIONS

To introduce the idea of constructions, consider the function:

$1 \rightarrow -2$
$2 \rightarrow 1$
$3 \rightarrow 6$
⋮ ⋮

This function can be reached by (infinitely) many different (mathematical) procedures. For instance, it is induced by multiplying an integer by itself and subtracting three from the result (i.e. by $(n \times n) - 3$) or by adding an integer to its square and subtracting what one gets by adding three to the integer from

[3] Ruzsa's type theory (*cf.* Ruzsa 1989, 3) does not allow some (types of) intensions which are admitted by Montagovians and Tichý. It should be added here that Tichý accepts not only functions of type $\xi_{\tau\omega}$ but also of type ξ_τ or ξ_ω (such functions are not called intensions in the present text).

the result (i.e. by $(n^2 + n) - (n + 3)$). To every such intuitive procedure, there corresponds a certain Tichý (numerical) construction. Tichý used λ-terms to record constructions and one may view constructions as so-called intensional (i.e. not extensional) senses of λ-terms. To get another analogy, recall hyperintensions ("structured meanings") often urged within the logical analysis of natural language. It seems also that Frege's *Sinn* or Russell's (structured) propositional functions are predecessors of constructions.

Unfortunately, a rigorous definition of constructions cannot be expounded here, see (Tichý 1988, 56-65) for that purpose. Omitting here so-called single and double execution, there are four kinds of constructions; their brief characterization is as follows. Let X be any object (a construction or non-construction) and C any construction; let v be any valuation (it is a field that consists of sequences of objects of given types):

1. *Trivialization* ${}^0X$ v-constructs X (i.e. ${}^0X$ takes X and leave it as it is).
2. *Variable* x_k v-constructs the kth member of the sequence of objects of a given type.
3. *Composition* $[CC_1 \ldots C_m]$ v-constructs the value of the function constructed by C at the string of objects (i.e. the argument for that function) which are constructed by $C_1, \ldots, C_m$; if C or C_1 (etc.) does not v-construct such object(s) or the function is undefined for that argument, $[CC_1 \ldots C_m]$ is v-*improper*—it does not v-construct anything at all.[4]
4. *Closure* $\lambda x[\ldots x \ldots]$ v-constructs, in a nutshell, the function which takes particular values of x to the objects v-constructed by $[\ldots x \ldots]$ on the respective valuations (e.g., $\lambda n[[n^0 \times n]^0 - {}^0 3]$ v-constructs the function sketched above).

One may thus say that these four kinds of constructions are objectual correlates of constants, variables (as letters), applications, and abstractions of λ-calculi. Realize, however, that constructions are not expressions—they are language-independent entities (the proper subject of Tichý's approach are constructions, not expressions of some formal language). For instance, the term '$\lambda n[[n^0 \times n]^0 - {}^0 3]$' denotes (stands for) the construction $\lambda n[[n^0 \times n]^0 - {}^0 3]$. Realize also that constructions are not set-theoretical entities. Note that the term '$\lambda n[[n^0 \times n]^0 - {}^0 3]$' denotes the procedure as such, not the aforementioned function constructed by $\lambda n[[n^0 \times n]^0 - {}^0 3]$ (analogously, '$[{}^0 8^0 \div {}^0 2]$' denotes the construction $[{}^0 8^0 \div {}^0 2]$, not its result—the number 4).

[4] The usual argument for the adoption of hyperintensions is this. Intensional semanticist suggests that all true mathematical sentences denote one and the same proposition (which is true in all possible worlds). Consequently, 'Xenia believes that 3+4=7', '49÷7=7' ∴ 'Xenia believes that 3+4=49÷7' is rendered as a valid inference which is obviously not. Hence more fine-grained entities than intensions are needed to be explications of meanings. For another reason consider 'Xenia calculates 3÷0'; the sentence surely describes the agent as related to a certain calculation, not to its (non-existing) result.

For non-circularity conditions, Tichý introduced a ramified theory of types. Its definition in (Tichý 1988, 66) has three parts: (a) types of ("classical") set-theoretic objects (*cf.* the simple-type theoretic part above), (b) types of constructions (some constructions are first-order constructions, belonging to the type $*_1$, other constructions are second-, third-, …, *n*-order constructions), (c) types of functions from/to constructions.

In the mid-1970s, Tichý already suggested that constructions are explications of (natural-language) meanings—having thus the following semantic scheme:

an expression *E*

expresses (means) in *L*

the construction, which is the *meaning* (or logical analysis) of *E* in *L*

constructs

an intension / non-intension / nothing (*cf.* '3÷0'), which is the *denotatum* of *E* in *L*

The value of an intension in a possible world w, time-moment t is the *referent* of an empirical expression *E* (such as 'dog', 'the king of France', 'It rains in London'); the denotatum and referent of a non-empirical expression are understood as identical.

For example (let w and t be variables v-constructing possible worlds and time-moments, respectively):

'The king of France is bald'	an expression *E*
$\lambda w \lambda t[{}^0\mathrm{Bald}_{wt}\ {}^0\mathrm{KF}_{wt}]$	the construction expressed by *E*[5]
$\langle w_1, t_1\rangle \rightarrow \mathrm{T}$	the proposition denoted by *E*
$\langle w_2, t_2\rangle \rightarrow$	(i.e. gap)
$\langle w_3, t_3\rangle \rightarrow \mathrm{F}$	
etc.	
T	the referent of *E* in w_1, t_1

It is not difficult to show that this semantic theory is capable to deal with puzzles created by "intensional" and "hyperintensional" contexts.

5 PARTIALITY AND FAILURE OF CLASSICAL LAWS

From the objectual viewpoint, logical laws are not strings of letters but constructions. It is clear that (let o be a variable v-constructing truth-values):

$$[{}^0\forall \lambda o[o^0 \vee [{}^0\neg o]]$$

[5] Trivializations of well-known mathematical or logical functions will be written in the infix manner (e.g., '$[{}^0 8^0 \div {}^0 2]$' instead of '$[{}^0 \div {}^0 8^0 2]$').

is tautological (the variable *o* is always a *v*-proper construction). However, this law is scarcely remarkable, one would rather declare that for any proposition, it obtains in w, t or it does not obtain in w, t (the excluded middle). Let p be a variable *v*-constructing propositions (i.e. objects of type $o_{\tau\omega}$). Then the following construction (which can be closed by $[^0\forall\lambda w[^0\forall[\lambda t$; similarly below) is contradictory:

$$[^0\forall\lambda p[p_{wt}{}^0\vee[^0\neg p_{wt}]]]$$

Since if some proposition is undefined in w, t, then $\lambda p[p_{wt}{}^0\vee[^0\neg p_{wt}]]$ *v*-constructs a partial class (partial characteristic function) which is empty, thus $\forall$ takes it to the truth-value F.

We get an analogous failure for the carelessly formulated De Morgan law for exchange of quantifiers. Let P be any construction of a proposition where P contains x as its free variable (e.g., $\lambda w\lambda t[x^0={}^0\mathrm{KF}_{wt}]$):

$$[[^0\neg[^0\exists\lambda xP_{wt}]]^0\leftrightarrow[^0\forall\lambda x[^0\neg P_{wt}]]]$$

The construction $\lambda w\lambda t[x^0={}^0\mathrm{KF}_{wt}]_{wt}$ (which is reducible to $\lambda x[x^0={}^0\mathrm{KF}_{wt}]$) can *v*-construct a partial class which is empty thus $\forall$ takes it to F, not to T as we wish.

To avoid the destructive power of partiality, formulating thus the correct versions of the laws, I suggest utilizing a "totalizer" overcoming the trouble. In Tichý's framework, there are three kinds of properties "be true" due to their applicability to (a) propositions, (b) constructions, (c) expressions (relatively to a given language *L*). Each kind has several variants; the (a)-kind has only two. The *"partial" truth property* of propositions (i.e. an object of type $(oo_{\tau\omega})_{\tau\omega}$) can be defined as:[6]

$$[^0\mathrm{True}^{\pi\mathrm{P}}{}_{wt}p]\equiv p_{wt}$$

thus certain propositions are not in the extension or the anti-extension of that property (in w, t). The *"total" truth property* of propositions can be defined as:

$$[^0\mathrm{True}^{\pi\mathrm{T}}{}_{wt}p]\equiv[^0\exists\lambda o[[p_{wt}{}^0=o]^0\wedge[o^0={}^0\mathrm{T}]]]$$

A partial proposition having no value in w, t belongs to the anti-extension of the property—it is not true in w, t.

Using $^0\mathrm{True}^{\pi\mathrm{T}}$ for "totalizing", the correct law of excluded middle is:

[6] 'C_{wt}' abbreviates '$[[Cw]t]$'. Of course, ^{0}KF is a simplification. The procedure consists in taking (a) the property "popular", (b) applying it to w and t (values of w and t), getting thus the extension of "popular", and then (c) taking "the king of France", (d) applying it to w and t, getting thus the individual who fills that office, and (e) asking whether that individual (if any) is in that extension—yielding thus T or F (analogously for other w's and t's).

$$[^0\forall\lambda p[[^0\mathrm{True}^{\pi\mathrm{T}}_{w\not t}p]^0\vee[^0\neg[^0\mathrm{True}^{\pi\mathrm{T}}_{w\not t}p]]]]$$

and the correct De Morgan law is:

$$[[^0\neg[^0\exists\lambda xP_{wt}]]^0\leftrightarrow[^0\forall\lambda x[^0\neg[^0\mathrm{True}^{\pi\mathrm{T}}_{wt}P]]]]$$

So it is clear that partiality affects the rules for substitutivity (e.g., whether o can be substituted by p_{wt}), which lead Tichý to the sophisticated theory exposed in (Tichý 1982) and in Indiscernibility of Identicals (Tichý 1986), where he paid closer attention to constructions involving identity.

6 BETA-REDUCTION, ETA-REDUCTION AND PARTIALITY

But there are more complications with partiality—even classical β-reduction fails (β-reduction rule says that $[\lambda x[\ldots x\ldots]C]$ is equivalent to $[\ldots C\ldots]$). Consider:

1. $\lambda w\lambda t[^0\neg[^0\mathrm{True}^{\pi\mathrm{T}}_{wt}\lambda w\lambda t[^0\mathrm{Bald}_{wt}\,^0\mathrm{KF}_{wt}]]]$
(the analysis of 'It is not true that the King of France is bald')

2. $\lambda w\lambda t[\,\lambda x[^0\neg[^0\mathrm{True}^{\pi\mathrm{T}}_{wt}\lambda w\lambda t[^0\mathrm{Bald}_{wt}x]]]\,^0\mathrm{KF}_{wt}]$
(the analysis of 'The King of France is such that it is not true that he is bald')

The two constructions v-construct distinct propositions because 1. v-constructs a total proposition whereas 2. v-constructs a partial proposition (if there is no king of France in w, t, $^0\mathrm{KF}_{wt}$ is v-improper, so the proposition is gappy). Thus 2. is not β-reducible to 1.[7]

In (Tichý 1982, 67), β-reduction and β-expansion are explained as deduction rules (The Rule of Contraction/Expansion). Tichý's rule of β-reduction contains an explicit condition that the construction C, which is substituted, is not v-improper. So conditioned, β-reduction preserves equivalence of constructions.[8]

Moreover, η-reduction (that $\lambda x[Cx]$ is reducible-equivalent to C) fails as well (Raclavský 2009, 283). Consider $[^0Fy]$, where ^0F v-constructs a function of type $((\theta\zeta)\xi)$ which is undefined for the ξ-object assigned to y by v. Thus $[^0Fy]$ is v-improper, it v-constructs nothing at all. But $\lambda x[[^0Fy]x]$ does v-construct an object, namely a function of type $(\theta\zeta)$ which is undefined for the ζ-object assigned to x by v. Hence $\lambda x[[^0Fy]x]$ cannot be equivalent to $[^0Fy]$. The remedy (*ibid.*) is the same as Tichý's conditioning of β-reduction.

[7] Here ≡ means inter-derivability of two constructions (*cf.* $\leftrightarrow_t$ in CP 489); the two constructions flanking ≡ v-construct one and the same object (a truth-value in this case) or they both v-construct nothing at all. The construction $[^0\mathrm{True}^{\varpi\mathrm{P}}_{w\not t}p]$ can be closed by $\lambda w\lambda t[\lambda p$ and then η-reduced to $^0\mathrm{True}^{\varpi\mathrm{P}}$ (which is used below). See (Raclavský 2008) for more.

[8] *Cf.* (Duží 2003) for more.

7 ANOTHER WAY OF REPAIRING PARTIALITY

When defining various concepts one sometimes needs to overcome partiality of some function. For the sake of illustration, imagine that you sum salaries of various people—including the king of France ("…+the salary of(KF)+…"). Since there is no king of France, "the salary of(KF)" returns no number. But you need a certain number (zero in this case) because you do not want the final sum ("…+…+…") to be undermined by the "local" partiality failure.

In Tichý's logic, the delivering of a "dummy value" (e.g. zero) can be easily managed in the following way (see Raclavský 2009, 243). Consider a partial function F from (type-theoretically appropriate) x's to (type-theoretically appropriate) y's. Since F is partial, $[^0Fx]$ can v-construct nothing; in such case, however, you need something—a dummy value. I suggest replacing $[^0Fx]$ in a construction C (which is affected by partiality of F) by the following construction which fulfils our demand:

$$\begin{array}{ll} [^0\mathrm{Sng}\lambda z[^0\mathrm{If}\_ & [^0\exists\lambda y[y^0=[^0Fx]]] \\ \quad \_\mathrm{Then}\_ & [^0\exists\lambda o[[o^0=[z^0=[^0Fx]]]^0\wedge[o^0={}^0\mathrm{T}]]] \\ \quad \_\mathrm{Else} & [z^0={}^0\mathrm{DummyValue}]]] \end{array}$$

(the singularization function, Sng^{ξ}, takes one-membered classes of ξ-objects to their sole members, it is undefined otherwise; if_then_else is the well-known ternary truth-function, its trivialization is written in parts; note that there is an analogue of "it is true$^{\pi\mathrm{T}}$ that $y=F(x)$" in the second line).

8 THREE-VALUED FUNCTIONS REPRESENTED BY PROCEDURES

To conclude this short paper, 2VP-logic incorporating procedures (constructions) is capable to capture the intuition which underlies 3V-logic, if 3V-functions are modelled not by (partial) functions but by procedures. For instance, the 3V-function T→F, F→T, U→U can be modelled by $[^0\neg p_{wt}]$ because this construction behaves in an analogous way as that 3V-function: it returns T when the proposition p has (in w, t) the truth-value F (and *vice versa*) but it returns nothing if the proposition p is undefined (in w, t). To define procedures representing other 3V-functions is usually more involved—one must utilize $^0\mathrm{True}^{\pi\mathrm{T}}$, often together with the dummy-value construction.

REFERENCES

Duží, M., 2003, Do we have to deal with partiality? *Miscellanea Logica* V, 45-76.
Raclavský, J., 2008, Explications of kinds of being true (in Czech). *SPFFBU* B 53(1), 89-99.
Raclavský, J., 2009, *Names and Descriptions: Logico-Semantical Investigations* (in Czech). Olomouc, Nakladatelství Olomouc.
Ruzsa, I., 1991, *Intensional Logic Revisited*. Budapest, published by the author.
Tichý, P., 1971, An approach to intensional analysis. *Noûs* 5(3), 273-297.[9]
Tichý, P., 1976, *Introduction to Intensional Logic*. Unpublished book manuscript.
Tichý, P., 1982, The foundations of partial type theory. *Reports on Mathematical Logic* 14, 57-72.
Tichý, P., 1986, Indiscernibility of identicals. *Studia Logica* 45(3), 257-273.
Tichý, P., 1988, *The Foundations of Frege's Logic*. Berlin, Walter de Gruyter.
Tichý, P., 2004, *Pavel Tichý's Collected Papers in Logic and Philosophy*. V. Svoboda, B. Jespersen, C. Cheyne (eds.), Dunedin, University of Otago Publisher / Prague, Filosofia.

[9] It seems that 2.4.2.5 in (Ruzsa 1989) captures an analogous restriction of β-reduction.

Márta Ujvári

Prior on Radical Coming-into-Being

Abstract. Chapter VIII of *Papers on Time and Tense* (1968) on 'Identifiable Individuals' is the locus where Prior elaborates his polemic on whether radical coming-into-being is a *genuine* de re *possibility* of the individual substances. He argues for the conclusion that there is no such possibility concerning radical coming-into-being as opposed to piecemeal change. Therefore the property or predicate about an individual's origin does not yield *de re* modality.

Prior considers the thought experiment of swapping the properties of two individuals through worlds. Say, we have in the actual world the person Julius Caesar with the usual Caesar-like properties. We have also Mark Antony as an inhabitant of the actual world with his own properties. Suppose that through chains of accessible worlds we reach a world such that Caesar has all and only the properties of Mark Antony and vice versa. Surely, this world is qualitatively indistinguishable from the original world. The thought experiment is typically explored for highlighting our intuitions whether the world resulting in complete property swap is different in any sense from the original world. Those who say that the two worlds are actually one, for all that matters only is what properties are instantiated and co-instantiated in a world and the instantiating entities are nothing over and above the properties they instantiate—are Leibnizians applying the criterion of L-indiscernibility. Others, who feel that there is a genuine difference between the two worlds, claim that Caesar and Antony still differ by their haecceistic properties of *being Caesar* and *being Antony*, respectively, whatever other properties of theirs are swapped. The usual retort from the Leibnizian camp is that haecceistic properties are not genuine properties.

The dispute about the identity condition of worlds and individuals cannot be settled in a short talk. What is interesting is how Prior finds his way out of the dilemma. He is neither Leibnizian nor haecceist. Moreover, he is not making the Kripkean point about the necessity of origin. Instead, he finds that the property which is necessarily exempt from property-swap is the property of originating from the actual ancestors one has. This property is resistant

to property exchange. If this is so, it helps to solve the problem of Leibniz-indiscernible worlds. And it has the virtue of not being a suspect haecceistic property, like that of being identical to Caesar; it is similar to it though in being an impure relational property.

After considering several property-swaps, Prior asks: 'can we not go further and suppose Caesar to have had the whole of Antony's life, including being born to Antony's parents?' (1968, 85.) The question is not about logical possibility since the proposition asserting Caesar's having born to Antony's parents is not inconsistent.

By contrast, construed as a question about a *temporal* possibility, we have a more substantial issue at hand. For now we can ask: '*when* was it possible?' And it is easy to see that '*after* his birth … it was clearly *too late* for him to have had different parents.' (ibid.)

This insight is fairly obvious. Why not ascribe, then, the *de re* possibility of having different parents '*before* Caesar existed'? Intuitively speaking, the remoter the present is, viewed from a distant future, the more possibilities are still open concerning that present (taken indexically). Alas, the crucial point, highlighted by Prior, is that however broad the general possibility is in this case, 'there would seem to have been no individual identifiable as Caesar, i.e. the Caesar who we are now discussing, who could have been the subject of this possibility.'(ibid.)

Now Prior's point can be generalized to the possibility of actual individuals, too. My contribution to the problem consists in this suggestion: let us extend Prior's point to the *de re* temporal possibility of the coming into being of *actual* individuals thereby bringing forth its full metaphysical significance. The generalization is this:

If Caesar (or any other actual individual) could not have been the subject, before his birth, of the (later) unrealized possibility, equally, *he* could not have been the subject of the *later realized* possibility either. This means that none of us who was going to be born could have been the subject of a *de re* possibility of being (going to be) born – *i.e.*, at least not before our conception. And this amounts to saying that what is once actual is always preceded by what is non-possible, contradicting thus the logic of propositional modalities. According to these modalities the actual is never preceded by something impossible. "Precedence" is taken here in the sense of yielding existential conditions through temporal priority.

So we have to rule out not only the unrealized *de re* possibilities concerning origins. The realized courses of originating by birth are not possible either, at least *not* in the *de re* sense. The reason being that there is no identifiable individual to ascribe the *de re* possibility to. Hence, there is no identifiable individual to ascribe the putative *de re* necessity of origin either. The realized and the unrealized *de re* possibilities/necessities are in a symmetric situation

relative to each other: both require the semantic precondition that the term in a referential position successfully refer, which is not fulfilled in either case. Let me note that cases of origin as temporal *de re* modalities are treated like cases of empty names in extensional contexts.

Prior rightly claims that any genuine *de re* possibility/necessity presupposes that the subject of the modal ascription, even if just a 'thin' individual, be fixed referentially. Since this is not satisfied in the case of the putative *de re* possibility of origin, the latter cannot be regarded as a case of *genuine de re* possibility. Further, I suspect that Prior is doing more than simply calling attention to the lack of a demonstrative reference when there is not yet anybody before birth to refer to. I take it that Prior is emphasizing the absence of a *descriptive* device of referring to individuals to consider *de re* possibilities. As Prior notes 'as new distinguishable individuals come into being, there is … a multiplication of distinguishable logical possibilities' (*op.cit.*, p.91.) Now it is fairly obvious that the basis of the distinction among individuals and their possibilities are of a qualitative nature. For example, there must be people identifiable qualitatively as different applicants to a job for opening the logical possibility of competition for that job. So I take it that Prior is implying some descriptive notion of individuals. It is a further question how 'thin' or 'thick' his implied notion of individuals is.

We have seen that the putative *de re* possibility of origin lacks the semantic precondition of reference. But then it is hard to see how one can satisfy the natural intuition that it is still meaningful to talk about the possibility of someone's having had a different origin.

The suggested solution is this: there is no subject, strictly speaking, of such *de re* possibilities. Our modal claims can be entertained only in the way that can be illustrated by Burleigh's example of a promise of giving someone a horse. Such promise may be disambiguated *not* as the promise of a specific horse but as the promise that can be fulfilled by giving someone *any* horse. So is the case with the possibility of origin. As Prior puts it, 'the possibility that *an* individual should begin to exist … is like a promise of the second kind'.(*op.cit.*, p.86.) In other words, this possibility is 'general' rather than specific. This is tantamount to saying that the possibility of origin is not a *de re* but a *de dicto* possibility. And this is the solution to the seemingly paradoxical situation. It is possible *that* someone be born to such and such parents, but it is not possible *of someone* that *he* should be born to these or other parents.

So, Prior's solution to the problem of *de re* possibility with radical coming-into-being consists in denying that there is such *de re* possibility and satisfying, instead, our modal intuition with the *de dicto* form of possibility. *De re* possibility with respect to radical coming-into-being is, at best, a *post factum* possibility: as Thomas Aquinas put it, quoted by Prior, it is an 'accident' which 'is subsequent to the thing' that has already come into existence.

The ramification of the topic worth considering is the issue of temporal vs logical possibilities. The temporal possibility just discussed is obviously different from possibility in the logical sense. In the logical sense of 'could' Caesar could have been born to persons who actually turned out to be the parents of Mark Antony for there is no inconsistency in the proposition stating this course of events. (Leibnizians would, perhaps, object that there is, for the 'complete concept' of Caesar excludes the relational property of being born to those other people.) Logical possibility seems then to be permissive about a case explicitly ruled out by temporal possibility.

The gap between the two kinds of possibilities is smaller though: for, both possibilities as *de re* possibilities have an existential precondition such that 'before [Caesar] existed it was not logically or in any other way possible that he should *come to have* those people, or any other people, as his parents'. (*op.cit.*, p.92.) As possibilities dependent on an (empirical) existential precondition, the logical and the temporal readings of possibility behave much the same way.

Another ramification would be to compare radical coming-into-being with piecemeal change. It is clear from the foregoing that radical coming-into-being, in Prior's view, is not change in the piecemeal sense of an existing thing acquiring/dropping properties over time. Coming-into-being should not be taken, as he points out, to mean that 'once *X*'s non-being was the case and now its being is'. It should be taken to mean, instead, that 'it is *not* the case that X *was*, but it *is* the case that X *is*', and this does not express a change but two contrasting present facts'. (*op.cit.*, p.88.) Clearly, radical coming-into-being does not have the features of a genuine change. Radical coming-into-being is not unique, however, in this respect: Cambridge changes are typically not taken to be genuine changes either. Elsewhere I have discussed them.

Conclusion: the property of origin should be exempt from the range of properties affording *de re* locutions. This logical insight is backed by the metaphysical insight that radical coming-into-being does not constitute genuine change. The common source of these insights is that the property of origin (unlike other properties of individuals) lacks the existential precondition. *After* this precondition being fulfilled, the property of origin is an accident subsequent to the individual. As a consequence, the property swaps between individuals through worlds must stop at the impure relational property of origin.

The broader metaphysical moral is this: we have seen the limits of *de re* talk drawn by the property of origin or radical coming-into-being. It is all too familiar that *de re* locutions are found suspect from Quine on, for the weird metaphysics one gets when one quantifies in modal contexts. The present talk is obviously not a follow up on Quine' discontent with the 'invidious attitude' (as he calls it)

of Aristotelian essentalism.[1] Rather, by drawing the lines of applicability one can assess the merits of the metaphysical doctrines. As to the significance of Prior's argument, we have seen that it offers a genuine novel suggestion to the polemic about property-indiscernible worlds. He does not seek haecceistic differences in order to explain the difference of Leibniz-indiscernible worlds. Rather, Prior shows that there is a further difference which is responsible for the difference between any two qualitatively indiscernible worlds: it is the *difference of origin* of the individual inhabitants of the worlds. To repeat, the *property of origin* is neither purely qualitative, nor haecceistic, but a peculiar accident that can be had only *post factum* and it is resistant to property-swap.

REFERENCES

Prior, Arthur, 1968, *Papers on Time and Tense*. Oxford, Clarendon Press.
Quine, W.O, 1953/1963, Reference and Modality. In id., *From a Logical Point of View*. Cambridge, Mass., Harvard University Press.

[1] Quine (1953) characterizes 'Aristotelian essentialism' by its 'adopting an invidious attitude towards certain ways of uniquely specifying' an object and seeing these ways 'as somehow better revealing [the object's] "essence"' than alternative specifications.

László Kálmán

Analogy in Semantics

Abstract. The principle of compositionality seems too trivial and too restrictive at the same time. I propose that this principle should be seen as (part of) a definition of what "meaning" is in a theory that posits a very abstract concept of "meaning", one very far from empirically testable reality. The principle of compositionality presupposes an analytic/atomistic approach to meaning, where the properties of a whole must be explained in terms of those of its constituent parts. In my paper, I will propose a *holistic* approach instead: I will emphasise the global features of signs. I will introduce the principle of *generalized compositionality*, which is based on the concept of similarities between forms and meanings. My generalized compositionality principle states that we interpret and produce complex signs by analogy, relying on our earlier experience on similar complex signs and their interpretation. This move, I believe, is necessary for moving towards a cognitively more realistic model, with a view to predicting frequency effects and other psychological factors of interpretation. As a side-effect, the dubious distinction between "literal" and "non-literal" interpretations disappears.

INTRODUCTION

In this paper I will argue for a model of natural-language semantics that diverges in important ways from the main-stream approach, but it fits well within a more general cognitive framework in which language is viewed as a system of habits, in much the same sense as it was viewed in the late 1800s and in European structuralism.

I will start from two different families of problems, namely, the distinction between "literal" meanings and actual interpretations, e.g., in the case of metonymical language use (section *1.1*), and an ambiguity related to disjunctions used in negative contexts (section *1.2*). I will show that both of these types of

problems pose serious challenges for the traditional machinery of semantics that posits a compositional "translation" process that produces logical formulae.

I then turn to the formulation of and the inherent problems with the principle of compositionality (section *2.1*), and I will propose a different principle, which I call *generalized compositionality*, which avoids the pitfalls of traditional compositionality (section *2.2*). I will argue that this alternative formulation can serve as the fundamental principle of *analogy-based semantics*, a view of semantics that does not require "underlying", abstract syntactic representations or a distinction between "literal" and "non-literal" interpretation. I will briefly return to the linguistic problems that constitute the starting-point of the discussion, and outline ways of approaching them in analogy-based semantics. I conclude, in section *2.6*, by sketching a model of associative memory that can serve as an implementation of the machinery of analogy-based linguistic theory.

1 TWO PROBLEMS IN LINGUISTIC SEMANTICS

1.1 Non-literal meaning and abduction

As pointed out in (Hobbs *& al.* 1993), even the simples sentences, like his *The Boston office called*, contain *implicit information*, i.e., propositions that are *plausible* for the audience to assume, and without which we cannot speak of making sense of, interpreting, or understanding what the sentence "literally" says. In this sentence, for example, the audience has to figure out (among other things) that *the Boston office* is an office located in Boston, and that someone working at that office (rather than the office itself) placed the call. (Hobbs *& al.* 1993) propose to consider this mechanism a case of abduction, i.e., to consider the "literal" interpretation of the sentence as a conclusion, and the plausible assumptions the audience has to take on as missing premises, which would then make the conclusion true.

Unfortunately, the two key terms that I have just employed, *plausibility* and *literalness* are very hard to grasp, and there is hardly a consensus surrounding them. For example, it is not clear why 'someone working at the Boston office called' is more plausible than, say, 'someone who happened to be at the Boston office called'; or, to take another example, it is not clear what the "literal meaning" of a compound like *Boston office* is: 'an office somehow related to Boston' (as (Hobbs *& al.* 1993) assumes, in which case we have to rely on abduction again to get the metonymical interpretation) or 'an office located in Boston', and it is also not clear whether this expression is ambiguous, i.e., whether it has more than one "literal meaning".

From the linguistic point of view, the approach taken in (Hobbs *& al.* 1993) illustrates nicely the paradigm going back at least to (Grice 1967), according to which utterances can be assigned context-independent, purely linguistic,

"literal" *meanings*, which leave much uncertainty as to the appropriate interpretation in a given context. In Grice's (1967) view, it is the anomalies and contradictions arising from this basic "meaning" that give rise to all sorts of implicit information that the audience tends to assume, in order to resolve the anomalies, and thereby arrive at a contextually determined "non-literal" interpretation. The process often involves ontological (real-world) knowledge and pragmatic considerations. In (Hobbs *& al.* 1993), no anomaly is needed to trigger this process: implicit information is always retrieved until a *maximal consistent* stock of background knowledge is incorporated into the interpretation through abduction (by assuming only *maximally plausible* pieces of information).

This model of understanding heavily relies on a dichotomy between "linguistic knowledge" (driving the translation mechanism that produces a "meaning" from a linguistic expression), on the one hand, and "background knowledge", from which plausible assumptions are retrieved by the abduction process. But this dichotomy is disputable. For example, in the case of *Boston office*, the abduction process must use an axiom in the knowledge base according to which "$\texttt{located\_at}(y, x) \rightarrow \texttt{rel}(x, y)$", where the translation of *Boston office* contains "$\texttt{office}(x) \wedge \texttt{boston}(y) \wedge \texttt{rel}(x, y)$", so "$x$ is located at y" is a plausible premise for x standing in the relation "$\texttt{rel}$" with y, and the conclusion is part of the "literal meaning" of *Boston office*. But it is not at all clear why just the use of the mysterious relation"$\texttt{rel}$" is part of the "linguistic" module, whereas possible interpretations of such compounds are located in the "background knowledge", although they are clearly of a linguistic character, too (as shown by the fact that the same type of compounds cannot be interpreted in this way in Hungarian: there, a suffix must be added to *Boston* in order for it to refer to the location of the office).

In general, there is no universal recipe for deciding whether a certain ingredient of understanding (in our example, 'the office is located in Boston') should stem from "linguistic knowledge" (in our example, this would amount to a particular "compounding rule" to this effect), or from "background knowledge" (in our example, this corresponds to the approach in Hobbs *& al.* 1993). There is no accepted philosophical, linguistic or psychological evidence for deciding either way for any particular case.

The shakiness of the distinction between "linguistic" and "background" knowledge is not a purely theoretical problem. In particular, it is challenging for any potential *mental model* of the mechanism of understanding, in particular a model of *acquisition*, on the one hand, and a model capable of predicting *performance data*, on the other. It is not obvious, to say the least, how the distinction between "linguistic" and "background" knowledge could be learned, and I do not believe there is any psycholinguistic evidence to the effect that the "translation" of an expression such as *the Boston office* (i.e., retrieving its

"literal meaning") should be separable from the process of interpreting it using background knowledge.

1.2 Disjunction of NP's

Let me now turn to a completely different problem, or rather a family of problems, unrelated at first sight to those explained in *1.1*. These problems, which have puzzled semanticists for a long time (e.g., Szabolcsi 2002; Szabolcsi & *al.* 2004), are connected to the interpretation of *disjunction* in various embedded positions. For example, consider the following ambiguity:

(1) *He didn't close the door or the window.*
   a. 'He left both the door and the window open'
   b. 'He left either the door or the window (or both) open'

Szabolcsi (2002) explains the ambiguity in an essentially syntactic way (although she relies on features that can be viewed as having semantic content). She claims that, under one interpretation, negation "has a wider scope than" disjunction, whereas the opposite holds when the other reading is to be obtained:

(1′) a. ¬('closed the door' ∨ 'closed the window')
   b. (¬'closed the door' ∨ ¬'closed the window')

Note that in order to obtain these readings, highly unnatural (abstract, covert) syntactic structures have to be assumed: both of these formulae are very distant from the surface structure of the sentence in (1). Thus, this mechanism requires a non-deterministic "translation" step from any theory that respects compositionality.

The situation is even worse when negation is not explicit, but inherent in the predicate, as in the following example:

(2) *You forgot to close the door or the window.*
   a. ¬('remembered to close the door' ∨
      'remembered to close the window')
   b. (¬'remembered to close the door' ∨
      ¬'remembered to close the window')

In order to get the interaction of negation and disjunction here, one has to assume a representation in which the negation inherent in *forget* is made explicit by converting it into *not remember*.

While the problem explained in *1.1* allegedly concerns how we arrive at the desired interpretation from the "literal", "linguistic" meaning, the problem of interpreting disjunctions seems related to the "translation" itself, i.e., to how we get from a surface utterance to the "literal", "linguistic" meaning.

Under the standard approach, in order to account for ambiguities like the one presented here, we must assign two different logical formulae to one and the same utterance, depending on factors that are not entirely clear.

In both cases, there is a gap between a surface expression and its interpretation, i.e., both problems challenge the traditional, *compositional* approach to semantics, which consists in assigning a logical representation to utterances in a systematic way, then interpreting that representation with respect to a model of the world. It looks like we have to revise the concept of compositionality in the sense it is normally used while, for obvious reasons, we do not want to deny the systematic character of interpreting natural-language utterances.

2 GENERALIZING COMPOSITIONALITY

2.1 Problems with compositionality

The principle of compositionality is usually stated as follows:

(3) *The principle of compositionality*
The meaning of a complex expression is a function of the meanings of the constituents which syntactically constitute it.

This principle, attributed to Frege, is considered the starting-point of modern semantics, aiming at establishing and examining the systematic relationship between form and meaning in natural language. It is commonly thought of as saying something fundamental (in fact, almost trivial) about meanings and the science of meanings. On the other hand, it has puzzled most modern semanticists in various ways because of the multitude of phenomena that it does not naturally apply to. To mention a couple of those, there is the context dependence of the use of many, if not all, linguistic expressions (so the context of utterance can be seen as an additional factor, not mentioned in the principle of compositionality, potentially influencing the possible use of a complex expression), and there are legions of obviously complex expressions in all natural languages the meaning of which is "non-compositional" or idiomatic to a smaller or greater extent (which means that their meanings are more or less unrelated to the meanings of their constituent parts).

As for context dependence, the received view seems to be that the principle of compositionality is a methodological axiom that demarcates what is covered by part of the definition of meaning exactly by eliminating those properties of linguistic use that depend on contextual information. That is, context dependence does not contradict compositionality; to the contrary, compositionality carves out just those aspects of the use of linguistic signs that are independent

of context,[1] and constrains the term *meaning* by excluding all other aspects (and tacitly relegating them to some other field of study, say, "pragmatics").

The problem of those expressions that are idiomatic to some extent is a much tougher one. We could approach this problem in the same vein as context dependence, by saying that the principle of compositionality limits semantic study to the so-called "compositional" (non-idiomatic) constructions, but it is not at all clear where the boundaries would lie, and what such a semantics could do with mixed cases when a complex expression has both idiomatic and non-idiomatic aspects. For example, an idiom like *spill the beans* has the same aspectual and thematic features as its main verb *spill* does (it is agentive and expresses an accomplishment or an achievement), so it is "transparent" in this respect. (As a matter of fact, the view that there is a concept of "meaning" that abstracts away all context dependence, can be challenged in a similar vein, but I will not go into that discussion here.)

So buying into the principle of compositionality seems to represent an important sacrifice for a truly scientific approach to meanings. On the other hand, it can be shown that it is ridiculously easy to still satisfy the principle, given the fact that the concepts of *meaning*, *syntactic constitution* and *function* figure in it without any further specification or constraint. The principle imposes no constraint on ambiguities (of the expressions constituting a complex one) and on the process on their resolution. For example, take the understanding of *The Boston office called:* the principle of compositionality does not constrain how *office* and/or *call* are to be disambiguated. For example, it is not a violation of compositionality if one uses "global" information on what the whole utterance (or even the text that it occurs in) is about, or what the speaker's intentions are. This means that, effectively, there can be wildly non-compositional steps in a strictly speaking compositional interpretation process.

So both "non-compositional" constructions and context dependence can be "explained away" by appealing to the presence of multiple (possibly infinite) ambiguities and obscure mechanisms for their resolution. Meanings can be of any nature whatsoever, so it is possible to encode not only semantic, but also formal properties in them. As an extreme case, we can posit that the "meanings" of constituent expressions specify their interpretation for each (type of) complex expression they occur in, which is compatible with the definition, yet it would make compositionality entirely empty. As for the concept of "function", a function can do just about anything and, by taking this to the extreme, one can arrive at wildly unnatural, yet compositional mechanisms of interpretation—like ignoring the meaning of a constituent; for other examples, see, e.g., (Zadrozny

[1]Sometimes context still plays a role, when reference to context is said to be built into "meanings". For example, the fact that the personal pronoun *I* refers to the speaker is said to be part of its "meaning", while the context determines who the speaker actually is. Cf. (Gendler Szabó 2000) for a summary.

1994)). Finally, the principle of compositionality contains no proviso to the effect that syntactic structure be determined independently of semantics. As a matter of fact, most syntactic constituency tests (such as substitutability or mutual information) are implicitly based on the greater semantic cohesion between certain structural elements than that between others within a construction. So there is ample space for "playing around" with syntactic structure in order to satisfy compositionality.

2.2 Generalized compositionality

As I have pointed out in the previous section, the principle of compositionality is both too strong and too weak, too restrictive and too liberal at the same time. I have also mentioned that, still, it is essential to adopt it in some form or another, inasmuch as it captures one of the leading ideas of the linguistics of the late 19th century, namely, that the productive aspect of natural languages is due to the *systematicity* of our ability to understand utterances. That is, we can only understand an utterance if it is built up in a similar way to utterances we have seen earlier, and we interpret it by *analogy* to how those earlier utterances were used. By the same token, we are able to communicate through utterances if we utter ones that are sufficiently similar to those the audience is familiar with, and we have to take into account that they will be interpreted in the same way as the audience's earlier experiences dictate.

The key concepts here are *recurrence* and *similarity* rather than "meaning", "function" and "constituency", but the basic idea, I believe, is the same as in the case of the principle of compositionality. Not only elementary signs, but also their combinations must show a sufficient amount of similarity to combinations seen earlier in order for them to be understandable, which means that the way they are combined must also be recurrent, and their interpretation is to be calculated *mutatis mutandis*, by a possible recombination of signs and associated uses seen earlier. This corresponds to the emphasis put by the compositionality principle on the role of syntactic combination.

However, the reformulation I am about to propose also differs from the compositionality principle in important respects. Both similarity and recurrence are *gradual* concepts, because one expression or meaning can be similar to another not only in various aspects, but also to varying degrees, and the recurrence of a pattern (the frequency of its previous occurrences or the strength of the memories we have of them) also comes in degrees. The linguistics of the mid-20th century was reluctant to use models involving gradualness, but since many human systems, especially mental ones, clearly show gradualness effects, it will be hard to avoid appropriate stochastic models in the long run. So my generalized principle can be formulated along the following lines:

(4) *The principle of generalized compositionality*
To achieve maximal understanding, if we want to express an idea I, then we had better use an expression E that is maximally similar to the most frequent expressions E' which, to our knowledge, express ideas I' maximally similar to I. Also, this is the strategy that a hearer assumes other speakers to use.

The reason why I dare call this principle a generalization of compositionality is that it also expresses that different forms that show some parallelism are interpreted in parallel ways. Under traditional compositionality, formal parallelism is restricted to constituent structure, and semantic parallelism is restricted to the identity of the functions that combine the constituents' meanings. Thus my reformulation of the compositionality principle, in addition to the gradualness inherent in it, also crucially lacks a reference to syntactic constituency. For example, in the example I have examined earlier, it is crucial that *Boston office* is similar to expressions frequently heard earlier, in the sense that, in those expressions, the first word was similar to *Boston* (not in the phonological sense, but in terms of being names of places) and the second was similar to *office* (again, not in terms of their form, but their function), and those expressions were interpreted analogously to *Boston office*. These similarities involve "constituents" (it does matter that *Boston office* is built up of two sub-expressions in a particular linear order), but the similarity between "constituent structures" is just one particular case of many types of similarity. For example, similarities between word forms (i.e., the morphological built-up of words) is often not restricted to linear, concatenative similarity. As a matter of fact, as we have shown elsewhere (Kálmán *& al.* 2005), it would be wrong to consider linear and concatenative morphology as "normal" and qualify all other cases as deviant.

2.3 Analogy-based semantics

My claim is that the principle of generalized compositionality in (4) not only amends the weaknesses inherent in the traditional concept of compositionality, but also solves problems like those presented in *1*. That is, it offers an alternative, more appealing approach to how linguistic and background knowledge contribute to possible interpretations, and to how ambiguities arise in various contexts.

To explore this alternative, I will assume that, as long as we lack evidence to the contrary, there is a single stock of knowledge for "linguistic" and "background" knowledge, i.e., an ontology that comprises knowledge about both linguistic and extra-linguistic entities. Syntactic structures are based on surface utterances, no "underlying" or "abstract" structure is posited. No "translation" occurs at all; the structure of an utterance plays a role only inasmuch as it expresses the formal *similarity* of one utterance to the other.

Instead of translating expressions into a logical language, we just retrieve information associated with the linguistic entities perceived, be it linguistic or extra-linguistic in character. Thus, information, linguistic and non-linguistic alike, is activated through *association*. In the examples quoted in *1.1*, the relevant information comprises everything related to *Boston* and *office* and their syntactic arrangement (i.e., this particular type of compound). Those pieces of information get activated, together with all the information originating from the rest of the utterance and the utterance context. In the case of the problem in *1.2*, the relevant information that gets invoked includes the formal and functional aspects of other instances of clauses embedded in (negative) predicates, other instances of NP disjunctions, etc.

2.4 The abduction process

Independently of whether we assume two separate modules like (Hobbs *& al.* 1993) or a unique network-like knowledge base, the abduction process must be an optimization of continuous variables, since plausibility is gradual. The task is a quite complex one: as I have pointed out earlier, we have to retrieve a *maximal consistent set* of *maximally plausible* abduced premises for a set of conclusions. For (Hobbs *& al.* 1993), the conclusions are constituted by the "translation" of the input utterance; in my view, they must consist of all we might consider *empirically maximally certain*, i.e., what has been perceived, in whatever way we model that. For the example of *Boston office*, this means that we can only take it for certain that the words *Boston* and *office* have been uttered in this order (plus whatever contextual information we want to take into account).

To be sure, plausibility must rest on probability or frequency: the more probable or frequent a state of affairs is, the more plausible it is to assume it (at least with a certain probability) given that we have some evidence for the truth of something that follows from it (again, the probability that this evidence is reliable may vary). Our aim is to maximize the joint probability of the entire set of propositions, perceived and abduced, taken together. We also have to take into account the *synergy* of perceived facts and plausible assumptions, in the sense that taking the joint probability of several antecedents taken together may yield a different result from just looking at their probabilities one by one. For example, take our example *The Boston office called*. Hobbs *& al.* (1993) argue that the most plausible assumption to make is 'someone working at the Boston office placed the call', because only humans can make telephone calls, and relating a human to an office can be done most plausibly by assuming that the person works there (how exactly this is achieved by (Hobbs *& al.* 1993) is irrelevant here). However, *call* is itself polysemous (animals or even machines can call as well). On the other hand, even if calling was uniquely human, this would not ensure for the above abduction to be correct: *The Boston office is on holiday* would

not be interpreted (if it is interpretable at all) as 'someone working at the Boston office is on holiday' but rather as 'everyone working at the Boston office is on holiday'. It is only the synergy of all more or less plausible premises, i.e., their high joint probability, that makes the abduction suggested in (Hobbs *& al.* 1993) really plausible.[2]

These conditions suggest a solution in the spirit of *Bayesian networks*—e.g., (Pearl 1985) or *Markov logic networks* (Richardson *& al.* 2006). Calculations using both of these models are extremely complex, which suggests that *massively parallel* computational mechanisms are required in order to deal with them with reasonable resources. The model that I will propose shortly, in *2.6,* is intended to solve this.

2.5 Accounting for NP disjunction

In accordance with the alternative approach proposed here, when interpreting sentences like (1) in *1.2*, we do not depart from their genuine (surface) syntactic structure, and try to derive their interpretations from what the sentence actually looks like. In particular, we must take into account how sentences with a negative content and, in particular, the arguments of negative predicates are interpreted in other cases. On the other hand, we also have to look at how the disjunction and, in general, the co-ordination of NP's is interpreted in other cases. Interpretation has to proceed by analogy to all those other cases.

This approach has two obvious consequences which I think are desirable. First, it predicts that uncertainty may arise in the interpretation of the utterances in question if there are analogous structures with conflicting interpretations. Such uncertainties are hard to explain if one assumes a compositional semantics producing a logical form, given the fact that, informally speaking, we are dealing with the simplest syntactic structures and the simplest logical connectives. Second, it predicts that the interpretation of such utterances may show peculiar differences from one language to the other, even if independent arguments for structural differences would be hard to find.

In sum, I will assume that all the examples discussed in *1.2* and here contain disjunctions of noun phrases (i.e., we must not convert the disjunctions into sentential ones), embedded into (explicit or implicit) negation (i.e., there is no "scope ambiguity" involved). The idea is somewhat similar to Scha's (1981), who argues against reducing plural or co-ordinated arguments to quantificational structures, and for conceiving of them as marking *predication about collections*. That is, when using a disjunction of noun phrases, one talks about collections the

[2]There are also additional propositions to be abduced that (Hobbs *& al.* 1993) fails to mention. For example, the person placing the call is not just anyone working there, but one who is somehow entitled to represent the Boston office; the office must belong to a company uniquely identifiable as familiar to both the speaker and the audience; and so on.

members of which are "alternatives" of each other (in some sense of the word), while the conjunction of noun phrases refers to collections that the predicate applies to "jointly" (in some sense of the word).

It is well-known that the exact interpretation of predicating about collections is, to a large extent, underdetermined by the linguistic structure used. One consequence of this is that ample space is left for contextual influences and uncertainty. In the case of "conjunctive" collections, possible interpretations include collective, distributive and cumulative readings (cf. Scha 1981); the possible interpretations of "disjunctive" ones are different: they express "choice" in a largely underspecified sense (including free choice, uncertainty and variation).

Obviously, the preferred interpretation of such an underspecified interpretation instruction may vary depending on what the disjunctive coordination is embedded into. For example, consider:

(5) *I closed the door or the window.*
   a. 'I (always) closed some opening, I forgot/it does not matter if it was the door or the window'
   b. 'Sometimes I closed the door, sometimes the window'

(6) *Close the door or the window.*
   a. 'Close that opening, whether it is the door or the window'
   b. 'Close either the door or the window'
   c. 'Always close an opening, sometimes the door, sometimes the window'

The same effect underlies, I believe, the well-known puzzle, originally noted in (Kamp 1973), which involves disjunction embedded in a permission predicate:

(7) a. *You can have soup or cookies.*
   '(7a)' $\models$ 'you can have soup'
   '(7a)' $\models$ 'you can have cookies'
   b. *You can have soup.*
   '(7b)' $\not\models$ 'you can have soup or cookies'

Although Fox (2007) and others have argued that the entailments in (7a) are in fact conversational implicatures, they are not cancellable as conversational implicatures are supposed to be, which makes this type of analyses untenable. Instead, we must explain them by looking at similar structures, and reconstruct what sense people can make out of a disjunctive collection in a permission context. It is easy to see that, in this special type of contexts, predication about such disjunctive collections must be interpreted in nearly the same way as if they were conjunctive (although exclusiveness may be understood):

(7′) *You can have soup or cookies.*
   'You can have whichever you want (but not both)'

As a matter of course, this sentence can also be interpreted by taking the disjunction to indicate uncertainty ('You can have food, but I don't know which kind is available'), it is simply less probable or plausible for real-world reasons.

As opposed to the permission context, the negative context easily yields both types of readings. The reading in (1′a), with "wide-scope negation", corresponds to the free-choice reading, whereas the reading in (1) corresponds to uncertainty:

(1) *He didn't close the door or the window.*
   a. 'He didn't close them; whichever you pick, it is true that he didn't close it'
   b. 'He didn't close one of them; I don't know which (or maybe both)'

(Szabolcsi 2002) claims that the Hungarian counterpart of (1) is less ambiguous than the English sentence. In particular, while reading (a) is preferred in English (if there is a preferred reading), reading (b) seems preferred in Hungarian. My own informal survey shows that, in fact, both readings exist in Hungarian, too, and they are more or less prominent depending on the context. (Note the difficulty of capturing how contextual factors can help disambiguation if the different readings are due to different syntactic structures.) For example, consider:

(8) *Nem csuktad be az ajtót vagy az ablakot.*
not closed-you in the door-acc or the window-acc
'You didn't close the door or the window'
   a. When explaining the source of the draught:
   'you left both the door and the window open'
   b. When explaining the source of the cold temperature:
   'you left the door or the window (or both) open'

In (a), the preferred interpretation is like in (1′a), because draught is due to two things open simultaneously, whereas (b) is more likely to be interpreted as (1′b), because one opening is sufficient for explaining the cold temperature.

If the difference between English and Hungarian is not as prominent as (Szabolcsi 2002) claims, then the syntactically grounded explanation faces a serious problem, because slight biases or weak tendencies are hard to account for in a theory of syntax like transformational generative grammar. Obviously, since we all believe that natural languages are systematic, it is essential that we look for parallel phenomena that can be related to the interpretation of disjunctions in negative contexts.

One obvious candidate for such a parallel phenomenon is the behaviour of *indefinites* in negative contexts since, in the logical sense, indefinites can be considered abbreviations of potentially infinite disjunctions. Indefinites can be interpreted in pretty much the same way as disjunctive collections:

(9) a. *He closed a window.*
b. *You can have a cookie.*
c. *I bought a car.*

The interpretation of "choice" and/or "uncertainty" varies from one sentence to the other: a genuine free-choice reading is only available in (9b); in the case of the indefinite in (9a), the "choice" and the "uncertainty" interpretations are almost undistinguishable. The sentence in (9c) seems to have an "uncertainty" reading only, namely, the addressee reading, there is no "choice" whatsoever involved, only uncertainty, namely, the addressee cannot be certain what car the speaker bought.

When an indefinite occurs in a negative context, it usually has both a "choice" and an "uncertainty" interpretation:

(10) *I didn't buy a car.*
a. 'I bought no car' (choice)
b. 'There is some car (it does not matter/you may not know which) I didn't buy' (uncertainty)

But English and Hungarian sharply differ in this respect:

(11) *I didn't buy a car.*

(12) a. # *Nem vettem egy autót.*
not bought-I a car-acc
b. OK*Nem vettem autót.*
not bought-I car-acc
'I didn't by a car (= I bought no car)'
c. OK*Nem vettem meg egy autót.*
not bought-I pref a car-acc
'I didn't by one of the cars, there is a car I didn't buy'

That is, in Hungarian the "choice" interpretation of indefinites within a negative context can only be achieved by using *bare nominals* like *autót* 'car-acc' in (12b); a genuine indefinite only tolerates the negative context in an "uncertainty" reading, as in (12c), and only if a prefixed version of the verb is available (in this case, *meg* + *vettem* 'pref-bought-I' is the prefixed version of *vettem* 'bought-I', and the prefix is relegated to a post-verbal position because of the presence of negation).

Under an analysis like Szabolcsi's (2002), indefinites and disjunctions share the property that they avoid "having smaller scope than negation" (except maybe certain circumstances). But such an analysis would predict, e.g., that (8a) should be as bad as (12a), which is far from true. On the other hand, if one interprets the relevant sentences based on analogies with other structures, then

the argument goes as follows. In English, free-choice indefinites in negative contexts are highly frequent (cf. (10a)), whereas in Hungarian they are much rarer (given the unacceptability of sentences like (12a)). This predicts a stronger tendency for a similar distribution of "choice" vs. "uncertainty" readings in general in the respective languages, including the interpretation of predication about disjunctive collections. The effect on the readings of disjunctions can well be a slight one, which cannot be obtained in a theory that attributes the difference to features that trigger or block various transformations. As a matter of fact, one could even give an estimation of the degree of these tendencies if one had a sufficiently rich semantically annotated corpus.

2.6 Machinery: An associative memory model

As I have suggested in section *2.4*, that solution to the logical problem outlined in *1.1* requires a formal model of *association* capable of embodying parallel stochastic algorithms. Such networks are well-known in the history of artificial intelligence, from the semantic nets of (Schank *&al.* 1969; Schank 1975, 1982) through Bayesian networks (e.g., Pearl 1985) to the countless versions of neural networks (for an overview see, e.g., Arbib 1995). It is not clear, however, how the type of analogical reasoning that I have outlined in the previous sections can be implemented in such systems. Purely symbolic networks such as Schank's can model associations between concepts, but they do not make it possible to decompose and create new combinations of concepts by analogy of combinations seen earlier; purely neural networks, on the other hand, which do not contain explicit representations of concepts, are ill-understood, and it is not clear whether and how they can solve a particular logical problem like the one I am concentrating on.

For this reason, I am proposing a *hybrid* conceptual/neural network, which stores *aspects* or *properties* of concepts/experiences/memories, together with their *probability* (depending on how frequently they have been activated, i.e., observed or used). Novel combinations (new "concepts", so to say) can arise in such a network, because we can interpret the simultaneous activation of properties as their recombinations. (For the problems discussed in this paper, the relevant properties include uses of place names and *office*, properties of compound nominals in English, negative predication, predicating about collections, "disjunctive" and "conjunctive" collections, and so on.) Since we intend to model association, the algorithm that operates on the network is *activation spreading*.

Obviously, the system is capable of *learning* by incorporating new properties (never observed earlier) and by updating the frequency of a property whenever it gets activated (for any reason, either by direct observation or by internal use). The main difference between this system and predecessors is that the *connections*, i.e., associations between properties, do not ever change: neither

new connections nor "connection weights" can be introduced (except when a new property gets incorporated in the network). Associations only exist between more and less *general* properties, just like in a classical *generalization network* (e.g., Hispanicus 1947; Levinson 1996), plus *inhibitory links* exist between more or less incompatible properties. The role of probability (frequency) is that the spreading of activation (and inhibition) is stronger from more frequent nodes. Crucially, activation also spreads more efficiently (with less loss) from more specific to more general nodes than the other way round. This corresponds to the idea that the decomposition of a more concrete experience (i.e., its association with its properties) is quasi-automatic, whereas the association from properties to the more concrete experiences that exhibit that property is less fluent.

The memory model outlined here is explained in detail in (Kálmán 2010). It has been successfully tested on relatively simple examples in morphology and syntax, but not yet in the much more complex realm of semantics. Therefore, for the moment, I can only speculate on how analogy can be made to work in the domain of semantics, as I have explained in sections *2.4* and *2.5*.

3 SUMMARY

We have witnessed various problems with the dominant view of understanding. This view posits a level of "literal" or "linguistic meaning", which is then supplemented by extra assumptions made by the audience using their "background knowledge". But the borderline between "linguistic" and "background" knowledge is questionable. On the other hand, the principle of compositionality, which underlies modern semantics in general, and formal semantic theory in particular, also posits a level of "linguistic", "context-independent meanings", tacitly relegating all other information used in the production and understanding of linguistic signs to other modules (namely, "pragmatics"). So I had to reformulate this principle in order to bring it into harmony with the homogeneous view of understanding that I advocate.

The reformulation of the compositionality principle (which I have dubbed *generalized compositionality*), is based on a general principle of human linguistic communication, namely, *systematicity*, which means that similar forms tend to be associated with similar functions, and vice versa. (It should be noted that, as a consequence of this general principle, linguistic systematicity also means that different forms tend to be associated with different functions, and vice versa.) Since, as a consequence, the principle of generalized compositionality relies only on similarity and recurrence, it incorporates a *holistic* approach to linguistic signs: as opposed to the traditional, *atomistic* view, which aims at deriving all properties of complex signs from the properties of their constituent parts, the holistic view emphasises global similarities which, in principle, need not always be defined

recursively, as a similarity of "constituent structure" and the similarity of the corresponding "constituents".[1]

REFERENCES

Arbib, M. A. (ed.), 1995, *The Handbook of Brain Theory and Neural Networks*, Cambridge (MA), MIT Press.

Fox, D., 2007, Free choice and the theory of scalar implicatures. In U. Sauerland and P. Stateva eds.), *Presupposition and implicature in compositional semantics*. Houndmills, Basingstoke, Palgrave Macmillan, 71–120.

Gendler Szabó, Z., 2000, *Problems of compositionality*. New York, Garland.

Grice, P. 1967, William James lectures on logic and conversation. In D. Davidson and G. Harman (eds.), *The Logic of Grammar*, Encino, Dickenson, 64–75.

Petrus Hispanicus, 1947, *Summulæ logicales*. Rome. (Edited by I. M. Bocheski; original from c. 1239)

Hobbs, J. R., Stickel, M. E., Appelt, D. E. and Martin, P. A., 1993, Interpretation as abduction. *Artificial Intelligence* 63 (1–2), 69–142.

Kálmán, L., forthcoming, Analogical reasoning using an associative memory model. Manuscript.

Kálmán, L., Rebrus, P. and Törkenczy, M., 2005, Hungarian linking vowels: An analogy-based approach. Unpublished manusript. 2nd Old World Conference in Phonology (OCP2), University of Tromso, Center for Advanced Study in Theoretical Linguistics (CASTL), Tromso, Norway.

Kamp, H., 1973, Free choice permission. *Proceedings of the Aristotelian Society* 74, 57–74.

Levinson, R. A., 1996, General game-playing and reinforcement learning. *Computational Intelligence* 12, 155–176.

Pearl, J., 1985, *Bayesian networks: A model of self-activated memory for evidential reasoning*. Technical Report CSD-850017, Los Angeles, UCLA.

Richardson, M. and Domingos, P., 2006, Markov logic networks. *Machine Learning*, 62, 107–136.

Scha, R., 1981, Distributive, Collective and Cumulative Quantification. In *Formal Methods in the Study of Language*, Amsterdam, Mathematisch Centrum, 483–512.

Schank, R. C. (ed,), 1975, *Conceptual information processing*. Amsterdam, North-Holland.

Schank, R. C., 1982, *Dynamic Memory*. New York, Cambridge University Press.

Schank, R. C. and Tesler, L. G., 1969, A conceptual parser for natural language. *Proc. IJCAI-69*, 569–578.

Szabolcsi, A., 2002, Hungarian disjunctions and positive polarity. In I. Kenesei (ed.), *Approaches to Hungarian* 8. Szeged, SzTE.

Szabolcsi, A. and Haddican, B., 2004, Conjunction meets negation: A study in cross-linguistic variation. *Journal of Semantics* 21(3), 219–249.

Zadrozny, W., 1994, From compositional to systematic semantics. *Linguistics and Philosophy* 17, 329–342.

[1]I am grateful to Zsófia Zvolenszky for her comments on the manuscript.

András Kornai

The Treatment of Ordinary Quantification in English Proper

Abstract. In this paper we bring together some well-known lines of criticism directed at Montague Grammar, such as (i) taking a stilted, highly regulated variety of language as the object of inquiry; (ii) ignoring the meaning of content words; and (iii) the failure to treat hyperintensionals; and suggest a coherent, and we believe much simpler alternative based on structured meanings.

INTRODUCTION

Among the many inventions that made PTQ such a monument of intellectual achievement was the insistence on presenting a significant fragment of *ordinary* English, as opposed to the semi-formalized (sometimes fully formalized) and regimented English-like sublanguages used in most works of philosophical logic at the time. Yet upon rereading the founding papers of Montague Grammar (MG), in particular (Montague 1970a, 1970b, 1973), all reprinted in (Thomason 1974), linguists brought up in a more descriptive tradition are inevitably struck by the stilted "English" of the fragments. The problem is not so much that the pioneering examples from *Every man loves a woman such that she loves him* to *John seeks a unicorn and Mary seeks it* could hardly be regarded as examples of ordinary language but that there has been an alarming lack of progress in this regard. Nearly forty years have passed, and papers discussing seminumerical puzzles like *At least three professors graded at most five exams from seven or fewer students* are abundant, while the interpretation of ordinary language makes little progress. It's not that people couldn't say things like *My grandmother has more plates than utensils* it's just that they don't.

In this paper, while still avoiding the rough and tumble of actual spoken English, we take a less regimented language variety, that of copy-edited journalistic prose, and investigate the semantics of *ordinary quantification* in this variety we

call *English proper* (rather than 'proper English' to avoid the association with primness and pedantry). In Section 1 we discuss how various uses of *for all* and *every* are to be sorted out, and describe informally how the dominant usage of *every* can be accounted for in a model that preserves most of the goals, but very little of the formal techniques, of MG. In Section 2 we suggest that the analysis of quantifiers and similar function words needs to be supplemented by a more detailed analysis of content words, if semantics is to cover more realistic examples—statetements and inferences of the sort found both in ordinary English and in philosophical discourse. Since this brings into sharp focus the traditional puzzles concerning hyperintensionals, we consider these in some detail, and arrive at the conclusion that standard intensional theory is not tenable.

1 ORDINARY QUANTIFICATION

Our sample of universally quantified expressions is drawn from an American newspaper, the San Jose Mercury News (Merc). We manually analyzed 300 issues, totaling some 45 million words, to demonstrate that ordinary quantification is not at all close to the logical variety described in MG. As we shall see in Section 1.1, the data is dominated by a large number of other patterns. Since these are rarely amenable to a purely lexical, purely grammatical, or purely pragmatic treatment, in Section 1.2 we introduce a new Principle of Responsibility that forces us to look at such constructions more closely.

1.1 Universal constructions in English

In our corpus we find over 2400 occurrences of the strings *for all* and *For all*. Many of these could be called idiom chunks: *For all the glamour of aerial fish planting, it was a mass production money-maker []*[1] clearly does not mean anything like $\forall x glamour(x)(\ldots)$. A descriptive label such as *idiom*, *(partially) lexicalized expression*, or *snowclone* already hints at the necessity of stepping outside 'pure' semantics, bringing resources of a lexical or pragmatic sort on the issue at hand. In what follows, we will describe constructions as (partially) fixed patterns in the spirit of (Fillmore and Kay 1997), and ask what expressions like *[the Clarence Thomas hearings], for all their import []* or *For all their efforts at parity and fairness, [NFL officials ...]* actually mean. This is not a problem somewhere on the fringes of the data—as a matter of fact, examples like these are considerably more frequent than those involving standard quantifier readings.

[1]For reasons of expository convenience we will often considerably simplify the raw examples, indicating inessential parts by [] wherever necessary. In doing so, we attempt to make sure the simplified example remains an instance of English proper, i.e. an example that could be produced by a reasonable writer of English and would be left standing by a reasonable copy-editor.

We define a *construction* as a string composed of nonterminals (variables ranging over some syntactic category) and terminals (fixed grammatical formatives and lexical entries) with a uniform compositional meaning, obtained by a fixed process whose inputs are the meanings of the nonterminals and whose output is the meaning of the construction as a whole. Completely productive and highly abstract grammatical patterns such as

(1) $NP\langle\alpha PERS\ \beta NUM\rangle\ VP\langle\alpha PERS\ \beta NUM\ \gamma TENSE\rangle$

and highly specified and almost entirely frozen idioms such as

(2) $NP\langle\alpha PERS\ \beta NUM\rangle$ *kick*$\langle\alpha PERS\ \beta NUM\ \gamma TENSE\rangle$ *the bucket*

will both be treated as constructions. On occasion, when we are interested in the substitution of one construction into another, it will be necessary to assign a grammatical category (defined as including morphosyntactic features specified in angled brackets) to the construction as a whole, so a syntactic theory roughly along the lines of GPSG (Gazdar *et al.* 1985) is presupposed. The combinatorial flexibility of the constructions suggests that more powerful theories, such as TAGs or LTAGs, (Joshi and Schabes 1997) may actually be a better choice, but for our current purposes, we are not particularly interested in transference across patterns, such as the phenomenon that the agreement portion of (2) is obviously inherited from that of (1). As a limiting case, entirely frozen expressions, i.e. those constructions that no longer contain open slots, like *go tell it to the Marines*, are simply taken as lexical entries, in this case, with meaning 'nobody cares if you complain'. (The indexicals implicit in the imperative *go* and explicit in the paraphrased *you* do not constitute open slots in the sense we are interested in here.)

Viewed from this perspective, the standard case, found in many examples like *the law [] makes helmets mandatory for all motorcyclists and passengers* or *[] lowers the quality of life for all concerned* is yet another construction:

(3) *X for all N*

where *N* is some (bare) noun phrase and X is some predicative element, often a verb or VP, as in *lowers the quality of life*, but, perhaps surprisingly, more often a nominal or NP, as in *a model for all nations*.

In many cases, the adjacency of *for* and *all* appear accidental, as in *[] honored for all his work* or *sell [] for all that the market will bear* even though in some of these cases the standard analysis ($\forall x$*his_work*$(x)\ldots$ or $\forall x$*market_will_bear*$(x)\ldots$) remains prima facie available.

Turning to *every*, of which there are about five times as many examples, here the dominant pattern is indeed one that lends itself to analysis in the standard terms: *every Californian with a car phone, every case, every famous star, ...* Remarkably, about 20% of these are time adverbials, *every day, every week, every time, every night* and so on. There are considerably fewer constructions with *every* than with *for all*, but they all share the property with (1) that they admit exceptions. In many cases, this is confirmed directly by the text: *every case, except that of Sen. Kennedy []*. In others, the text offers no overt exceptions, but it is clear that exceptions can be made: *every Californian with a car phone* except, of course, drivers of emergency vehicles...

Manual inspection of a large number of *every N<BAR 1>* constructions makes clear that their meaning is really 'every non-exceptional N' rather than 'every N'—in fact, in the whole Merc corpus we could not find a single example of the latter. The results would have been very different with a corpus based on calculus textbooks, and we do not deny that the episodic reading routinely analyzed in MG exists, at least in a regimented variety of technical English—the claim is simply that it is not a part of English proper.

We should add here that we claim no originality in recognizing the problem, as the defeasibility of natural language statements has already given rise to a wide variety of *non-monotonic logic* approaches (for an overview see (Ginsberg 1987)) and the fact that generics admit exceptions is often viewed as one of their defining properties since (Jespersen 1924). If there is an original claim to be made in this area, it is that universal quantification, as the term is understood in predicate calculus, plays no role whatsoever in ordinary English or, indeed, in any natural language.

To put this finding in the harshest possible terms, PTQ fails to deliver on its major promise to treat quantification in ordinary English, concentrating on the jargon of mathematics instead. While subsequent work in the MG tradition such as (Moltmann 1995) and (Lappin 1996) have clearly recognized, and to some extent resolved the local problem of exceptionality, the global problem of dealing with the large variety of relevant constructions sampled here remains as acute as ever.

1.2 The meaning of constructions

Among the thousands of constructions used in English only a handful like *(go) tell it to the Marines* are amenable to a purely lexical treatment, and only a handful like $S \rightarrow NP\ VP$ are purely compositional. In between, there is a vast range of expressions containing one or more open slots, and our primary interest in doing semantics lies with interpreting these. There is a clear intuition that nonce phrases such as *California driver* differ from lexicalized forms such as *Rottweiler dog* only in that the latter is part of the lexicon, and the basic

components of its meaning is no longer surface accessible. California drivers are obviously humans with two properties that are true by definition, namely, that they live in California and that they drive a car, and many others that are derivable from these, such that they are above the California driver age limit or that they are featherless bipeds. In the formal model that we outline in Section 3, the meaning of nominals will be taken as a bundle (unordered conjunction) of predicates that correspond to the *essential* properties of the nominal in question. It is clear that both 'being Californian' and 'being a driver' are definitely part of the bundle of properties that form the base of the semantic model for California drivers, and we leave open the possibility that several other predicates are also part of the bundle. For example, it is hard to know whether the stereotype that California drivers are polite is part of the lexicon (viewed as a purely grammatical construct) or belongs in some nebulous encyclopedia of world knowledge. But to the extent speakers of English can and do pursue inferences on this basis, we view it as part of the task of semantics to account for these. We state this as our *Principle of Responsibility:*

> The semantics of any expression must be fully accounted for by the lexicon and the grammar taken together.

The Principle of Responsibility is only slightly stronger than the standard Principle of Compositionality which takes the semantics of any expression to be determined by the semantics of its lexical components and by the grammatical way those are combined. The additional requirement it imposes is that the 'pragmatic wastebasket' remain empty at all times: it doesn't matter whether we call ordinary inferences grammatical, lexical, or pragmatic (and perhaps extragrammatical), the overall system needs to account for these, either in one specific component, or by means of tracing the inference process through several components.

Let us begin with the *for all NP<+DEF>, S* construction. Clearly, this means something like '*S*, in spite of the usual implications of *NP<+DEF>*'. In the case of *the glamour of aerial fish planting*, the implication that needs to be defeased is that glamorous things are restricted to the few, a notion incompatible with *mass production*. The lesson from the example is already clear: to make sense of the construction we need to use a great deal of lexical information. Without doing so, the clear difference between the acceptability of the Merc examples and *???For all their protein content, eggs are shaped so as to ease passage through the duct* would remain completely mysterious.

As for non-exceptionality, mathematics offers two significantly different formal reconstructions of this notion. One approach, exemplified in (4), relates non-exceptionality to probability:

(4) *The typical number is irrational*

and can be rephrased as 'the set of exceptions has measure zero'. The other, exemplified in (5), relates to the satisfaction of no extra predicates:

(5) *The typical square matrix over Q has unequal eigenvalues*

Statements like (4) occur so frequently in mathematical discourse that they have a terminus technicus of their own: we say *almost all [numbers are irrational]*. Statements like (5) are interpreted more in terms of dimension than in terms of measure, and we speak of *lower dimension* when reducing these to more primitive notions.

These two approaches to non-exceptionality are not incompatible, but it may take very significant work to establish, as Martin-Löf (1966) did, that the statement *The typical binary string is not compressible* yields the same definition from both the measure-theoretic and the no-extra-predicates standpoint. Here we choose the second approach in light of the fact that there is no obvious way to define measure spaces over semantic objects like *legal cases* or *California drivers with car phones*.[2] Thus, to say that *Geraldo Rivera [[] reveals that he is an extremely attractive virile hunk of man who] has had sex with [] every famous star in the entertainment industry []* is to say that for all x such that x has no extra properties beyond being a famous star in the entertainment industry, Geraldo Rivera has had sex with x.

A less clumsy translation, very much in the spirit of generalized quantification, would be to say that the property of having had sex with Geraldo Rivera is implied by the property of being a famous star in the entertainment industry, and this is what we adopt for ordinary quantification: we say that *every N* is the set of *typical* properties that N has, where typicality is defined in the lexical entry of N. Since having four legs is typical of donkeys, *every donkey has four legs* will be true by definition, and cannot be falsified by the odd lame donkey with three or fewer legs.

But if having four legs is an analytic truth for donkeys, what about counterfactuals where five-legged donkeys can appear easily, or the rather clear intuition, not disputed here, that being four-legged is a contingent fact about donkeys, one that can be changed e.g. by genetic manipulation? The answer offered here is that to reach these, we need to change the lexicon. Thus, to go from the historical meaning of Hungarian *kocsi* 'coach, horse-driven carriage' to its current meaning '(motor) car' what is needed is the prevalence of the motor variety among 'wheeled contrivances capable of carrying several people on roads'. A 17th century Hungarian would no doubt find the notion of a horseless coach just as puzzling as the notion of flying machines or same-sex marriages. The key

[2]Not only do we need a measure space, we would need substantive agreement that this particular measure is the one that is "natural" to the domain, just as Lebesgue-measure is agreed to be *the* natural measure for real numbers. While defining measures for semantic objects can be done many ways, arguing for any of these as being natural is much harder.

issue in readjusting the lexicon, it appears, is not counterfactuality as much as rarity: as long as cloning remains a rare medical technique we won't have to say 'a womb-borne human'.

A notable consequence of our definition by typical properties is that the translation of *every donkey* will not differ significantly from that of *any donkey*, *a donkey*, *donkeys* or even *the donkey:* the typicality restriction pertains to them all. This is as we want it for cross-linguistic purposes, since the clearly generic readings are not tied to the same varieties of quantified NPs in all languages.

To summarize what we have so far: *every man loves a woman* means neither $\forall x man(x) \exists y woman(y) loves(x, y)$ nor $\exists y woman(y) \forall x man(x) loves(x, y)$—it means that woman-loving is a typical property of men, just as donkey-beating is a typical property of farmers. Importantly, it requires evidence beyond what is available in the example sentences to know whether farmers beat every donkey they can lay their hands on or just their own, and whether men love every women or just one.

2 CONTENT WORDS

Legend has it that once in a semantics class a student asked Barbara Partee *What is the meaning of life?*, and she responded, after a moment of thought, by writing **life´** on the blackboard. Since a key goal of the whole semantics enterprise is to provide a more satisfactory answer, we begin with analyzing two well known approaches. The first one, conventionally attributed to Koheleth, the author of Ecclesiastes, is that life is vanity, entirely devoid of meaning or purpose. According to Macbeth, "life ... is a tale told by an idiot, full of sound and fury, signifying nothing" (Act 5, Scene 5). Perhaps the most articulate exponent of this position is Schopenhauer, but we find many thinkers expressing the same idea before and after him.

The other well known answer is the religious one, that the meaning of life is to serve God. Somewhat surprisingly, given the magnitude and importance of the problem, this answer is relegated to a subordinate clause of a longer story concerned with something else, hidden in a book generally regarded minor, Isaiah 43.6-7, wherein the standard Judeo-Christian approach is spelled out as follows: "bring my sons ... every one that is called by my name: *for I have created him for my glory,* I have formed him; yea, I have made him." Lest the reader feel disappointed we should emphasize at the outset that our primary interest here is not so much with exegesis as with lexicography. Instead of attacking the major problem posed by the student and many before him, we merely seek a technical approach that at least makes it possible to formulate the traditional answers sketched above.

We see in Partee's witty response a deeper truth, namely that Montague semantics lacks entirely the resources to approach issues of word meaning. The

problem, from our standpoint, is not so much that we don't know the meaning of life; rather, the problem is that even if we did, we couldn't express it within the standard framework. Assume, for a moment, that the first answer is correct, that life has no meaning. Does this mean **life'** $= \emptyset$, and if so, how can this be derived along the lines proposed by Koheleth, from the observation that "All go unto one place; all are of the dust, and all turn to dust again"? Perhaps we need a subsidiary axiom that meaning is a permanent, unchanging and unchangeable thing, so that if something is not eternal it must be meaningless. But is it the object that must be eternal or is it its meaning, and how would we distinguish the two cases? Or assume, for the sake of argument, that the second answer is the correct one, that God has created Man for His glory. Really, what does this mean? Does it mean, on account of the masculine pronoun being used, that Woman is exluded? And what is *glory* so that even God cares to have more of it? The whole MG framework, which treats the meaning of everything other than a few function words as an unanalyzed set, is incapable of formulating, let alone resolving, such questions.

The issue is really not the meaning of the word *meaning*. The main question could be recast in many other ways, as an inquiry concerning the *purpose* of life, the *goal* of life, and so forth. The technical reconstruction of 'meaning' used by MG as the set of instances in this world or other possible worlds, is quite satisfactory. But what is the set of purposes, the set of goals, or even the set of living things? We do not wish to create a mystery where there isn't one, for there are perfectly reasonable commonsensical answers which accord well both with everyday and with philosophical usage, and every dictionary will have some version of these, stated in terms of a rather simple theory of lexical semantics wherein the meaning of nominals is conceived of as a bundle of properties.

Historically such a theory can be traced back to Leibnitz, the Schoolmen, and eventually to Aristotle. In contemporary semantics this idea is at the foundations of both the Semantic Web (there called Web Ontology Language or OWL) and of the influential WordNet approach to the lexicon (Miller 1983). Yet in contemporary philosophy such theories have little credibility since Russell's (1905) critique of Meinong, with (Parsons 1974) definitely remaining a minority view. The technical difficulty is in formulating inferences based on such dictionary definitions, because the theory that can sustain definitions for nouns such as *fox* composed of properties such as `animal, four-legged, red, clever` can also sustain definitions of inconsistent and/or non-existent objects, to which we now turn.

2.1 Hyperintensionals

Russell had two major objections: first, a technical one concerning the existence predicate. Let us take any nonexistent object of Meinong's Jungle,

such as the gold mountain. Such an object doesn't exist, but if Meinong is right, and any combination of properties can be construed as an object, the existent gold mountain is also part of the Jungle, and as such it is not only gold, and mountain, but also existent, contradicting our earlier assertion that it was nonexistent. This objection is easily defeased by splitting the existence predicate in two, 'exists in the mind', and 'exists in reality', and exempting the second meaning from the class of predicates that can be used to describe (or in the mind, create) noun objects. Russell's second point, concerning inconsistent bundles of properties, such as triangular circles, was that they violate the law of non-contradiction. Before going any further, let us consider some examples.

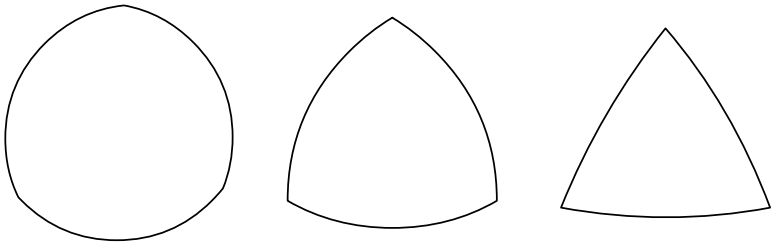

Figure 1. The Reuleaux triangle and its cousins

The figure shows on the left a slightly triangular circle, and on the right a slightly circular triangle. Whether the object in the middle, known as the "Reuleaux triangle", is considered a triangle, a point of view justified by its having three distinct vertices, or a circle, a point of view justified by its having a constant diameter, is a matter of perception.

What is clear from the linguistic standpoint is that adadjectives like *slightly*, *seemingly*, *very* attach to adjectives like *circular*, *triangular*, *equal* that have a strict mathematical definition just as easily as they attach to adjectives like *red*, *large*, *awful* that lack such a definition. Clearly, what these adadjectives modify is the "everyday" sense of these terms—the mathematical sense is fixed once and for all and not subject to modification. Just as we were interested in the everyday sense of *all* and *every* and found that these are distinct from the standard mathematical sense taken for granted in MG, here we are interested in the ordinary sense of *circular*. Working backwards from typical expressions like *circular letter*, *circular argument*, we find that the central aspect of the meaning is not 'a fixed distance away from a center' or even 'fixed diameter', but rather 'returning to its starting point', 'being cyclic'.

In these examples, the morphologically primitive forms are nominal: the adjectival forms *circular*, *triangular* are clearly derived from *circle*, *triangle* and not the other way around. Since derivations of this sort change only the syntactic category of the expression but preserve its meaning, we can safely conclude that

circle in the everyday sense is defined by some finite conjunction of essential properties that includes 'being cyclic' and that the mathematical definition extends this conjunction by 'staying in an (ideal) plane, keeping some (exact) fixed distance from a point'. Similarly, *triangle* simply means 'having three angular corners' rather than the exact configuration of points and lines assumed in geometry. Taking these notions together, the predicate bundle of *triangular circle* will contain the properties `curve, cyclic, has(three vertices)` and all three shapes depicted above will fit this definition.

A more general example of the same mechanism is provided by the adjectival modification of proper names. Who is *the Polish Shakespeare* that the "*Looking for the Polish Shakespeare*" *Contest for Young Playwrights* wants to find? Clearly, not some British subject born in Stratford-upon-Avon but a brilliant playwright who is a Polish national. Once we recognize that adjectival modification is not simply a conjunction of some new property to the set of essential properties, but one that may interpose a higher predicate that is only implicit (as is `nationality` in this case), a whole range of otherwise puzzling constructions become transparent. Altogether, English proper has far fewer hyperintensional constructions than hitherto assumed: certainly triangular circles and immaculate conceptions do not give rise to logical contradictions. This renders the well-known problem with hyperintensionals in the MG account of opacity far less urgent, as we now only have to deal with cases in which the essential meaning of the adjective is in strict contradiction to the essential meaning of the noun it modifies, and the latter is given by a single conjunct. Thus we need to consider examples like

(6) *Mondays that fall on Tuesdays*

(7) *Mondays that fall on Wednesdays*

Does it follow that a (rational) agent who believes in (6) must also believe in (7)? While the matter is obviously somewhat speculative, we believe the answer to be negative: if we learn that *Peter believes Mondays can be really weird—he actually woke up to one that fell on Tuesday* it does not follow that he also believes himself to have woken up on one that fell on Wednesday. Since he has some sort of weird experience that justifies for him a belief in (6), he is entitled to this belief without having to commit himself to (7), as the latter is not supported by any experience he has. If this is so, the hyperintensional problem is still relevant, and the intensional treatment of opacity cannot be maintained for all cases. But if it cannot be maintained for all cases, there does not seem to be a compelling reason to maintain it at all, since a simpler alternative treatment, based on structured meanings (Cresswell 1985), is available (see (Ojeda 2006) for detailed argumentation why a non-intensional treatment is to be preferred both on grounds of simplicity and grounds of adequacy). Being algebraic, the

structured meanings we use (see (Kornai 2009)) are somewhat different from those used by Cresswell, but all the reasoning that leads to structured meanings remains applicable. In particular, the theory presented here puts fiction on a par with factual discourse: when we assert truthfulness we do this relative to a particular model, so that *Anna Karenina commits suicide* can be true while *Jan Valjean commits suicide* is false, relative to their respective models.

According to Thomason (1977) a heavy price must be paid for adopting a semantic theory based on structured meanings: in brief, such a theory is incompatible with a nontrivial theory of truth. In his discussion of Montague (1963), Thomason argues that any direct theory of propositional attitudes is bound to be caught up in Tarski's (1935) Theorem of Undefinability. However, as Thomason is careful to note, the conclusion rests on our ability to pass from natural language to the kinds of formal systems that Tarski and Montague consider: first order theories with identity, strong enough to model arithmetic. Tarski himself was not sanguine about this: he held that in natural language "it seems to be impossible to define the notion of truth or even to use this notion in a consistent manner and in agreement with the laws of logic". Russell held similar views, calling natural language "a rough and ready instrument incapable of expressing Truth with a capital T".

To replicate Tarski's proof, we first need to supplement natural language with variables. The basic idea—to formalize the semantics of a predicate like *subject owns object* by a two-place relation $\zeta(s,o)$—is fairly standard (although there are significant alternatives that do not rely on variables at all). But the proposed paraphrases for first order formulas, such as replacing $\forall x[\exists y\zeta(x,y) \rightarrow \exists z\zeta(z,x)]$ by *for everything x, either there is not something y such that x owns y or there is something z such that z belongs to x* clearly belong to an artificially regimented extension of English, rather than to English proper. Second, we must assume that the language can sustain a form of arithmetic, e.g. Robinson's Q.

The universality of natural language (or English proper) as a means of supporting logical or arithmetic calculi, is highly doubtful, and using Q we can pinpoint the source of these doubts more narrowly: several of the axioms in Q appear untenable for natural language. Interestingly, the key issues arise long before we consider exponentiation (a central feature for Gödel numbering) or ordering. Q comes with a signature that includes a successor s, addition +, and multiplication $\cdot$. By Q2 we can infer $x = y$ from $sx = sy$, Q4 provides $x + sy = s(x + y)$, and Q6 gives $x \cdot sy = xy + x$. All of these axioms are gravely suspect in light of the following Principle of Non-Counting:

$$\text{If } \alpha p^n \beta \in L \text{ for } n > 4, \alpha p^{n+1} \beta \in L \text{ and has the same meaning}$$

The notion that linguistic structures are non-counting goes back at least to Chomsky (1965, 55) and pervades every variety of syntax that we know

of.[3] There are many ways we can *start* counting in natural language: we can look at quotations of quotations *(Joe said that Bill said...)*, emphasis of emphasized material *(very very...)*, but there is not a single way that takes us very far—whichever way we go, we reach the top in no more than four steps, and there Q2 fails. Since on this key point, the semantics of natural language expressions parts with the semantics of mathematical expressions, the empirical underpinnings of the Tarski/Montague/Thomason argument are missing: there is no loss entailed by the use of structured meanings, as there was no chance to maintain a Tarski-type theory of truth in natural language to begin with.

3 CONCLUSIONS

Historically, MG started out as an ambitious but quite reasonable research program, explicitly moving away from the stilted examples of an earlier generation of logic textbooks toward ordinary natural language. Indeed, many of the pivotal examples motivating much subsequent work, e.g. Bach-Peters sentences, have an immediate impact, clearly comprehensible to any native speaker. We believe that increasingly arcane examples with increasingly contrived readings reasserted themselves as the primary focus of interest simply because MG and its modern descendants, starting with Dowty (1979), concentrated on elucidating the semantic analysis of those expressions for which the underlying logic had the resources. Since Montague's intensional logic IL includes a time parameter, in depth analysis of temporal markers (tense, aspect, time adverbials) becomes possible. But as long as the logic lacks analogous resources for space, kinship terms, sensory inputs, or obligations, this approach has no traction, and heaping all these issues on top of what was already a computationally intractable logic calculus has not proven fruitful. The goal of this paper was to drastically realign the focus of formal semantics from interesting puzzles and a Turing-complete higher order intensional apparatus to data of the simple and frequent kind that is likely to dominate the language acquisition process. We must crawl before we walk, and if we cannot account for the data that is likely seen during language acquisition our account of more complex phenomena is in doubt.

ACKNOWLEDGMENTS

This paper builds extensively on ideas presented in Kornai (2008). We thank Károly Varasdi (HAS Institute of Linguistics), Anna Szabolcsi (NYU), and Zoltán Gendler Szabó (Yale) for penetrating comments on earlier drafts.

[3]Some limited counting, such as the building of binary (and perhaps ternary) feet, is generally assumed in phonology.

REFERENCES

Chomsky, N., 1965, *Aspects of the Theory of Syntax*. Cambridge (MA), MIT Press.
Cresswell, M. J., 1985 *Structured Meanings*. Cambridge (MA), MIT Press.
Dowty, D., 1979, *Word Meaning and Montague Grammar*. Dordrecht, Reidel.
Fillmore, Ch. and P. Kay, 1997, *Berkeley Construction Grammar*. `http://www.icsi.berkeley.edu/~kay/bcg/ConGram.html`
Fine, K., 1985, *Reasoning with Arbitrary Objects*. Oxford, Blackwell.
Gazdar, G., E. Klein, G. K. Pullum and I. A. Sag, 1985, *Generalized Phrase Structure Grammar*. Oxford, Blackwell.
Ginsberg, M. L. (ed.), 1986a, *Readings in Non-monotonic Reasoning*. San Mateo(CA), Morgan Kauffman.
Jespersen, O., 1924, *The Philosophy of Grammar*. London, George Allen and Unwin.
Aravind K. Joshi, A. K. and Y. Schabes, 1997, Tree-adjoining grammars. In *Handbook of Formal Languages*, volume 3: Beyond Words. Berlin, Springer.
Kornai, A., 2003, The algebra of lexicography. In G. Jager and J. Michaelis (eds), *Proceedings of the 11th Mathematics of Language Workshop*, FoLLI Lecture Notes in Artificial Intelligence. Berlin, Springer Verlag.
Kornai, A., 2008, *Mathematical Linguistics*. Berlin, Springer Verlag.
Lappin, Sh., 1996, Generalized quantifiers, exception phrases, and logicality. *Journal of Semantics* 13, 197–220.
Martin-Löf, P., 1966, The definition of random sequences. *Information and Control* 6, 602–619.
Miller, G. A., 1995, Wordnet: a lexical database for english. *Communications of the ACM* 38, 39–41.
Moltmann, F., 1995, Exception phrases and polyadic quantification. *Linguistics and Philosophy* 18, 223–280.
Montague, R., 1963, Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philosophica Fennica* 16, 153–167.
Montague, R., 1970a, English as a formal language. In (Thomason 1974), 188–221.
Montague, R., 1970b, Universal grammar. *Theoria* 36, 373–398.
Montague, R., 1973, The proper treatment of quantification in ordinary English. In (Thomason 1974), 247–270.
Ojeda, A., 2006, Discontinuous constituents. In *Elsevier Encyclopedia of Languages and Linguistics*. Amsterdam, Elsevier.
Parsons, T., 1974, A prolegomenon to Meinongian semantics. *The Journal of Philosophy* 71, 561–580.
Russell, B., 1905, On denoting. *Mind* 14, 441–478.
Tarski, A., 1956, The concept of truth in formalized languages. In A. Tarski, editor, *Logic, Semantics, Metamathematics*. Oxford, Clarendon Press, 152–278.
Thomason, R. H. (ed.) 1974, *Formal Philosophy: Selected papers of Richard Montague*. New Haven, Yale University Press.
Thomason, R. H., 1977, Indirect discourse is not quotational. *The Monist* 60, 340–354.

Péter Mekis

Atomic Descriptions in Dynamic Predicate Logic

Abstract. We introduce a version of dynamic predicate logic (*DPL*, see (Groenendijk *& al.* 1991) as a framework to model in a compositional way the dynamics of definite descriptions put forward in (Lewis 1979). The resulting system, *dynamic predicate logic with descriptions* (*DPLD*) borrows the concept of a referent system from an upgraded version of *DPL* introduced in (Groenendijk *& al.* 1996). It is an interesting feature of *DPLD* that not only formulas but also individual terms are capable of updating discourse information.

1 INTRODUCTION

David Lewis, in his seminal paper (Lewis 1979) that served as a philosophical basis for the project of dynamic semantics, put forward a cluster of linguistic phenomena that deserve dynamic treatment. One of them is the anaphoric use of definite descriptions. Strangely enough, in more than thirty years since Lewis's paper this phenomenon has apparently not been studied within the framework of dynamic semantics.[1]

According to Lewis, examples like "The dog got in a fight with another dog" show that "[i]t is not true that a definite description 'the F' denotes x if and only if x is the one and only F in existence. Neither is it true that 'the F' denotes x if and only if x is the one and only F in some contextually determined domain of discourse." (348.) Instead, "[t]he proper treatment of descriptions must be more like this: 'the F' denotes x if and only if x is the most salient F in the domain of discourse, according to some contextually determined salience ranking." (ibid.)

[1]On the other hand, it has been studied to a certain extent within the *DRT* framework; see (Eijck *& al.* 1997, 189ff.) The *DRT* treatment is completely different from ours.

This view contradicts the classical view on descriptions dominant in logic ever since Frege and Russell at crucial points.

Salience may be determined by the context of utterance or antecedent expressions. Both can be found in Lewis's following example:

(1) The cat is in the carton. The cat will never meet our other cat, because our other cat lives in New Zealand. Our New Zealand cat lives with the Cresswells. And there he'll stay, because Miriam would be sad if the cat went away.

In Lewis's analysis, the referent of the first occurrence of "the cat" is determined by external conditions, and the second one refers back to the first. Then, the term "our other cat" introduces another discourse referent, and "our New Zealand cat" refers back to "our other cat", which has meanwhile been associated with the term "New Zealand".

All of these four description occurrences deserve special attention. The semantic phenomena involved in this sample discourse are very complex and have far-reaching implications, only a small portion of which can be discussed in this paper. In particular,

(a) we are going to focus on atomic descriptions like "the cat" or "the man", since they seem to be more capable of linking anaphorically, while complex descriptions tend to behave in the Frege–Russell way;
(b) we attempt to model the anaphoric use of atomic descriptions, like "the cat" referring back to "a cat" and "the one that lives with us" referring back to "the cat" in the discourse "We have a cat that lives with us, and another one that lives in new Zealand. The one that lives with us is old";
(c) we also attempt to model the way atomic descriptions take discourse referents from external sources (or even introduce new discourse referents) when they have no salient antecedent, like "the cat" in the opening sentence of Lewis's example "the cat is in the carton":

The technical problem to be solved is that of representing the way salience information is gathered, updated and put to use in a discourse. In our approach, salience is a relation between discourse referents and predicates. Since the former are semantic entities, while the latter syntactic ones, this is a mixed type of information. Clearly, a predicate itself is a better choice as an object of salience information than its extension. But we cannot rule out the possibility that intension plays a role in salience ranking, as the following example may suggest:

(2) A bachelor has a date with an old maid. ??The man is nervous. The woman is bored.

Whether and how anaphoric linking in this example works is beyond the scope of this paper.

We take Lewis's statement "[i]t is not true that a definite description 'the F' denotes x if and only if x is the one and only F in existence" as an exaggeration. An account of definite descriptions that takes the uniqueness and existence conditions attached to "the smallest prime number" or "the author of *Counterfactuals*" as merely extreme cases of salience ranking would be highly counterintuitive. The use of descriptions that involves uniqueness and existence (henceforth, we refer to a description interpreted this way as a *classical description*) is just as legitimate as the anaphoric one, and it may even have priority over that. But descriptions, like so many other natural language expressions, are tools that—in spite of being of the same form—have different uses. In this paper we attempt to model one of these uses.

2 THE DYNAMICS OF ATOMIC DESCRIPTIONS

Cross-sentential anaphora is one of the two semantic challenges that *DPL* took on (the other being the anaphoric relations in donkey-sentences, which we will not discuss here). Let us see an example of it and the way *DPL* treats it.

(3) A man walks in the park. He meets a woman.

Standard first-order logic (henceforth, *PL*) translates this discourse as

(3a) $\exists x \left(\mathtt{man}(x) \;\&\; \mathtt{walk\_in\_the\_park}(x) \;\&\; \exists y \left(\mathtt{woman}(y) \;\&\; \mathtt{meet}(x, y)\right)\right)$[2]

This translation reflects the anaphoric link between "he" in the second sentence and "a man" in the first by placing all the occurrences of x within the scope of $\exists x$. The price is that the translation is not compositional; no syntactic component of the formula translates the first sentence of (3). On the other hand,

(3b) $\exists x \left(\mathtt{man}(x) \;\&\; \mathtt{walk\_in\_the\_park}(x)\right) \;\&\; \exists y \left(\mathtt{woman}(y) \;\&\; \mathtt{meet}(x, y)\right)$

is not a correct translation, since the last occurrence of x is not bound by the quantifier $\exists x$. In *PL*, (3a) is not equivalent with (3b).

DPL offers a framework in which the syntactic scope of a quantifier occurrence (that is, the formula to which the quantifier is prefixed in the syntactic construction) separates from its semantic scope (that is, the part of the discourse in which it is able to bind variable occurrences). In *DPL*'s semantics, due to the dynamic rules governing existential quantification and conjunction, (3a) is equivalent with (3b).

Moreover, in *DPL* anaphoric linking is possible even between the conclusion of a consequence and one of its premises. Consider the following example and its attempted translations:

[2]As a tribute to professor Ruzsa, we are going to use his versions of the logical constants, wherever it is possible, including $\&$ as conjunction, $\sim$ as negation, $\supset$ as implication, $\equiv$ as the biconditional and I as the descriptor. See, for example, (Ruzsa 1981, 16ff.)

(4) *Witness*: A man entered the house and switched the light on. He had a knife. *Inspector*: So, he switched the light on.

(4a) $\{\exists x\,(\texttt{man}(x)\ \&\ \texttt{enter\_the\_house}(x)\ \&\ \texttt{switch\_the\_light\_on}(x),\ \texttt{have\_a\_knife}(x))\} \models \texttt{switch\_the\_light\_on}(x)$

(4b) $\forall x\,(\texttt{man}(x)\ \&\ \texttt{enter\_the\_house}(x)\ \&\ \texttt{switch\_the\_light\_on}(x)\ \&\ \texttt{have\_a\_knife}(x)) \supset \texttt{switch\_the\_light\_on}(x))$

This is an example that cannot be properly translated into *PL* even if we give up compositionality. In standard semantics, (4a) is not a valid consequence, and (4b) is not even an consequence. On the other hand, *DPL* has a dynamic consequence relation that makes "cross-inferential" anaphora possible, and makes (4b) valid.

In *DPLD* we want to deal with similar examples involving atomic definite descriptions. The first one and its compositional translation are as follows:

(5) A man walks in the park. He meets a woman. The man hugs her. A man watches from a distance. He walks a dog. The dog is bored. The man is jealous.

(5a) $\exists x\,\texttt{man}(x)\ \&\ \texttt{walk\_in\_the\_park}(x)\ \&\ \exists y\,\texttt{woman}(y)\ \&\ \texttt{meet}(x,y)\ \&\ \texttt{hug}(\mathrm{I}w\ \texttt{man}(w), y)\ \&\ \exists z\,\texttt{man}(z)\ \&\ \texttt{watch\_from\_a\_distance}(z)\ \&\ \exists v\,\texttt{dog}(v)\ \&\ \texttt{walk}(z,w)\ \&\ \texttt{bored}(\mathrm{I}w\ \texttt{dog}(w))\ \&\ \texttt{jealous}(\mathrm{I}w\ \texttt{man}(w))$

In this discourse, all occurrences of descriptions refer back anaphorically to indefinite ones. One particularly interesting point is that the two occurrences of "the man" have two different antecedents, both of which are occurrences of the same expression "a man". Let us call d_1 and d_2 the discourse referents introduced by the first and the second occurrences of $\exists x$, respectively; and let us call s_1 the salience information that the predicate man is associated with d_1, and s_2 the salience information that the predicate man is associated with d_2. Now, *DPLD* has to interpret this formula in such a way that s_1 is passed to the first occurrence of $\mathrm{I}w\ \texttt{man}(w)$, but it is replaced with s_2 in the course of the updating process before the second occurrence of $\mathrm{I}w\ \texttt{man}(w)$.

DPLD also has to treat descriptions that lack possible antecedents, like Lewis's

(6) The cat is in the carton. She is asleep.

(6a) $\texttt{in\_the\_carton}(\mathrm{I}x\ \texttt{cat}(x))\ \&\ \texttt{asleep}(x)$

If we meet an atomic description at the beginning of a discourse, there are two possible scenarios for evaluation. One is that—like Lewis suggests—we have salience information from some external sources. This means that the evaluation of a discourse may not start with an empty discourse information state. The other possibility is that, for some reason or another, we lack the salience information needed; for example, this is the opening sentence of a novel. In this case, the description is most likely to introduce a new discourse referent, behaving the same way as existential quantification. In either of the scenarios, the occurrence of x in the second formula is bound by the description operator in the first.

We have to deal with "cross-inferential" anaphora, too:

(7) *Witness*: A woman entered the house. She switched the light on. A man waited outside.
Inspector: So, the one that switched the light on was a woman.

(7a) $\langle \exists x\, \mathtt{woman}(x) \;\&\; \mathtt{enter\_the\_house}(x), \mathtt{switch\_the\_light\_on}(x), \exists y\, \mathtt{man}(y) \;\&\; \mathtt{wait\_outside}(x) \rangle \models \mathtt{woman}(\mathrm{I}z\; \mathtt{switch\_the\_light\_on}(z))$

As the example suggests, *DPLD*'s consequence relation has to make it possible for a description in its conclusion to be bound by an existential quantifier in one of its premises. The example has another peculiarity that deserves attention. The description $\mathrm{I}z\; \mathtt{switch\_the\_light\_on}(z)$ uses a predicate that was associated with the discourse referent in one sentence later than it was introduced. This is another phenomenon in the dynamics of salience ranking to be taken account of.

Finally, let us consider the behavior of a description with respect to negated sentences.

(8) John doesn't own either a car or a motorcycle. ??The car is too expensive for him.

(8a) $\sim \big((\exists x\, (\mathtt{car} \;\&\; \mathtt{own}(\mathtt{John}, x)) \vee \exists y\, (\mathtt{motorcycle} \;\&\; \mathtt{own}(\mathtt{John}, x))\big) \;\&\; \mathtt{too\_expensive}(\mathrm{I}x\; \mathtt{car}(x), \mathtt{John})$

Clearly, the description "the car" in the second sentence cannot refer back to the indefinite expression "a car" in the first. Instead, if the definite description is acceptable here at all, it introduces a new discourse referent. On the other hand, an atomic description in a negated sentence seems to be able to bind later variable occurrences:

(9) It is not the case that the cat is in the carton. She is in the garden.

(9a) $\sim \mathtt{in\_the\_carton}(\mathrm{I}x\; \mathtt{cat}(x)) \;\&\; \mathtt{in\_the\_garden}(x)$

Even if the referent of "the cat" is determined by external conditions, the pronoun "she" in the second sentence refers back to the description. Thus, negation seems to be externally dynamic with respect to atomic descriptions. This is essentially different from the analogous case with existential quantification:

(9′) It is not the case that there is a cat in the carton. ??She is in the garden.

(9a′) $\sim (\exists x\, \mathtt{cat}(x) \;\&\; \mathtt{in\_the\_carton}(x)) \;\&\; \mathtt{in\_the\_garden}(x)$

Here, the pronoun "she" appears to behave deictically, and thus the corresponding occurrence of x in the second formula is best interpreted as free.

We choose not to take this last phenomenon into account in *DPLD*. The reason is simple; no proper dynamic version of negation is known in a first-order framework. This is a fundamental problem of dynamic semantics, already observed in (Groenendijk *& al.* 1991, 99ff.), and to all appearances, it has not been solved yet.

3 INTERLUDE: THE DYNAMICS OF CLASSICAL DESCRIPTIONS

After making clear what we are interested in, a quick word on what we are not interested in. There is another way of using dynamic semantics in the analysis of definite descriptions; it concerns classical descriptions instead of Lewis's anaphoric ones. It is well-known that standard first-order logic cannot deal with Russell's analysis of descriptions in a compositional way.[3] It is less well-known that dynamic predicate logic solves this problem. This is an obvious consequence of its refined manner of handling anaphoric relations.

(10) The present king of France is bald. He shaves himself.

(10a) $\exists x \left(\forall y \left(\texttt{present\_king\_of\_France}(y) \equiv y = x\right)\right.$
$\left. \& \ \texttt{bald}(y) \ \& \ \texttt{shave}(x, x)\right)$

(10b) $\exists x \, \forall y \left(\texttt{present\_king\_of\_France}(y) \equiv y = x\right)$
$\& \ \texttt{bald}(y) \ \& \ \texttt{shave}(x, x)$

While "the present king of France" is a syntactic component in (10), its first-order translation, $\exists x \, \forall y \left(\texttt{present\_king\_of\_France}(y) \equiv y = x\right)$ is not a component in (10a). Nevertheless, it is present in the formula (10b), and in *DPL* (10a) is equivalent with (10b). Similarly, the translation of the first sentence is a component of (10b), but not of (10a). However, all occurences of x in $\texttt{bald}(y)$ and $\texttt{shave}(x, x)$ are anaphorically linked to $\exists x$. Thus, unlike *PL*, *DPL* offers a proper treatment of Russellian definite descriptions. Compositionality also makes the introduction of a classical description operator straightforward, either with a Russellian or with a Strawsonian semantics, and thus makes the semantic modelling of the dynamics of existence and uniqueness conditions straightforward. This idea is developed in (Eijck 1993), but it has little in common with our project.

4 THE SYNTAX AND SEMANTICS OF *DPL* AND *DPLR*

We are going to present dynamic predicate logic in two versions, which have the same syntax, but differ in their semantics. One of them is the well-known *DPL* presented in (Groenendijk *& al.* 1991). The second version, which we refer to as *DPLR*, differs from the original one at one important point: the use of referent systems. Referent systems were introduced in the semantics in (Groenendijk *& al.* 1996), along with the concept of a *peg*. (In fact, *DPLR* is the extensional part of the system presented in (Groenendijk *& al.* 1996).) Pegs are the technical equivalents of discourse referents; that is, abstract entities which are introduced by indefinite descriptions like "a man", and to which discourse information is attributed instead of being attributed to the variables themselves,

[3] Cf. e.g. (Gamut 1991, 164.))

or the elements of the domain of discourse. We are going to refer to them as discourse referents instead of pegs.[4]

Syntax

The syntax of *DPL* and *DPLR* is identical to that of standard predicate logic with identity. For simplicity, the only function symbols allowed are individual constants. Individual terms (henceforth, terms) are either variables or constants. Besides these, we have n-ary predicates in the non-logical vocabulary. Logical primitives are $\sim, \&, \exists, =, (,)$; te rest of the logical symbols are defined in terms of these.

i $\varphi \vee \psi \iff_{\mathrm{d}} \sim(\sim \varphi \,\&\, \sim \psi)$
ii $\varphi \supset \psi \iff_{\mathrm{d}} \sim(\varphi \,\&\, \sim \psi)$
iii $\varphi \equiv \psi \iff_{\mathrm{d}} (\varphi \supset \psi) \,\&\, (\psi \supset \varphi)$
iv $\forall x\, \varphi(x) \iff_{\mathrm{d}} \sim \exists x \sim \varphi$

Now, we define formulas as usual:

(1) If P is an n-place predicate and $t_1, \ldots, t_n$ are terms, then $P(t_1, \ldots, t_n)$ is an atomic formula.
(2) If t_1 and t_2 are terms, then $t_1 = t_2$ is an atomic formula.
(3) Nothing else is an atomic formula. Atomic formulas are formulas.
(4) If φ and ψ are formulas, then $(\varphi \,\&\, \psi)$ is a formula.
(5) If φ is a formula, then $\sim \varphi$ is a formula.
(6) If φ is a formula and x is a variable, then $\exists x\, \varphi$ is a formula.
(7) Nothing else is a formula.

The semantics of DPL

Although the orignal version of *DPL* is well-known, we reintroduce it in a nutshell. Dynamic semantics is semantics of information state updating. Discourse information is based on an ordinary first-order structure $\mathcal{M} = \langle U, \varrho \rangle$, U being the universe of discourse and ϱ being the interpretation function. If a is an individual constant and P is an n-place predicate, then $\varrho(a) \in U$ and $\varrho(P) \subseteq U^n$. This structure is not updated in the process of evaluating a discourse; it is not part of the discourse information.

We will use cylindrification as a basic operation on assignment sets, defined as

$$V[x] =_{\mathrm{d}} \{v : \text{ for some } v',\, v' \in V \text{ and } v[x]v'\},$$

[4]The term *discourse referent* was coined in (Karttunen 1975), along with the idea that discourse referents should be identified with natural numbers. We find this expression more expressive and less idiosyncratic than *peg*.

where

$$v[x]v' \iff_{\mathrm{d}} \text{ for all variables } y \text{ different from } x,\ v'(y) = v(y).$$

The value $|t|^{DPL}_{\mathcal{M},V}$ of a term t is $\varrho(t)$ if it is a constant, and $v(t)$ if it is a variable. Thus, the evaluation function $[\![\varphi]\!]^{DPL}_{\mathcal{M}}$ of formulas updates information states as follows (we omit the upper index wherever it is not confusing):[5]

(1) $[\![P(t_1,\ldots,t_n)]\!]_{\mathcal{M}}(V) =_{\mathrm{d}} \{v \in V : \langle |t_1|_{\mathcal{M},V},\ldots,|t_n|_{\mathcal{M},V}\rangle \in \varrho(P)\}$;
(2) $[\![t_1 = t_2)]\!]_{\mathcal{M}}(V) =_{\mathrm{d}} \{v \in V : \langle |t_1|_{\mathcal{M},V} = |t_2|_{\mathcal{M},V}\rangle\}$;
(3) $[\![(\varphi \ \&\ \psi)]\!]_{\mathcal{M}}(V) =_{\mathrm{d}} [\![(\psi)]\!]_{\mathcal{M}} \circ [\![(\varphi)]\!]_{\mathcal{M}}(V)$;
(4) $[\![\sim \varphi]\!]_{\mathcal{M}}(V) =_{\mathrm{d}} \left\{v \in V : [\![\varphi]\!]_{\mathcal{M}}(\{v\}) = \emptyset\right\}$;
(5) $[\![\exists x\, \varphi]\!]_{\mathcal{M}}(V) =_{\mathrm{d}} [\![\varphi]\!]_{\mathcal{M}}(V[x])$.

Finally, a dynamic consequence relation $\langle \varphi_1,\ldots,\varphi_n\rangle \models_{DPL} \psi$ is defined as

$$\langle \varphi_1,\ldots,\varphi_n\rangle \models_{DPL} \psi \iff_{\mathrm{d}} \text{ for all } \mathcal{M} \text{ and } V,$$

$$\text{if } [\![(\varphi_n)]\!]_{\mathcal{M}} \circ \cdots \circ [\![(\varphi_1)]\!]_{\mathcal{M}}(V) \neq \emptyset,$$

$$\text{then } [\![(\varphi_n)]\!]_{\mathcal{M}} \circ [\![(\varphi_n)]\!]_{\mathcal{M}} \circ \cdots \circ [\![(\varphi_1)]\!]_{\mathcal{M}}(V) \neq \emptyset.$$

That is, $\langle \varphi_1,\ldots,\varphi_n\rangle \models_{DPL} \psi$ is valid iff whenever the evaluation process of the discourse $\langle \varphi_1,\ldots,\varphi_n\rangle$ does not result in an empty set of assignments, neither does the evaluation of $\langle \varphi_1,\ldots,\varphi_n,\psi\rangle$. Note that, unlike in *PL*, the premises of a consequence relation form an ordered sequence.

The semantics of DPLR

DPLR is a variant of *DPL* that applies a more detailed model of discourse information. In *DPL*, an information state is a set of assignments, and assignments are total functions from the set of variables to the domain of discourse. In this semantic framework discourse referents do not play an explicite role. *DPLR* differs from *DPL* in making the use of discourse referents transparent, by means of referent systems.

Discourse information is represented in *DPLR* as a triple $\mathcal{I} = \langle d, r, V\rangle$. d is a natural number, and serves as the set of the discourse referents that are in use at a certain point of a discourse. As it is standard in set theory, numbers are identified with the sets of their predecessors, that is, $0 = \emptyset$ and $n = \{0,\ldots,n-1\}$. The elements of d serve as discourse referents in a given information state. r is a partial function from the set of variables to n. The pair $\langle d, r\rangle$ is called a referent

[5]Instead of updating a set of assignments as a whole, (Groenendijk & *al.* 1991) updates each assignment separately. Our version is standard in dynamic systems that give an explicit definition of information state. We first saw *DPL* presented this way in (Kálmán & *al.* 2001, 62ff.)

system. V is a set of evaluation functions from d to U. Thus, if a variable x has a value, then its value is given by $v(r(x))$ for each $v \in V$. But it is not the case that every variable has a value; in fact, the evaluation of any discourse only necessitates that a finite number of them has. A variable that does not occur in a discourse does not have a value. Discourse information is treated in a dynamic way; in the the course of evaluating a discourse, each occurrence of a term updates the actual referent system, the same way as each occurrence of a formula does.

Some technical concepts will be useful in the definition of semantics. The first is that of updating of a referent system with a new variable. This will be used in the evaluation of quantified formulas.

$$\langle d, r\rangle[x] =_{\mathrm{d}} \langle d+1, (r \setminus \{\langle x, i\rangle : i < d\}) \cup \{\langle x, d\rangle\}\rangle$$

That is, a new discourse referent is introduced into the referent system as the value of the variable x. Meanwhile, if x already had a value, the old value is deleted. A set of assignments V is updated with a new discourse referent in a similar way:

$$V[d] =_{\mathrm{d}} \left\{v : (\exists v' \in V)\, (\exists u \in U)\, v = v' \cup \{\langle d, u\rangle\}\right\}$$

(Note that unlike variables, discourse referents are not reused in the evaluation process; and although we don't rule out the possibility that different variables are associated with the same discourse referent, nothing in our semantics enforces such a situation.)

To make the following definitions more transparent, we will follow a simple notational rule for information states and first-order structures: $I = \langle d, r, V\rangle$, $I' = \langle d', r', V'\rangle$, $I'' = \langle d'', r'', V''\rangle$ etc; and $\mathcal{M} = \langle U, \varrho\rangle$, $\mathcal{M}' = \langle U', \varrho'\rangle$ etc. This way it will be easy to identify the semantic components.

A discourse information state $\mathcal{I}$ is updated by updating both $\langle d, r\rangle$ and V:

$$\mathcal{I}[x] =_{\mathrm{d}} \langle \langle d, r\rangle[x], V[d]\rangle^{6}$$

We call a discourse information state *empty* iff it has an empty set of assignments. An empty information state may or may not have an empty referent system.

The value $|t|^{DPLR}_{\mathcal{M},\mathcal{I}}$ of a term t is $\varrho(t)$ if it is a constant, and $v(r(t))$ if it is a variable with a previous occurrence. Variables with no previous occurrences have no value. Thus, the evaluation function $[\![\varphi]\!]^{DPLR}_{\mathcal{M}}$ of formulas is essentially partial. (We omit the upper indices wherever it is not confusing.) Formulas with semantically free variables have no semantic value. The semantic value of a formula, if it exists, is a function from discourse information states to discourse information states, defined in the following clauses:

[6] As is customary, for convenience, we identify $\langle\langle a, b\rangle, c\rangle$ with $\langle a, \langle b, c\rangle\rangle$. Also, we identify $\langle a\rangle$ with a.

(1) $[\![P(t_1,\ldots,t_n)]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}}$
$\langle d, r, \{v \in V : \langle |t_1|_{\mathcal{M},\mathcal{I}}, \ldots, |t_n|_{\mathcal{M},\mathcal{I}}\rangle \in \varrho(P)\}\rangle$;
(2) $[\![t_1 = t_2)]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} \langle d, r, \{v \in V : \langle |t_1|_{\mathcal{M},\mathcal{I}} = |t_2|_{\mathcal{M},\mathcal{I}}\rangle\}\rangle$;
(3) $[\![(\varphi \;\&\; \psi)]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} [\![(\psi)]\!]_{\mathcal{M}} \circ [\![(\varphi)]\!]_{\mathcal{M}}(I)$;
(4) $[\![\sim \varphi]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} \left\langle d, r, \left\{v \in V : [\![\varphi]\!]_{\mathcal{M}}(\langle d, r, \{v\}\rangle) \text{ is empty}\right\}\right\rangle$;
(5) $[\![\exists x\, \varphi]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} [\![\varphi]\!]_{\mathcal{M}}(\mathcal{I}[x])$.

Finally, the definition of dynamic consequence is as follows:

$\langle \varphi_1, \ldots, \varphi_n\rangle \models \psi$ iff for all $\mathcal{M}$ and $\mathcal{I}$, if

$$[\![\psi]\!]_{\mathcal{M}} \circ [\![\varphi_n]\!]_{\mathcal{M}} \circ \cdots \circ [\![\varphi_n]\!]_{\mathcal{M}}(\mathcal{I})$$

exists and

$$[\![\varphi_n]\!]_{\mathcal{M}} \circ \cdots \circ [\![\varphi_n]\!]_{\mathcal{M}}(\mathcal{I})$$

is nonempty, then

$$[\![\psi]\!]_{\mathcal{M}} \circ [\![\varphi_n]\!]_{\mathcal{M}} \circ \cdots \circ [\![\varphi_n]\!]_{\mathcal{M}}(\mathcal{I})$$

is nonempty.

The problem of partiality

We have seen how *DPLR* makes the semantic role of a discourse referent explicit. The price of this is partiality. In *DPLR*, a semantically free variable occurrence—that is, an occurrence of a variable that has not been associated with a discourse referent—results in a semantic value gap. This is not necessarily a problem. As it turns out from the above definitions, the evaluation process of a discourse $\langle \varphi_1, \ldots, \varphi_n\rangle$ does not necessarily starts with an empty referent system. If a variable x occurs in a discourse without an antecedent $\exists x$, it may still have a value, provided by the initial information state. This way *DPLR* gives an account of the deictic use of pronouns of natural language.

However, a natural language discourse does not always come to a halt when a new pronoun is introduced without salient discourse referent. It is more likely that a new discourse referent is tacitly introduced. This resembles the case of using an atomic description without any salient referent, like in example (6) above. For this reason, in *DPLD* we do not follow (Groenendijk *& al.* 1996) in making semantic rules partial.

5 THE SYNTAX AND SEMANTICS OF *DPLD*

In this section we introduce *DPLD*, a system of dynamic predicate logic with atomic descriptions, as a modified version of *DPLR*. The main differences are the use of the descriptor in the syntax, and salience ranking as a component of discourse information in semantics.

Syntax

In all our examples, we used one-place predicates in the descriptions. We restrict ourselves to them in the syntax, because we do not intend to model the dynamics of embedded and open descriptions, like $\mathrm{I}x\ R(x, \mathrm{I}y\ P(y))$ ("the man's friend") and $\mathrm{I}x\ R(x, y)$ ("his friend"). Adapting our semantic rules to atomic descriptions with many-place predicates is straightforward but slightly complicated.

We define terms by the following clauses:

(1) Variables and individual constants are terms.

(2) If P is a one-place predicate and x is a variable, then $\mathrm{I}x\ Px$ is a term.

The rest of the syntax remains intact.

Semantics

As in *DPLR*, the definition of semantics begins with the concept of a discourse information state. We use referent systems and assignment sets as in *DPLR* (but as we will see, they are updated differently). We enrich discourse information states with a new component, salience ranking; $\mathcal{I} = \langle d, r, S, V \rangle$, where S is a salience ranking.

A salience ranking is an ordered tuple of salience information bits. Salience bits, on their turn, are ordered pairs. If P is a one-place predicate and d is a discourse referent, then $\langle P, d \rangle$ is a salience bit. (Note that salience bits have both syntactic and semantic elements.) Let us explain its use through simple examples.

Let P be a one-place predicate, d a discourse referent, and let the salience information bit $s = \langle P, d \rangle$ be at the head of our actual salience ranking. Now, if a description $\mathrm{I}x\ Px$ occurs in the course of evaluation, then our actual referent system $\langle d', r \rangle$ is updated with the pair $\langle x, d \rangle$. On the other hand, if we find a number of bits of the form $s = \langle P, d \rangle$ in our actual salience ranking, then the referent system is updated with the discourse referent in the leftmost bit. And finally, if there is no bit of the form $\langle P, d \rangle$ on the list, then x is associated with a new discourse referent, just like in the case of *DPL*'s existential quantification.

The formal definition of updating a referent system is

$$\langle d, r \rangle [x/d'] =_{\mathrm{d}} \langle \mathsf{max}(d, d' + 1), r \setminus \{\langle x, i \rangle : i < d\}) \cup \{\langle x, d' \rangle\} \rangle$$

The difference between this update definition and the one given in the last section is that x is not necessarily associated with a new discourse referent. As a consequence, different variables can be associated with the same discourse referent. An occurrence of $\mathrm{I}x\ P(x)$ can refer back to $\exists y\, P(y)$ in this way; although x is not bound by $\exists y$, x and y are associated with the same discourse referent, and hence they have the same value.

Let, again, P be a one-place predicate, and t a term associated with the discourse referent d; that is, $r(t) = d$. Let the atomic formula $P(t)$ occur at a certain point of the discourse. Then the salience ranking S is updated with a new bit $\langle P, d\rangle$; that is, this pair is attached to the head of the ranking:

$$S[P(t)] =_{\mathrm{d}} \left\langle \langle P, r(t)\rangle, S\right\rangle$$

Since the last information bit is always the leftmost one, the discourse referent this bit offers is more salient than its rivals that are offered further down in the ranking. In simple terms, the last thing mentioned is the most salient.

It will be useful to define a function $S(P)$ that gives back the most salient discourse referent associated with a predicate P in a given salience ranking S. If there is no salient discourse referent, then the value of $S(P)$ is P. We define $S(P)$ recursively with respect to the length of S.

(1) If $S = \emptyset$, then $S(P) = P$;
(2) if $S = \left\langle \langle P, d\rangle, S'\right\rangle$ for some d and S', then $S(P) = d$;
(3) if $S = \left\langle \langle Q, d\rangle, S'\right\rangle$ for some Q that is different from P, d and S', then $S(P) = S'(P)$.

The last component of a discourse information state is a set V of assignments. It is updated with a discourse referent the same way as in *DPL*. There are two ways of updating the discourse information state with a new variable x. In the first case, x is associated with an existing discourse referent:

$$\mathcal{I}[x/d'] =_{\mathrm{d}} \langle \langle d, r\rangle [x/d'], S, V\rangle$$

In the second case, a new discourse referent is introduced along with a new variable:

$$\mathcal{I}[x] =_{\mathrm{d}} \langle \langle d, r\rangle [x/d], S, V[d]\rangle$$

We have seen that terms are capable of updating discourse information. Now we define this update function $[\![t]\!]^{DPLD}_{\mathcal{M}}$.

(1) $[\![x]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} \begin{cases} \mathcal{I}[x] & \text{if } x \notin \mathsf{dom}(r); \\ \mathcal{I} & \text{if } x \in \mathsf{dom}(r); \end{cases}$
(2) $[\![a]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} \mathcal{I}$;
(3) $[\![\mathrm{I}x\ P(x)]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} \begin{cases} \mathcal{I}[x] & \text{if } S(P) = P; \\ \mathcal{I}[x/d'] & \text{if } S(P) = d'. \end{cases}$

That is, an occurrence of a variable x that has not yet been associated with a discourse referent behaves like an existential quantifier in the sense that it introduces a new discourse referent. Occurrences of x that are already associated with a discourse referent leave discourse information unchanged. Individual constants do not change discourse information either. An occurrence of an atomic description $\mathrm{I}x\ P(x)$ for which there is no salient referent introduces

a new one. Otherwise the x of $\mathrm{I}x\ P(x)$ is associated with the most salient discourse referent.

The value $|t|^{DPLD}_{\mathcal{M},\mathcal{I}}$ of a term t is defined as follows:

(1) $|a|_{\mathcal{M},\mathcal{I}} =_{\mathrm{d}} \varrho(t)$;
(2) $|x|_{\mathcal{M},\mathcal{I}} =_{\mathrm{d}} v(r(x))$;
(3) $|ixP(x)|_{\mathcal{M},\mathcal{I}} =_{\mathrm{d}} v(r(x))$.

Thus every occurrence of any term has a value in *DPLD*. Unlike *DPLR*, the semantics of *DPLD* is not a partial one.

Now we are ready to evaluate formulas. The information update function $[\![\varphi]\!]^{DPLD}_{\mathcal{M}}$ is defined in the following clauses:

(1) $[\![P(t)]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} \langle d', r', S'[P(t)], \{v \in V' : |t|_{\mathcal{M},\mathcal{I}} \in \varrho(P)\}\rangle$,
where $\langle d', r', S', V'\rangle = [\![t]\!]_{\mathcal{M}}(\mathcal{I})$;
(2) $[\![P(t_1, \ldots, t_n)]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}}$
$\langle d', r', S', \{v \in V' : \langle |t_1|_{\mathcal{M},\mathcal{I}}, \ldots, |t_n|_{\mathcal{M},\mathcal{I}}\rangle \in \varrho(P)\}\rangle$,
where $n > 1$ and $\langle d', r', S', V'\rangle = [\![t_n]\!]_{\mathcal{M}} \circ \cdots \circ [\![t_1]\!]_{\mathcal{M}}(\mathcal{I})$;
(3) $[\![t_1 = t_2)]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} \langle d', r', S', \{v \in V' : \langle |t_1|_{\mathcal{M},\mathcal{I}} = |t_2|_{\mathcal{M},\mathcal{I}}\rangle\}\rangle$,
where $\langle d', r', S', V'\rangle = [\![t_2]\!]_{\mathcal{M}} \circ [\![t_1]\!]_{\mathcal{M}}(\mathcal{I})$;
(4) $[\![(\varphi\ \&\ \psi)]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} [\![(\psi)]\!]_{\mathcal{M}} \circ [\![(\varphi)]\!]_{\mathcal{M}}(I)$;
(5) $[\![\sim \varphi]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} \left\langle d, r, S, \left\{v \in V : [\![\varphi]\!]_{\mathcal{M}}(\langle d, r, S, \{v\}\rangle) \text{ is empty}\right\}\right\rangle$;
(6) $[\![\exists x\, \varphi]\!]_{\mathcal{M}}(\mathcal{I}) =_{\mathrm{d}} [\![\varphi]\!]_{\mathcal{M}}(\mathcal{I}[x])$.

Thus, the definition deviates from the one of *DPL* only in the evaluation of atomic formulas. Dynamic inference is defined the same way as in *DPL*.

To see how discourse information updating works, it will be instructive to see the evaluation of a particular formula. Let us consider the first three sentences of example (5). To make the formula shorter, we abbreviate the predicates `man`, `walk_in_the_park`, `woman`, `meet` and `hug` as P, Q, T, R and S, respectively.

(5′) A man walks in the park. He meets a woman. The man hugs her.
(5a′) $\exists x\, P(x)\ \&\ Q(x)\ \&\ \exists y\, T(y)\ \&\ R(x, y)\ \&\ S(\mathrm{I}w\ P(w), y)$

We present the evaluation process of the formula in an informal fashion.

(1) We start with an empty referent system, an empty salience ranking and an empty assignment set. That is, we do not make use of an external context.
(2) $\exists x\, P(x)$ introduces a new discourse referent, and associates x with it. Salience information associates the new discourse referent with the predicate P. $d_1 = 1$; $r_1 = \{\langle x, 0\rangle\}$; $S_1 = \langle P, 0\rangle$; $V_1 = \{\{\langle 0, u\rangle\} : u \in \varrho(P)\}$.
(3) $Q(x)$ updates salience information. $d_1 = d_1$; $r_2 = r_1$; $S_2 = \langle\langle Q, 0\rangle, < P, 0\rangle\rangle$; $V_1 = \{\{\langle 0, u\rangle\} : u \in \varrho(P), u \in \varrho(Q)\}$.
(4) $\exists y\, T(y)$ introduces a new discourse referent, associates y with it, and updates salience information. $d_3 = 2$; $r_3 = \{\langle x, 0\rangle, \langle y, 1\rangle\}$;

$S_3 = \langle\langle T, 1\rangle, \langle Q, 0\rangle, \langle P, 0\rangle\rangle$; $V_3 = \{\{\langle 0, u\rangle, \langle 1, u'\rangle\} : u \in \varrho(P), u \in \varrho(Q), u' \in \varrho(T)\}$.

(5) $R(x, y)$ does not change either the referent system or the salience ranking. $d_4 = d_3$; $r_4 = r_3$; $S_4 = S_3$; $V_4 = \{\{\langle 0, u\rangle, \langle 1, u'\rangle\} : u \in \varrho(P), u \in \varrho(Q), u' \in \varrho(T), \langle u, u'\rangle \in \varrho(R)\}$.

(6) Iw $P(w)$ introduces a new variable in the referent system. Since P occurs only once in S_4, this occurrence determines the discourse referent associated with w: $r_5(w) = S_4(P) = 0$. Thus, $d_5 = d_4$; $r_5 = \{\langle x, 0\rangle, \langle y, 1\rangle, \langle w, 0\rangle\}$; $S_5 = S_4$; $V_5 = V_4$.

(7) Finally, $S(\text{I}w\ P(w), y)$ does not change either the referent system or the salience ranking. $d_6 = d_5$; $r_6 = r_5$; $S_6 = S_5$; $V_6 = \{\{\langle 0, u\rangle, \langle 1, u'\rangle\} : u \in \varrho(P), u \in \varrho(Q), u' \in \varrho(T), \langle u, u'\rangle \in \varrho(R), \langle u, u'\rangle \in \varrho(S)\}$.

REFERENCES

Eijck, J. van, 1993, The dynamics of description. *Journal of Semantics* 10, 223–267.

Eijck, J. van and Kamp, H., 1997, Representing discourse in context. In J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*. Elsevier, Amsterdam & MIT Press, Cambridge (MA), 179–237.

Gamut, L. T. F., 1991, *Logic, Language, and Meaning*. University of Chicago Press, Chicago. (L. T. F. Gamut is a collective pseudonym for J. van Benthem, J. Groenendijk, D. de Jongh, M. Stokhof and H. Verkuyl.)

Groenendijk, J. and Stokhof, M., 1991, Dynamic predicate logic. *Linguistics and Philosophy* 14, 39–100.

Groenendijk, J., Stokhof, M. and Veltman, F., 1996, Coreference and modality. In S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*, Blackwell, 179–214.

Kálmán, L. and Rádai, G., 2001, *Dinamikus szemantika* (*Dynamic Semantics*; in Hungarian). Budapest, Osiris.

Karttunen, L., 1975, Discourse referents. In J. McCawley (ed.), *Syntax and Semantics 7: Notes from the Linguistic Underground*. Academic Press, New York, 363–385.

Lewis, D., 1979, Scorekeeping in a language game. *Journal of Philosophical Logic* 8, 339–359.

Ruzsa, I., 1981, *Modal Logic with Descriptions*. Martinus Nijhoff, De Hague.

Zoltán Gendler Szabó

Tasks and Ultra-tasks

Abstract. Can we count the primes? There is a near unanimous consensus that in principle we can. I believe the near-consensus rests on a mistake: we tend to confuse counting the primes with counting each prime. To count the primes, I suggest, is to come up with an answer to the question "How many primes are there?" because of counting each prime. This, in turn requires some sort of dependence of outcome on process. Building on some ideas from Max Black, I argue that—barring very odd laws of nature—such dependence cannot obtain.

1 COUNTING THE PRIMES

There are countably many primes but this does not settle the question whether the primes can be counted. "Countably infinite" is a technical term which applies to multitudes equinumerous with the natural numbers. The question I want to pursue here is whether any such multitude can be counted. I will talk about counting the primes but that is just an example—I could talk about counting any countably infinite set.

The obvious difficulty with counting the primes is that even if you could go on forever, it seems there would always remain more to count. But what if you could count faster and faster as you proceed? Then, we are taught, there would be no problem. You could call the first prime within the first 1 minute, the second within the next 1/2 minute, the third within the next 1/4 minute, … the nth within the next $1/2^{n-1}$ minute, … and so on. After 2 minutes you would have counted *each prime*. But would you thereby have counted *the primes*?

There is a difference between counting each prime and counting the primes. If you count the natural numbers you have counted each prime but you have not counted the primes. In general, if all *F*s are *G*s, then in counting the *G*s you count each *F* but not necessarily the *F*s. Should we perhaps say that counting the *F*s is counting each *F* without counting *non-F*s? This won't do, as the following

example illustrates. Suppose there is a birthday party at your house for your six year old and you ask him to count his guests. He counts them in the living room, in the dining room, in the kitchen, and so on for all the rooms of the house. He is lucky—the children don't move around and he doesn't miss any. He is also smart—knowing that he is not too good with large numbers, whenever he is done with a room he writes down the result, moves to the next room, and starts the count from 1 again. Now suppose you asked him after all the numbers are written down but before they are added up "Did you count the guests?" The natural response would be "Not yet." Each and only the guests have been counted but the guests have not been counted.

Why not? Because the child doesn't know how many guests there are, one might think. But this is not exactly right: coming to know the number of guests is neither necessary nor sufficient for counting them. If the child makes a mistake in adding up the numbers on the paper he will have counted the guests without coming to know their number; if he is told how many guests there are before he adds up the numbers on the paper he will know the number of guests without having counted them. To count the *F*s, I suggest, is to come up with some sort of (perhaps incorrect, perhaps poorly justified) answer to the question "How many *F*s are there?" *as a result of* counting each *F*.[1] Counting each guest in the usual way *leads to* an answer to the question "How many guests are there?"; counting each guest the way the child did it does not.

I believe the near-consensus that the primes could in principle be counted rests on a mistake: we tend to confuse counting the primes with counting each prime. It may be obvious that we can count each prime in the exponentially accelerating way described above but it is not obvious that this process leads to an answer to the question "How many primes are there?". When we count the primes below 1000 in the usual way, at each step we come to have an answer to the question "How many primes have you already counted?" When we reach what we take to be the largest prime below 1000 we *ipso facto* arrive at an answer to the question "How many primes are there below 1000?" But when we count each prime there is no last step in the count. Accordingly, it is anything but obvious that we can come to have an answer to the question "How many primes are there?" simply as a result of counting each prime one after the other.

The expressions "as a result of" and "leads to" are not among the clearest in our philosophical vocabulary. If I had an analysis of the concepts they express, I would gladly dispense with them. Unfortunately, I have no analysis and I am not optimistic about finding one anytime soon. (In section 3, I will suggest a necessary condition for a process to lead to a state.) One thing is clear: these

[1] This is compatible with the possibility of counting the *F*s when one already has an answer to the question "How many *F*s are there?" We can come up with an answer to a question as a result of a procedure even if we already have an answer to that question.

locutions express some sort of dependence, presumably a causal one, of outcome on process. I will argue that the outcome of an infinite count (having an answer to a certain cardinality question) cannot depend in the right way on the process of counting. If I am right, the conventional wisdom is wrong: the primes (or, for that matter, any other countably infinite multitude) cannot be counted.

2 BLACK'S ARGUMENT

More than half a century ago, Max Black argued that certain infinitary tasks are impossible to perform (Black 1951). When Black's article originally appeared, it produced a flurry of reactions, including Thomson 1954, where the famous lamp is discussed. (Benacerraf 1962) was seen as a decisive refutation of Black's argument, which in turn has largely faded from philosophical discussion. I think the dismissal was too quick: there is something important Black was right about, though it is not exactly what he thought he was right about.

The target of Black's criticism is the standard "mathematical" resolution of Zeno's paradox. According to this line of thought, Achilles can catch up with the tortoise by reaching the place p_0 where the tortoise had started the race, then reaching the place p_1 the tortoise had reached by the time Achilles reached p_0, then reaching the place p_2 the tortoise had reached by the time Achilles reached p_1, … and so on. This amounts to performing an infinite number of tasks: moving first to p_0, then from p_0 to p_1, then from p_1 to p_2, … and so on. If Achilles runs faster than the tortoise, the amount of time it takes to perform all these tasks one after the other is finite. Hence, we are told, there is no problem with catching up with the tortoise in this manner.

Black disagrees. His central complaint is that the "mathematical" resolution of the paradox "tells us, correctly, when and where Achilles and the tortoise will meet, *if* they meet; but it fails to show that Zeno was wrong in claiming they *could not* meet" (Black 1951, 93). Not that anyone should doubt that Achilles can catch up with the tortoise – Black certainly does not. His question is whether Achilles can catch up with the tortoise by performing an infinite series of tasks. Black thinks there is a serious difficulty with that idea "and it does not help to be told that the tasks become easier and easier, or need progressively less and less time in the doing" (Black 1951, 94).

What is the difficulty? Black asks us to imagine a mechanical scoop with an infinitely long narrow tray to its left and another to its right. The scoop is capable of moving marbles from one tray to the other at any finite speed. Let machine *Alpha* work as follows: Initially there are infinitely many marbles in the left tray and the right tray is empty. *Alpha* moves the first marble from left to right in 1 minute and then rests for 1 minute, it moves the second marble from left to right in 1/2 minute and then rests for 1/2 minute, it moves the third marble from left

to right in 1/4 minute and then rests for 1/4 minute, … , it moves the *n*th marble from left to right in $1/2^{n-1}$ minute and then rests for $1/2^{n-1}$ minute, … and so on. After 4 minutes, *Alpha* stops. Black argues that it is impossible for *Alpha* to move the marbles from left to right.

Since there is nothing special about moving marbles, if Black is right about *Alpha* then all sorts of infinitary tasks are impossible to perform. Imagine that whenever *Alpha* moves a marble from left to right, Achilles traverses one of the intervals mentioned in the "mathematical" resolution to the paradox. If *Alpha* cannot move the marbles left to right by moving them one by one, then Achilles can presumably also not catch up with the tortoise by traversing the intervals one after the other. *Alpha* is a good stand-in for counting the primes as well. Imagine that a prime is written on each of the marbles and that we call a prime when *Alpha* moves the appropriate marble from left to right. If the primes can be counted at all, then presumably they can be counted in this way—at least, those who think otherwise should explain why counting the primes by calling each prime in an exponentially accelerating way is possible, but moving the marbles from one tray to another by moving each marble in the same exponentially accelerating way is not.[2] I will assume that if *Alpha* cannot move the marbles from left to right, then the primes cannot be counted either.

So why does Black think that *Alpha* cannot move the marbles from left to right? The argument is presented in a somewhat oblique fashion and it can be interpreted in more than one way; I will give my best and most charitable reconstruction. The first thing to note is that in order to move the marbles from left to right, *Alpha* has to bring it about that at some time, the marbles are on the right. (Suppose it is your custom to remove no more than a couple of books at a time from your library to your study. In the evening, you always return them. Still, over the years you have carried each of the hundreds of books from your library into the study at one time or other. Then saying "I have moved the books from the library to the study" would be false, given the most natural reading of this sentence. Similarly, if the marbles are never all on the right then *Alpha* did not move them to the right.[3]) The second thing to note is that *Alpha* cannot bring it about that the marbles are all on the right unless it brings that about

[2] It is not enough to point out that counting is a mental process, and moving marbles a physical one. Some dualists believe that there are mental processes fundamentally different from all physical ones. Even if they are right, it does not follow that infinite counts are among the mental processes that lack a physical analogue. It is beyond doubt that one can model finite counts with finite marble transfers—why think that we cannot model infinite counts with infinite marble transfers?

[3] I don't deny that "move the marbles from left to right" has a distributive reading that is synonymous with "moving each of the marbles from left to right," just as "counting the guests" has a distributive reading that is synonymous with "counting each of the guests." But the dominant readings are the collective ones, and those are the ones that I am concerned with.

when it stops. (Before *Alpha* stops there are always marbles on the left. After it stops it cannot bring anything about.) It follows that *Alpha* cannot move the marbles from left to right unless it can bring it about that they are all on the right when it stops. Black seeks to show that Alpha cannot bring this about.

Black goes on to consider two other machines, *Beta* and *Gamma*. They are similar to *Alpha* but they share their trays and they work in tandem: Initially there is a single marble in the left tray and the right tray is empty. *Beta* moves the marble from left to right in 1 minute while *Gamma* rests, then *Gamma* moves the marble from right to left in 1 minute while *Beta* rests, then *Beta* moves the marble from left to right in 1/2 minute while *Gamma* rests, then *Gamma* moves the marble from right to left in 1/2 minute while *Beta* rests, … , then *Beta* moves the marble from left to right in $1/2^n$ minute while *Gamma* rests, then *Gamma* moves the marble from right to left in $1/2^n$ minute while *Beta* rests, … and so on. After 4 minutes, *Beta* and *Gamma* stop.

We can think of *Alpha*, *Beta* and *Gamma* as having the same kind of task: to bring it about that all the marbles they are working with are on one side when they stop. *Alpha* accomplishes its task just in case it brings it about that the infinitely many marbles that are initially on the left are all on the right when it stops, *Beta* accomplishes its task just in case it brings it about that the single marble that is initially on the left is on the right when it stops, and *Gamma* accomplishes its task just in case it brings it about that the single marble that is initially on the left (but is subsequently moved to the right by *Beta*) is on the left when it stops. Now, we can reason as follows:

(1) Necessarily, *Alpha* accomplishes its task if and only if *Beta* does.
(2) Necessarily, *Beta* accomplishes its task if and only if *Gamma* does.
(3) Necessarily, either *Beta* or *Gamma* does not accomplish its task.

Therefore,

(4) Necessarily, *Alpha* does not accomplish its task.

Alpha's task was to bring it about that the marbles are on the right when it stops. Since it cannot accomplish this task, it cannot bring it about the marbles are on the right, and hence, it cannot move the marbles from left to right. This is Black's argument, as I understand it.

3 IN SUPPORT OF THE ARGUMENT

The argument is clearly valid; the question is whether it is sound. Black supports (1) by emphasizing that *Alpha* and *Beta* are intrinsically the same: if you put them side by side their moves would be entirely parallel. The difference is

merely that while *Alpha* transfers an infinite number of qualitatively identical marbles, *Beta* transfers the same marble an infinite number of times. Something similar can be said in defense of (2). While there certainly is an intrinsic difference between *Beta* and *Gamma* – after a move that took $1/2^n$ minute the former rests $1/2^n$ minute, while the latter only $1/2^{n+1}$ minute – this seems negligible in the light of the fact the two machines go through an identical sequence of pairs of moves and rests. The difference is in the direction of the moves and in the order within the pairs: *Beta* moves first and rests afterwards, while for *Gamma* it is the other way around. Regarding (3), Black simply points out that after 4 minutes the marble must end up somewhere, so either *Beta* or *Gamma* must fail to accomplish its task.

How strong these considerations are depends on what kind of necessity is at stake. Physical necessity won't do—it makes the premises as well as the conclusion trivial. The workings of *Alpha*, *Beta*, and *Gamma* are all physically impossible: they require motion faster than the speed of light. Black is clear that he has logical necessity in mind, but that won't do either. It is not logically necessary for the marble to end up in just one of the trays—bilocation of an object is not only possible but perhaps even actual at the micro-physical level.

So Black's argument is uninteresting when the modality is construed as physical necessity and unsound when it is construed as logical necessity. But ordinarily when we wonder whether *Alpha* can accomplish its task we have neither the physical not the logical interpretation of the modal auxiliary in mind—we wonder whether there are more or less homely possible worlds where *Alpha* succeeds. The fact that nothing moves faster than the speed of light in the actual world and the fact that marbles are at different places at the same time in some remote possible worlds are irrelevant to the question we are after. This is not a special feature of the problem at hand: in general when we are asked whether something can be done we are supposed to ignore parochial limitations and far-fetched possibilities.[4] I interpret the question whether *Alpha* can move the marbles from left to right as asking whether it is possible—as we ordinarily understand what is possible in the sorts of contexts set by the description of the machine—for *Alpha* to accomplish its task.

Given the ordinary understanding of necessity (3) is in good shape: if *Beta* and *Gamma* both accomplish their tasks, then the marble is both on the left and on the right, which is certainly a far-fetched possibility. But the other two premises remain problematic even under the ordinary construal of the modality. The problem is that as long as the machines operate independently of each

[4] Suppose someone asks you the following: "Can you swim across this river?" If you say "No, I don't have my goggles with me" you misconstrue the question by failing to ignore parochial limitations. If you say "Yes, but I would need to learn how to swim first" you misconstrue the question by failing to ignore remote possibilities. What counts as parochial limitation or remote possibility depends on the context.

other, something can interrupt the working of one but not the other. It could happen, for example, that before the full 4 minutes elapse someone crushes *Beta* with a hammer. Then *Beta* surely does not accomplish its task but we are given no reason to think that *Alpha* fails too. Alternatively, it might be that while *Gamma* is making one of its moves and *Beta* is at rest, the latter is replaced by a duplicate machine which picks up the moving of the marble where *Beta* left off. Then *Beta* does not accomplish its task but for all we know *Gamma* might. To do justice to the intuition behind Black's argument, we must ignore not only far-fetched possible worlds but also nearby ones where something interferes with the proper functioning of the machines.

Suppose we do that; then the following case can be made for (2). If the machines work uninterrupted, the case of *Beta* and *Gamma* exhibits global symmetry. If the world is not too far-fetched, we should not expect anything to break this symmetry. While the description of the case leaves it open where the marble is located after the full 4 minutes has elapsed, reasonable guesses do not privilege any particular position: the chance that the marble should end up on the left equals the chance that it should end up on the right. Suppose then that the marble ends up on the right. Then *Gamma*, the machine whose task it was to bring it about that the marble ends up on the left, has failed. But *Beta* has not succeeded either. *Beta*'s moves did not raise the chances of the marble ending up on the right above the chances of it *not* ending up there, so the moves did not *bring it about* that the marble ended up on the right.[5] If we assume the marble ends up on the left (or somewhere other than the left or right trays), an analogous argument shows that neither *Beta* nor *Gamma* accomplishes its task.

I think these considerations show that, if we interpret the modality appropriately, premises (2) and (3) of Black's argument come out true: in nearby worlds where their work remains uninterrupted, neither *Beta* nor *Gamma* succeeds. However, many would insist, the case of *Alpha* is different. *Beta*'s work is constantly undone by *Gamma*, and vice versa, while nothing whatsoever interferes with the work of *Alpha*. This might explain how *Beta* and *Gamma* could fail while *Alpha* succeeds. While I don't think this is correct, I agree that (1) is in need of support.

We need a way to bolster the intuition that the makeup and task of *Alpha* and *Beta* are sufficiently similar that if they both work uninterrupted, the possibilities where one succeeds without the other are far-fetched. Here is a somewhat analogous problem. Consider two pebbles *Aleph* and *Beth*, the former being heavier than the latter. Suppose we drop both at the same time from the same height. There is a powerful intuition that *Aleph* will reach the ground

[5] I am assuming here only a necessary (but not sufficient) condition for a process bringing about a state: if the process *P* brings about the state *S*, then the objective chance of *S* holding after *P* is higher than the objective chance of *S* not holding after *P*.

before *Beth*. To combat this intuition, Galileo invited us to consider a case where we tie *Aleph* and *Beth* together and drop them from the same height. Call the tied up object *AlephBeth*. *AlephBeth* can be seen in two different ways—as a unit or as two separate objects. If heavier objects in general fall faster than lighter ones, *AlphBeth* as a unit would have to fall faster than *Aleph* (since its weight exceeds the weight on *Aleph*) but *AlephBeth* as two separate objects would have to fall slower than *Aleph* (since it is held back by the slower *Beth*). But it make no real difference whether we consider *AlephBeth* as one object or two, and so the intuition that *Aleph* reaches the ground before *Beth* must be mistaken.[6]

I will try to follow in Galileo's footsteps and argue that the intuition that *Alpha* might succeed while *Beta* fails is similarly mistaken. Consider *AlphaBeta*, a machine that works as follows. There are two trays on its left—one above the other—and a single tray on its right. *AlphaBeta* moves marbles from the lower left hand tray to the right hand tray. Initially there are infinitely many marbles in the upper left tray, a single marble in the lower left tray, and no marble on the right. Whenever *AlphaBeta* moves a marble, a hole opens up at the end of the upper tray and a single marble drops into the lower tray. *AlphaBeta* moves the marbles in the same pattern as *Alpha* and *Beta*, and it has the same task: to bring it about that all the marbles it is working with are on the right when it stops. It seems to me that (again, ignoring far-fetched possibilities and interruptions) the following claims are true:

(1') Necessarily, *AlphaBeta* accomplishes its task if and only if *Alpha* does.
(1'') Necessarily, *AlphaBeta* accomplishes its task if and only if *Beta* does.

One way to think about *AlphaBeta* is this: it moves infinitely many marbles from left to right in exactly the same pattern as *Alpha*. It makes no real difference whether the marbles that are waiting to be moved on the left will have dropped a bit just before the scoop picks them up. Another way to think about *AlphaBeta* is this: it moves a marble from the lower left tray to the right and while it rests, a marble shows up in the lower left tray, and this happens infinitely many times in the same pattern as with *Beta*. It makes no real difference whether the marble that shows up after each move is taken from the right hand tray or from some other place. But if both ways of thinking are legitimate then (1') and (1'') are true, and since they jointly entail (1), we now have an intuitive justification for the first premise.

In response, one could point out that there are possible laws that could guarantee that one of these machines fails without guaranteeing that the other does. For example, it could be a law that although things can move at any finite

[6] The idea of thinking about Galileo's thought experiment along these lines is from (Gendler 1998).

speed horizontally, they cannot move faster than the speed of light vertically; in a world governed by such a law *AlphaBeta* fails but *Alpha* might still succeed. Alternatively, it could be a law that after a single object oscillates between two locations infinitely many times, it goes out of existence; in a world with such a law, *Beta* fails but *AlphaBeta* might still succeed. But I assume that worlds governed by such recherché laws are beyond the scope of worlds we consider in making ordinary judgments of necessity. Note that Galileo's argument is subject to similar objections: it could be a law that whenever two objects are tied together they fuse into an extended simple, or go out of existence. All we can say in defense of the Galilean thought experiment is that such laws are far-fetched enough to be properly ignored.

I concede that there is a fairly natural law that distinguishes between *AlphaBeta* and *Beta*: the law of the continuity of motion. Let's assume that the law holds in an all nearby worlds. It follows that in nearby worlds, at the moment *Beta* and *Gamma* stop, the marble goes out of existence. (For if it existed, it would presumably have to be at a single location *l*. And at times arbitrarily close to the time when *Beta* and *Gamma* stopped, the marble had been at some fixed distance from *l*, which violates the continuity of motion.) On the other hand, there seems to be nothing that forces any of the marbles *AlphaBeta* moved from the left to the right to go out of existence. So, one might argue, in some nearby worlds *AlphaBeta* succeeds even though *Beta* fails, and so (1'') is false. But the last step in this reasoning is fallacious: even if we grant that when *AlphaBeta* stops the marbles are all on the right, *it does not follow* that *AlphaBeta* accomplished its task. There is a logical gap between the claim that *AlphaBeta* moved each marble from left to right and the claim that it moved the marbles from left to right (just as there is a logical gap between the claim that someone counted each guest and the claim that he counted the guests). Given the intuitive plausibility of (1''), I suggest that in nearby worlds *AlphaBeta* fails even though when it stops, the marbles are all on the right. So, accepting the claim that motion is necessarily continuous does not undermine the argument.

Before moving on, I'd like to restate Black's argument in terms of objective chances; this brings out the role of considerations about laws. Let P_x be a proposition describing the process machine x goes through and let S_x be a proposition describing the state the marbles must be in when x stops if x accomplishes its task. If a state is a result of a process, then the chances of the state holding must be higher than it not holding, given the process.[7] This means that *Alpha* cannot accomplish its task unless:

(5) $$Pr\,(S_{Alpha} \mid P_{Alpha}) > Pr\,(\neg S_{Alpha} \mid P_{Alpha}).$$

[7] I am assuming here only a necessary (but not sufficient) condition for a state holding as a result of a process: if the state *S* holds as a result of the process *P*, then the objective chance of *S* holding after *P* is higher than the objective chance of *S* not holding after *P*.

I am using *Pr* for objective chances set by the laws. Worlds where the laws distinguish between *Alpha* and *Beta* are remote. So, in a world relevant for assessing ordinary necessity, (5) and (6) have the same truth-value:

(6) $$Pr\,(S_{Beta} \mid P_{Beta}) > Pr\,(\neg S_{Beta} \mid P_{Beta}).$$

Now, by the logic of probability, (6) is equivalent to (7):

(7) $$Pr\,(S_{Beta} \,\&\, \neg S_{Gamma} \mid P_{Beta}) + Pr\,(S_{Beta} \,\&\, S_{Gamma} \mid P_{Beta}) > Pr\,(\neg S_{Beta} \,\&\, \neg S_{Gamma} \mid P_{Beta}) + Pr\,(\neg S_{Beta} \,\&\, S_{Gamma} \mid P_{Beta})$$

Since *Beta* and *Gamma* work in tandem they go through the very same process. Accordingly, $Pr\,(S_{Beta} \,\&\, \neg S_{Gamma} \mid P_{Beta})$ is the objective chance the marble ends up on the right and not on the left, given that these machines go through their moves. Similarly, $Pr\,(\neg S_{Beta} \,\&\, S_{Gamma} \mid P_{Beta})$ is the objective chance that the marble ends up on the left, not on the right, given that these machines go through their moves. In worlds governed by normal laws these are the same. So, in those worlds (7) and (8) have the same truth-value:

(8) $$Pr\,(S_{Beta} \,\&\, S_{Gamma} \mid P_{Beta}) > Pr\,(\neg S_{Beta} \,\&\, \neg S_{Gamma} \mid P_{Beta})$$

But (8) is false in worlds governed by normal laws: given that *Beta* and *Gamma* go through their moves, the objective chance that the marble ends up in *both* trays is certainly not higher than the objective chance that it ends up in *neither*. So, in the non-too-distant worlds relevant for assessing ordinary necessity (5) is also false, which means that *Alpha* does not accomplish its task.

4 SUPER-TASKS AND ULTRA-TASKS

The strength of Black's argument has not been widely appreciated. This is, to a large extent, his own fault: he repeatedly misstated its conclusion. He says that the argument shows that it is logically impossible to perform an infinite series of tasks. He even says that "the notion of an infinite series of act is self-contradictory" (Black 1951, 101). The argument, as I presented it, shows no such thing. We are given no reason to think that *Alpha* cannot move *each marble* from left to right—the conclusion is merely that it cannot move *the marbles* from left to right. Suppose *Alpha* moves each of the marbles and after it is done the marbles are all on the right. This could happen. But if it does, the outcome does not come about as a result of the infinite series of tasks *Alpha* performs. Perhaps it happens as a result of some interference, or as a result of some other thing *Alpha* does. Or perhaps it happens not as a result of anything in particular.

The point can be clarified by distinguishing between two notions of an infinitary task. One is that of a *super-task* – a series of tasks of type ω performed in a finite amount of time. The other is that of an *ultra-task* – a single task performed by performing a super-task.[8] Each individual move performed by *Alpha* is a task. If *Alpha* performs each of its moves it performs a super-task. What we earlier called the task of *Alpha* is an ultra-task: moving the marbles from left to right by performing this super-task. What Black's argument shows is that *Alpha* cannot perform this ultra-task.[9] And since there is nothing special about *Alpha*, a reasonable conjecture is that the same holds for ultra-tasks *tout court*: without there being some very odd laws, they cannot be performed at all.

The literature that followed Black's paper tended to focus on the question whether super-tasks are possible. The consensus appears to be that they are—and I share this view. But can one perform some other task by performing a super-task in the sense in which one can cross the street by making a series of steps or one can draw a picture by connecting a series of points? Considerations inspired by Black suggest a negative answer to this latter question. Super-tasks are *possible* but they are *inert*—when you perform a super-task you cannot thereby perform something over and above the individual tasks included in the super-task.

Hercules's second labor was to kill the Lernean Hydra, a nine-headed monster. The Hydra was a tough opponent: whenever Hercules cut off one of its heads, two new heads grew in its place. Killing such a creature by exponentially accelerating decapitation is an ultra-task, and as such, it is impossible to perform.[10] This is not to say that the supertask of cutting off infinitely many heads cannot *in principle* be performed, or that it is impossible for the Hydra to end up dead after this super-task is performed. Remarkable though it is, all this can happen: after the number of the Hydra's heads grows steadily beyond any finite limit, the Hydra suddenly finds itself headless. But this would not be a killing of the Hydra: the beast would not die as a result of what Hercules did. Setting aside the possibility of strange laws, some intervening force or miracle was also needed. Or perhaps its death was not the result of any process whatsoever.

[8] The notion of a *super-task* is due to (Thomson 1954). (Benacerraf 1962) introduced the notion of a *super-duper-task*: a series of tasks of type ω + 1. Peter Clark and Stephen Read (1984) suggested the notion of a *hyper-task*: series of uncountably many tasks. While I am almost out of adjectives, an *ultra-task* is fundamentally different from all of these. It is not a series of tasks, just one task—constituted by infinitely many.

[9] I don't think Black saw clearly the distinction between super-tasks and ultra-tasks: this is the fundamental confusion in his paper.

[10] According to Apollodorus's account, Hercules manages to kill the Hydra with the help of his trusted nephew Iolaus. Every time Hercules cut off one of the heads, Iolaus held a torch to the stump preventing the growth of the new heads. So, Hercules killed the Hydra by performing a finite series of tasks.

Catching up with the tortoise by traversing infinitely many distinct intervals would also be an ultra-task—something that cannot be done. Of course, Achilles can catch up with the tortoise and —contra Black—he can also traverse infinitely many distinct intervals in a finite time. What Achilles cannot do is perform the former task by performing the latter. The "mathematical" resolution of Zeno's paradox is wrong, even though both common sense and mathematics are vindicated. If you want an answer to the question "In virtue of what does Achilles catch up with the tortoise? (and it is by no means clear that you should want an answer to such a question) you need to appeal to something other than a super-task he performs. You can, for example, truthfully say that Achilles caught up with the tortoise by running faster that the animal, or by moving his legs one after the other, or by traversing a distance of some specific length.

5 TELICITY

I have argued that ultra-tasks are impossible (although not logically impossible) to perform. We can gain a better understanding of why this is so by examining what tasks are.

Let's start with some examples. Catching up with a tortoise, moving a marble from one tray to another, killing the Hydra—these are all tasks. One thing they all have in common is that they all happen *in* a time and are not going on *for* a time. For example, we say that Achilles caught up with the tortoise *in* two minutes, but not that he did that *for* two minutes. By contrast, we say that Achilles ran *for* an hour, but not that he ran *in* an hour. This is the classic test for telicity: it shows that 'caught up with a tortoise' a *telic* verb phrase and 'run' an *atelic* one.

Tasks are events described by telic verb phrases. Such events include a result, or *telos*, whose obtaining marks the end of the task. The *telos* of catching up with a tortoise is being lined up with the tortoise, the *telos* of moving a marble from one tray to another is for that marble to be in the latter tray, the *telos* of killing the Hydra is for the Hydra to be dead, and so on. I suggest that tasks are compound events consisting of a process leading up to a result state. Telic verb phrases describing a single task[11] say of an object that it is involved in a certain process that led to its *telos*. To catch up with a tortoise is to be doing something in a way that leads to being lined up with the tortoise, to move a marble from one tray to another is to be moving the marble from one tray in a way that leads to its being in the other, to kill the Hydra is to be killing it in a way that leads to its death, and so on. The phrase "counting the guests" in its most natural reading

[11] A verb phrase can be used to describe a single event or a multitude of events. 'John traveled in four countries' can mean that there was an event of him traveling in four countries or that four countries are such that there was an event of him traveling in them.

describes a single task; its *telos* is having an answer to the question "How many guests are there?" and the counting process leads up to the state of having this answer.

There is a distinction due to (Vendler 1967) within the category of telic verb phrases between *accomplishments* and *achievements*. The former are said to describe events that are extended in time (like stealing a car or building a house), and the latter events that are near-instantaneous (like reaching the peak or finding a key). It is sometimes said that achievement verb phrases don't allow the progressive, but this is not a reliable criterion (e.g. 'Jack was finding his key' is indeed odd, but 'Jill was reaching the peak is just fine.) I think the real difference between them is that in the case of accomplishments, the process that leads up to the *telos* is properly described with the *progressive from of the verb phrase*, while this is not so in the case of achievements. So if Mary crossed the street (accomplishment) then she was crossing it and this particular process led to her being across the street. But if Mary got across the street (achievement), then something was going on—perhaps she was crossing the street, perhaps she was being carried by someone else, perhaps she was being teleported, it does not matter—and this process, whatever it was, led to her being across the street.

If these ideas are on the right path, then it is part of the meaning of telic verb phrases that a process leads to its natural result.[12] Moving infinitely many marbles from the left hand tray to the right hand tray means moving them from the left hand tray in a way that leads to their being in the right hand tray. This is no doubt possible: one could pick up all the marbles at once from the left hand tray and place them in the right hand one. But moving each of the marbles individually would be a super-task and, I argued, such a super-task does not lead to the marbles being in the right hand tray. The marbles may each be moved from the left and they may all end up on the right, but the latter would not happen as a result of the former. It may happen as a result of something else or not as a result of anything at all.

To count the *F*s, I suggested, is to come up with an answer to the question "How many *F*s are there?" *as a result* of counting each. Now we can see that this suggestion is a consequence of a general semantic thesis about telic verb phrases and a specific proposal about the *telos* of counting. Counting the guests at a birthday party is a straightforward task; counting the primes is not. The latter is an ultra-task, and as such, impossible. You could, in principle, count each prime and at the end come to have the answer that there are infinitely many of them. But your counting would not lead to your having that answer. The "countable" multitudes cannot be counted after all.*

[12] For a sketch of a semantic account along these lines see (Szabó 2004) and (Szabó 2008).

* A version of this paper was presented at the *Semantics and Philosophy in Europe* conference in Paris, at the *Archél/CSMN Graduate Conference* in Oslo, at the University of Connecticut

REFERENCES

Benacerraf, P., 1962, Tasks, Super-tasks and the Modern Eleatics. *Journal of Philosophy 59*, 765–84.
Black, M. 1951, Achilles and the Tortoise. *Analysis 11*, 91–101.
Clark, P. and S. Read, 1984, Hypertasks. *Synthese 61*, 387 – 390.
Gendler, T. S. 1998, Galileo and the Indispensability of Scientific Thought Experiments. *The British Journal for the Philosophy of Science 49*, 397 – 424.
Szabó, Z. G. 2004, On the Progressive and the Perfective. *Nous 38*, 29 – 5.
Szabó, Z. G. 2008, Things in Progress. *Philosophical Perspectives 22*, 499 – 525.
Thomson, J. 1954, Tasks and Super-tasks. *Analysis 15*, 1 – 10.
Vendler, Z. 1967, Verbs and Times. In *Linguistics in Philosophy*, Ithaca, Cornell University Press, 97 – 121.

at Storrs, and at the memorial conference honoring Imre Ruzsa in Budapest. I thank the participants at each of these occasions for a lively and constructive discussion. I also thank Cian Dorr, Tamar Szabó Gendler, Daniel Rothschild, Ted Sider, and Brian Weatherson for written comments and criticism.

Gábor Forrai

What Mathematicians Say Means: In Defense of Hermeneutic Fictionalism[1]

Abstract. Hermeneutic fictionalism about mathematics maintains that mathematics is not committed to the existence of abstract objects such as numbers. Mathematical sentences are true, but they should not be construed literally. Numbers are just fictions in terms of which we can conveniently describe things which exist. The paper defends Stephen Yablo's hermeneutic fictionalism against an objection proposed by John Burgess and Gideon Rosen. The objection, directed against all forms of nominalism, goes as follows. Nominalism can take either a hermeneutic form and claim that mathematics, when rightly understood, is not committed to the existence of abstract objects, or a revolutionary form and claim that mathematics is to be understood literally but is false. The hermeneutic version is said to be untenable because there is no philosophically unbiased linguistic argument to show that mathematics should not be understood literally. Against this I argue that it is wrong to demand that hermeneutic fictionalism should be established solely on the basis of linguistic evidence. In addition, there are reasons to think that hermeneutic fictionalism cannot even be defeated by linguistic arguments alone.

Fictionalism is a general term for approaches which analyze a particular discourse or a particular idiom in terms of fictions. Take, for example, the sentence 'The average star has 2.4 planets'. Given the logical form of sentences involving definite descriptions, this sentence seems to assert that there is one and only one object which is the average star. But there is no such object, so the sentence is false. How come, then, that we find it true? The fictionalist says that in using this sentence we engage in a sort of game. We pretend that there is such an object and use this pretense to express a truth, namely, that if divide the number of planets with the number of stars we get 2.4.

[1] The research leading to this paper was supported by OTKA (National Foundation for Scientific Research), grant no. K 76865

Fictionalism can be pursued in a hermeneutic and in a revolutionary spirit.[2] Hermeneutic fictionalism seeks to uncover how the given discourse or idiom is in fact understood, i.e. to bring to the fore the meaning which has been there all along. The example just used is an instance of hermeneutic fictionalism. It does not tell us that we should stop believing in the existence of an average star, for we have never believed that. It tells us that instead of looking for a novel construal of the logical form of the sentence which would make it literally true, we should accept that it has the logical form it seems to have and it is not literally true.[3] Revolutionary fictionalism, in contrast, claims to reveal that what we took to be real is in fact a piece of fiction. It opens our eyes to the fact that we were wrong, and calls on us to change our commitments. Such is Field's attempt to counter Quine's and Putnam's indispensability argument, according to which we cannot but accept that the abstract objects of mathematics exist, because physics cannot do without them.[4] He attempts to show that physics can be pursued without numbers, so we do not have to put up with their existence.[5]

Stephen Yablo advocates hermeneutic fictionalism with respect to mathematics, and his theory has many attractions. It is nominalistic, so it can avoid the epistemological problem raised by Benacerraf. (A note of clarification: by 'nominalism' I mean the rejection of abstract objects and not the rejection of universals; nominalism so conceived is compatible with *in re* realism about universals.) In addition, it promises to explain why mathematics is necessary, how we can know it a priori, why we feel that mathematics is absolute in the sense that there cannot be an alternative arithmetic or set theory, why mathematics can be applied to the physical world, and many other things, including certain features of mathematical language. I will not elaborate on these, I will simply assume that it can deliver what it promises. In this paper I attempt to defend hermeneutic fictionalism against an objection first formulated by John Burgess, which he repeated several times, sometimes together with Gideon Rosen. I will start by a brief sketch of the account, which certainly will not do justice to its full complexity. Then I respond to the objection in two steps. Burgess and Rosen claim that the fate of hermeneutic fictionalism should be decided solely on the basis of empirical linguistic evidence. I argue first that the supportive evidence may come from philosophical considerations as well. Then I suggest, somewhat tentatively, that linguistic evidence alone might not even be sufficient for refutation.

[2] The hermeneutic-revolutionary distinction was introduced in (Burgess 2008a) and is first applied to fictionalism in (Stanley 2001).

[3] For a criticism of the fictionalist analysis of 'the average' example see (Stanley 2001, 54-58). For a response see (Yablo 2001, 93-96).

[4] (Quine 1980a, 1980b, 1981a, 1981b), (Putnam 1979a, 1979b).

[5] (Field 1980).

So let me start with Yablo. Quine has taught us that ontological commitment is marked by quantification. The entities whose existence we are committed to are the ones which we quantify over. Mathematics abounds with theorems which quantify over numbers, e.g. 'Any two numbers have a product'. It seems then that the truth of mathematical theorems implies that numbers exist. Yablo claims that quantifying over numbers incurs no such commitment just as by asserting that 'The average star has 2.4 planets', we do not incur commitment to the existence of the average star. But how can we quantify over numbers and yet abstain from ontological commitment?

Here is how. Number words have a use which is ontologically innocent, namely when they occur as devices of numerical quantification, like in 'There are twelve apostles'. Here the number word can be resolved into the standard devices or first order predicate logic with identity.[6] Starting from this innocent use we can get to quantification over numbers which is just as innocent by adopting a rule, which licenses the expression of the content of sentences involving numerical quantification in terms of quantification over numbers. Stated in a preliminary form, the rule says: if there are n Fs, imagine there is a thing n which is identical with the number of Fs. Using *S* as notation to be read 'imagine/suppose that S', the rule can be written as follows:

($N_{\text{preliminary}}$) if $\exists_n x\ (Fx)$, then *there is a thing n (n = the number of Fs)*[7]

F is a predicate applicable to ordinary objects, and in the antecedent, we have a simple numerical quantification that does not assume the existence of numbers as objects. In the consequent we have quantification over numbers, but the quantification is ontologically innocent, since it occurs in the scope of the 'imagine that' operator. When we merely imagine that something exists, we are not committed to its existence. What the rule says is not that whenever a specifiable real world condition obtains, there exists a given number; it says that whenever a certain real world condition obtains we are allowed to engage in a game of make-belief and pretend that a given number exists.

This rule, however, will not quite do, because it does not allow us to assign numbers to numbers, like when we say 'The number of even primes equals 1'. 'Even' and 'prime' are predicates applicable to numbers, not to ordinary objects, so they cannot occur in the antecedent of the rule. We need to liberalize the rule and allow such predicates in the antecedent. But if we deny that numbers exist, we must also deny that the properties even and prime are instantiated. However, if we may imagine that numbers exist, we may also imagine that these properties are instantiated. This gives us a clue as to how the rule should be amended:

[6] There are n Fs can be defined recursively as follows: $\exists_0 x\ Fx =_{\text{df}} \forall x\ (Fx \supset x \neq x)$, and $\exists_{n+1} x\ Fx =_{\text{df}} \exists y\ (Fy\ \&\ \exists_n x\ (Fx\ \&\ x \neq y))$.

[7] The following account is based primarily on (Yablo 2002).

(N) if *$\exists_n x$ (Fx)*, then *there is a thing n (n = the number of Fs)*

This rule says that if you imagine that there are n Fs, where F may be a property of ordinary objects or numbers, you may also imagine that there is an object which is the number of Fs. This rule includes the preliminary one as a special case: if the antecedent of ($N_{preliminary}$) is satisfied, i.e. if there are indeed a certain number of ordinary objects which are F, you are certainly entitled to imagine that. [8]

But why is it worth pretending that numbers exist? Because of the expressive power the quantificational idiom brings. Without this idiom, it would not be possible, for example, to formulate the laws of physics. Instead of Newton's second law, we could only formulate a huge conjunction with conjuncts of the form 'if a force F is exerted on a body with the mass M, it produces acceleration A'. But we would need an infinite number of conjuncts. Worse, since the magnitudes in question can take real numbers as values, the number of conjuncts should have to be uncountably infinite. If we are allowed to quantify over numbers, we can simply say, 'For all real numbers F, M and A, if F = the force acting on a body with the mass = M, and A = the acceleration produced, then $F = M \times A$'.

It is exactly because of the expressive power of quantification over numbers that Quine believes that mathematical objects are indispensible for physics. Whereas Field accepts that the quantificational idiom yields ontological commitment, and tries to show that we can achieve the same expressive power without quantifying over numbers, Yablo maintains that we may quantify over numbers and yet avoid commitment. We simply pretend that there are mathematical entities. He points out that the use of fictions for purposes of representation is very common. For instance, you may describe a certain bodily feel of nervousness by saying 'There are butterflies in my stomach'. Of course, you do not believe that there are. But if there were, you think that would feel in this way. So you call us to imagine a fictitious state of affairs in order to describe a state of affairs which is real. Indeed, this is the way in which metaphors usually work. Metaphors, read literally, are typically false, but they call us to imagine something. If the call is accepted, the features of what is imagined point us to certain features of reality. One may describe the location of the city of Crotone saying 'It is on the arch of the Italian boot'.[9] Italy is not a boot, but if you are willing to pretend that it is, the sentence tells us where the city is to be found. It is because mathematics shares this feature of figurative speech that Yablo prefers to call his approach 'figuralism'.

[8] Once (N) is in place, we can have infinitely many numbers even if there are only finitely many ordinary objects. 0 is the number of things not identical to themselves, n is the number of numbers smaller than n.

[9] The example from (Walton 1993) 40-41, whose work is a major source of inspiration for fictionalism.

We have seen that real contents of sentences of applied mathematics are states of affairs which include nothing mathematical. But what about pure mathematics? What is, for instance '3 + 5 = 8' about if not about numbers? Yablo shows how sentences of pure mathematics can be recast in the ontologically innocent idiom of numerical quantification. The basic idea is to use rule (N) backwards. What the previous sentence really says is something like this: 'If there are exactly three *F*s and there are exactly five *G*s, and no *F* is a *G*, then there are exactly eight objects which are *F*s or *G*s'. This is a logical truth. Yablo goes on to show how to reconstruct all sentences of arithmetic, including the ones which quantify over numbers, as logical truths, and he does the same for set theory. You can already see how Yablo can explain why mathematics is necessary and how it can be known a priori.

This should suffice to give us a flavor of Yablo's approach. Let us now see why Burgess and Rosen believe that an account along these lines is untenable. The objection is not directed specifically against fictionalism but against nominalism in general. The nominalist denies the existence of abstract objects, so he does not accept that the mathematical sentences apparently asserting the existence of such objects are literally true. At this point, he has two options. To admit that these sentences are true and deny that they are understood literally, or to admit that they are understood literally and deny that they are true. The former is the hermeneutic, the latter is the revolutionary position. Burgess and Rosen argue that both are untenable. The hermeneutic position fails because it is not supported by *scientific* evidence. The revolutionary position fails because there are no sound scientific reasons to challenge the truth of mathematics or to replace current mathematics with a nominalistic alternative such as Field's or Chihara's. I emphasize "scientific", because Burgess and Rosen are of the conviction that purely philosophical considerations can never take precedence over scientific reasoning. For example, epistemological worries about how we can acquire knowledge of the abstract entities of mathematics are not sufficient to discredit mathematicians' claims to knowledge, and *a fortiori*, the truths of mathematics.[10] I grant this.

Nonetheless—and now I am starting with the response—when it comes to arguing against the hermeneutic approach, the point that purely philosophical considerations cannot trump scientific ones is replaced by something stronger, namely that philosophical considerations are simply irrelevant and carry no weight at all. They write "no nominalists favoring such a reconstrual have ever published their suggestions in a linguistics journal with evidence such as a linguist without ulterior ontological motives might accept".[11] At another place Burgess briefly responds to those criticisms which allege that nominalists can have a third alternative in addition to hermeneutics and revolution.

[10] (Burgess and Rosen 2005, 520-523.)
[11] (Burgess and Rosen 2005, 525.)

> [I]t is sometimes said that a nominalist interpretation represents "the best way to make sense of" what mathematicians say. I see in this formulation not a third alternative, but simply an equivocation, between "the empirical hypothesis about what mathematicians mean that best agrees with the evidence" (hermeneutic) and "the construction that can be put on mathematicians' words that would best reconcile them with certain philosophical principles or prejudices" (revolutionary).[12]

What these remarks indicate is that the evidence for a nominalist interpretation of mathematics, such as Yablo's, should be purely empirical and should not rely on philosophical considerations. This is actually how Burgess and Rosen proceed when they take up Yablo's position. [13] They systematically ignore the philosophical benefits Yablo's account may bring, and focus on the evidence from linguistic behavior. E.g. Yablo claims that the ease with which we pass from ontological innocent number talk to the quantificational formula, that we do not demand a proof existence, suggests that the latter idiom does not carry ontological commitment either. Or: if the Oracle mentioned in Burgess' and Rosen's book,[14] who knows exactly what exists, would proclaim that only concrete objects exist, mathematicians would not renounce their existence claims. I do not want to discuss Yablo's linguistic arguments and Burgess' and Rosen's rejoinders. Suffice it to say that I do not find the rejoinders convincing, and I will later argue that a knockdown linguistic counterargument might not be that easy to formulate.

What I contend is that in assessing the case for hermeneutic fictionalism, it is wrong to disregard philosophical considerations.[15] I do not base this on the intrinsic importance of philosophy but on two facts about interpretation. First fact: interpretation—be it the interpretation of a text, of the behavior of a person, of a set social practices—is aimed at making sense, i.e. showing how the various parts hang together, how they cohere. The pursuit of coherence is checked against the empirical facts. Here is an example. Before the elections, a politician promises not to raise taxes, he comes to power, then raises them. There are several ways this may make sense. One: he believed he would not

[12] (Burgess 2008b, 51.)
[13] (Burgess and Rosen 2005, 528-534.)
[14] (Burgess and Rosen 1997, 3.)
[15] If I succeed, I shall have also disposed of Mark Balaguer's objection. In Balaguer's taxonomy there is no room for hermeneutic fictionalism. He defines fictionalism as the view that mathematical sentences should be taken at face value and are false. Yablo believes that mathematical sentences are true, so he is what Balaguer calls a paraphrase nominalist. Paraphrase nominalism is wrong because the empirical evidence suggests that mathematicians understand mathematical sentences literally and not according to the nominalist paraphrase. To me, this sounds like the same complaint as the one raised by Burgess and Rosen. (Balaguer 2008), (Balaguer 2009, 152, 158).

have to raise taxes and later found, to his dismay, that he was mistaken. Two: he knew all too well that he could not avoid raising taxes and calculated that the loss of credibility would be acceptable price for the increase of popularity the false promise would bring. Three: something in between; he was not certain, but he hoped he would not have to and took a calculated risk. Which is right? Empirical evidence decides. We have to find out what information he had about the state of the economy, how well he understood the information he had, what his advisors said, how often he kept his earlier promises, etc. And there are also several ways the story does not make sense (or at least does not make sense without further assumptions). One: he believed he would not have to raise taxes, and indeed he did not have to, still he raised them just for the fun of it. Two: he made a sincere promise and intended to keep it, just did not realize the legislation he passed was about tax raises. So an interpretation can fail in two ways: by conflicting with the empirical evidence and by violating the demand for coherence.

Second fact: judging whether or how much certain patterns are coherent draws heavily on the interpreter's own beliefs. This element of subjectivity is ineliminable, because there is no universal manual for identifying coherent patterns. The closest we have to such a manual is logic, but in matters of interpretation, logic might not have the last word. An interpretation which involves the attribution of inconsistency, might, on the whole, be better than one which involves the attribution a very far-fetched idea which happens to restore consistency. And to tell whether an idea is indeed far-fetched one has to rely on his own beliefs. Let me illustrate the same fact with the earlier example. Suppose you are thinking black and white. Then you will think that our politician either made a sincere promise but was unlucky, or he lied, and there are no other options. If you do think that, then, of course, you are a lousy interpreter. A good understanding of the field, human psychology and politics in this case, is necessary for a good interpretation. So the element of subjectivity does not imply arbitrariness.

How does this all bear on hermeneutic fictionalism? A philosopher, whose purpose is to interpret mathematics as a cognitive enterprise, wants to find out how various things in and around mathematics hang together. In deciding whether certain ideas cohere, he cannot but rely on what he believes. Suppose he believes that knowledge presupposes some kind of causal access. In that case, he would find it difficult to conceive how the Platonist account of mathematics, according to which mathematics provides literally true descriptions of abstract objects, which are not located in space-time and which are causally inert, may rationally cohere with the fact we do have mathematical knowledge. Or he may wonder how mathematics, alleged to describe causally inert objects, can benefit physics, which provides causal explanations.

If this is right, and the philosopher's interpretation of mathematics is a genuine interpretative enterprise, it cannot make do without reliance on the philosopher's

own convictions. So Burgess and Rosen are wrong when they demand that the interpretation of mathematics is to be based purely on empirical evidence, and should be free of philosophical considerations. Interpretation is never based purely on empirical evidence. It is in the business of uncovering coherence, rational connections between parts—and whether the parts are indeed rationally connected, is not something that can be empirically determined. The objection rests on a misunderstanding of what interpretation involves.

I want to emphasize that the above view of interpretation does not mean that philosophers are entitled to read into mathematics whatever philosophical views they happen to have. In order to see that, it is worth taking a look at how the empirical evidence and the interpreter's convictions interact in the course of interpretation. Suppose a historian is writing a book on Kepler. The dates when Kepler's books were published can be determined empirically. Once again, empirical evidence shows that the astronomical theory of *Harmonice Mundi*, which includes what we now call Kepler's laws, is superior to the astronomical theory in his first book, *Mysterium Cosmographicum*. But it is not empirical evidence which says that it was extremely odd of Kepler to republish his first book two years after the publication *Harmonice Mundi*. This judgment draws on the historian's own understanding that science aims primarily at empirically accurate theories. Given this understanding, the publication of an empirically inferior theory just does not make sense. The historian needs to find a coherent pattern which Kepler's actions fit. He may, for example suggest, that Kepler does not share the current view that empirical accuracy has exclusive importance. Kepler was a Platonist and held that that the world should exhibit an impressive mathematical order. Now *Mysterium Cosmographicum* is superior to *Harmonice Mundi* in terms of mathematical order. (Its leading idea is that orbits are circular and their distances are regulated by the five platonic solids: a platonic solid circumscribed around the orbit a planet closer to the Sun is inscribed in the orbit of the planet farther from the Sun.) Now, the historian who proceeds like this does not simply impute his own beliefs to Kepler, since he admits that Kepler's vision of science is different from his own. But he does not put his own beliefs aside either. After all, it is in terms of a belief he shares with Kepler that he makes sense of Kepler's actions, namely that it is right to publish what one believes to be good science. It is in the light of this conviction that Kepler's actions turn out in a way rational. So the way to conceive the role of the interpreter's own convictions is this. The interpreter's convictions provide ways in which what is interpreted can be construed as exhibiting coherence. The role of empirical evidence is to determine which ones of these coherent patterns are, in fact, exhibited.

I have been arguing so far that Burgess and Rosen cannot rule that evidence from philosophical considerations is inadmissible. This may strengthen the case for hermeneutic fictionalism. Now I want to go further and suggest that it is not entirely clear that hermeneutic fictionalism can be refuted at all solely

by non-philosophical considerations. Suppose we consider only arguments from the mathematicians' linguistic behavior and in the interpretation of what mathematicians say and write we consciously abstain from relying on philosophical considerations. I will consider two scenarios in which the result of such non-philosophical arguments is apparently unfavorable to hermeneutic fictionalism and claim that these scenarios do not suffice to refute hermeneutic fictionalism.

The first scenario is that we find that mathematicians do not believe that mathematical objects are fictions because they do not have beliefs about their ontological status. For instance, an empirical survey shows that the overwhelming majority of mathematicians say that they have not thought much about this question, they are not particularly interested in it, or claim to be ignorant about it, or are ready to adopt any position recommended to them; and the minority which displays interest consist of two groups. Members of the one have views which are vague, ambiguous, inconsistent or otherwise unsatisfactory. Members of the other minority group are very sophisticated but cannot agree among themselves. If this were the case, the hermeneutic fictionalist would have to choose carefully the way in which he formulates his position. In particular, he should make it very clear that he is not offering a psychological description of what mathematicians think. He should possibly avoid talking about mathematicians' beliefs, or explain that what he calls beliefs are the views which make best sense of what mathematicians do rather than the dispositional mental states they have. Or he should prefer to talk about mathematics and mathematical practice rather than of mathematicians.

This would not be an ad hoc maneuver. Interpretations often involve elements which are not meant to be psychologically faithful.[16] As a first example, take some current interpretations of Descartes which allege that ideas are to be understood as intentional contents. Viewed as a psychological statement ,this would involve some distortion, because Descartes did not possess the concept of intentional content. Today's concept of intentional content is informed by the tradition of Brentano, Husserl, Frege and Chisholm, which emerged only much later. Instead, advocates of this interpretation should be viewed as claiming that understanding Descartes's concept as intentional content is consistent with what Descartes actually says, and sheds light on how various elements of Descartes's thought hang together. For a more dramatic example, take the interpretation of potlatch as a means of maintaining hierarchical relations between clans or villages. Surely, when the Indians of the Pacific Northwest gather to give away

[16] For the purposes of discussion, I assume two things. First, that facts about beliefs are as "hard" as any physical fact. Second, that people are not mistaken about their beliefs. Giving up either assumptions would give me more room to maneuver but would also invite several objections.

and often destroy vast amount of goods, they do not think of this as a way of reinforcing their social status. It is not just that they do not possess the concepts of social science. Even if they did, the social scientists' explanation, which is thoroughly secular, would not occur to them, because in their eyes, potlatch has a religious character.

It is important to see that interpretations which fail in terms of psychological faithfulness may be fully legitimate as interpretations—they are not abnormal or deviant. Interpretation is in the business of making sense, displaying how things rationally cohere. Now sometimes we cannot capture coherence in terms of the actual beliefs of the people we interpret. Descartes's concept of idea is not sufficiently clear to make the coherence of his thought transparent. The people practicing potlatch explain this custom in terms of following the law. But we believe that laws must serve some purpose, so we need a rationale, and the people do not provide one. If an interpreter finds that coherence cannot be captured by psychologically faithful descriptions, he forgoes psychological faithfulness. Similarly, if hermeneutic fictionalism succeeds in making sense of mathematics and its use in physics, it should not be faulted on grounds that it does not represents mathematicians' beliefs.

Let us move over to the second scenario. Here, the interpretation of the linguistic behavior of mathematicians—which relinquishes philosophical considerations—makes it clear that mathematicians reject fictionalism. Imagine, it turns out, they are all Platonists.

Notice that this would not automatically refute fictionalism. It might be the case that the fictionalists are right, and the mathematicians are wrong. This is Mark Balaguer's favored response to the Burgess-Rosen argument.[17] He claims that revolutionary fictionalism, which accepts that mathematical statements are understood literally and are false is tenable. It would be admissible to overrule mathematicians' judgments concerning the ontological status of mathematical entities, for two reasons. First, such a decision would be of little significance for mathematical practice. Second, mathematicians' professional expertise, which a philosopher cannot question, does not extend to the issues of ontological status.

However, hermeneutic fictionalism holds that mathematical statements are true, but are not understood literally, and it is hermeneutic fictionalism I wish to defend. There are two forms the defense can take. One is to reevaluate the mathematician's alleged commitment to Platonism. Suppose mathematicians explain why they take mathematical sentences literally true in the following way. "Look, we know how to tell metaphors from literal speech. We speak literally when we use the words as we ordinarily do. Now the word 'Sun' normally refers

[17] (Balaguer 2009, 153-157.); he believes though that there might also be a way to reject the hermeneutic-revolutionary distinction, 157-161.

to a hot ball of gas. When Romeo calls Juliet the Sun, he cannot be talking literally, since he cannot possibly believe that Juliet is a hot ball of gas. But as opposed to the word 'Sun', mathematical terms do not have an established use with which our use could be contrasted. So we are talking literally." In response to this, the hermeneutic fictionalist may point out that certain expressions are inherently metaphorical in the sense that they do not have literal uses. Take the word 'Vulcan' introduced in Star Trek. If you call someone who always behaves in a cool, emotionally detached and highly logical fashion a Vulcan, you do not mean that he comes from a humanoid race which evolved on the planet Vulcan, since you know all too well that he does not. Or if you describe someone prone to emotional and illogical behavior as not being a Vulcan, you do not mean to assert that he does not from that race. And even if you call Captain Spock a Vulcan, you do not mean in all seriousness that there is an individual bearing this name who comes from the planet Vulcan. This example is meant to illustrate that when mathematicians confess to Platonism, that may be due to the fact that they misconstrue 'literal' or construe it in a way that differs from the hermeneutic fictionalist's intention.[18]

But suppose no such maneuver is possible. Mathematicians happen to be very sophisticated in matters of linguistics, they do not misconstrue hermeneutic fictionalism, but they reject it in full knowledge of what it involves. That alone would still not be enough to refute hermeneutic fictionalism. When defending revolutionary fictionalism, Balaguer considers the idea that his revolutionism might not concern mathematics at all.[19] He envisages a version of Platonism which runs as follows. Mathematical facts are compounded of two sorts of facts: ontologically neutral facts about the correctness of mathematical sentences construed in fictionalist terms, and platonic facts to the effect that the abstract objects mathematical sentences seem to describe exist, which make it the case that the sentences which are correct in the fictionalist terms are actually true. It is only these platonic facts which on the fictionalist view do not obtain. Balaguer wonders if the platonic facts are mathematical facts at all. If not, the fictionalism he proposes would amount to a revolution in philosophy rather than mathematics. He admits that he does not know how to show that the alleged platonic facts are not mathematical in nature, and neither do I.

I believe, however, that the hermeneutic fictionalist can make a similar move and is in a position to argue for it. Suppose that if we take into account philosophical considerations and no others, fictionalism scores better than other alternatives. This should be granted for the sake of argument, since if fictionalism fails on philosophical grounds, it fails, and there is no point in trying to show that it can be maintained in the in face of its rejection by mathematicians.

[18] For a more inclusive discussion see (Yablo 2000, 221-224.)
[19] (Balaguer 2009, 156.)

Then from the hermeneutic fictionalist's point of view, the situation looks as follows. Certain things mathematicians say, e.g. 'For every prime number there is a larger one', are true, even though not in a literal sense. Other things they say, e.g. 'Numbers are abstract objects and they do exist' are false in the literal sense. (If mathematicians did not intend these sentences in the literal sense, they would not be contradicting the fictionalist.) For sentences in the first group, they have arguments, which are virtually impossible to resist, and these arguments apply a small group of very special methods, such as deduction from axioms. Arguments for the sentences in the second group are not based on these special methods, and they can and should be resisted. Add to these certain behavioral or, if you wish, sociological facts. The professional training mathematicians receive prepares them to deal with the first group. The scholarly journals they publish in are devoted to the first group. One may gain recognition as a great mathematician only by establishing claims in the first group. Those who are exclusively concerned with the second group are typically not regarded as mathematicians, and the list may be continued. All in all, we find that the distinction between the two groups of sentences is not a local phenomenon but is manifested in many ways. Given the significance this distinction seems to have, an interpretation of mathematical practice has to account for it. And the easiest way to account for it is to say that sentences in the first group are the only ones that genuinely belong to mathematics. If this is right, then the mathematicians' uniform commitment to Platonism envisaged in this second scenario does not provide much of an argument against hermeneutic fictionalism, because this commitment falls outside territory of mathematics.

Let me summarize. I argued that Burgess and Rosen are wrong when they demand that hermeneutic fictionalism should be established purely by linguistic considerations. This argument was based on the nature of interpretation. I also raised doubts whether hermeneutic fictionalism can be defeated purely by linguistic considerations. I did that by considering two scenarios which might have seemed to support decisive linguistic objections. This latter argument was not meant to be conclusive. Perhaps one may develop a very well motivated account of fictional talk and use this to show that hermeneutic fictionalism is untenable.

REFERENCES

Balaguer, Mark, 2008, Fictionalism in the philosophy of mathematics. in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <http://plato.stanford.edu/archives/fall2008/entries/fictionalism-mathematics/>.

Balaguer, Mark, 2009, Fictionalism, theft, and the story of mathematics. *Philosophia Mathematica* 17, 131-162.

Burgess, John P., 2008a, Why I am not a nominalist. Reprinted in *Mathematics, Models, and Modality: Selected Philosophical Essays*, Cambridge, Cambridge University Press, 31-45.

Burgess, John P., 2008b, Mathematics and bleak house. Reprinted in *Mathematics, Models, and Modality: Selected Philosophical Essays*, Cambridge: Cambridge University Press, 46-65.

Burgess, John P. and Gideon Rosen, 1997, *A Subject With No Object*, New York, Oxford University Press.

Burgess, John P. and Gideon Rosen, 2005, Nominalism reconsidered. In S. Shapiro (ed.), *The Oxford Handbook of Philosophy and Mathematics and Logic*. Oxford, Oxford University Press, 515- 535.

Chihara, Charles, 2005, Nominalism. In *The Oxford Handbook of Philosophy and Mathematics and Logic*, Oxford, Oxford University Press, 483-514.

Field, Hartry, 1980, *Science Without Numbers*. Princeton (NJ), Princeton University Press.

Putnam, Hilary, 1979a, What is mathematical truth? In *Mathematics, Matter and Method: Philosophical Papers* vol 1, 2nd ed., Cambridge, Cambridge University Press, 60-78.

Putnam, Hilary, 1979b, Philosophy of logic. Reprinted in *Mathematics, Matter and Method: Philosophical Papers*, vol 1, 2nd ed., Cambridge, Cambridge University Press, 323-357.

Quine, Willard Van Orman, 1980a, On what there is. Reprinted in *From a Logical Point of View*, 2nd ed., Cambridge (MA), Harvard University Press, 1-19.

Quine, Willard Van Orman, 1980b, Two dogmas of empiricism. Reprinted in *From a Logical Point of View*, 2nd ed., Cambridge (MA), Harvard University Press, 20-46.

Quine, Willard Van Orman, 1981a Things and their place in theories. In *Theories and Things*. Cambridge (MA), Harvard University Press, pp. 1-23.

Quine, Willard Van Orman, 1981b. Five milestones of empiricism. In *Theories and Things*, Cambridge (MA), Harvard University Press, 67-72.

Stanley, Jason, 2001. Hermeneutic fictionalism. In P A. French and H. K. Wettstein (eds.), *Midwest Studies in Philosophy* 25: *Figurative Language*, 36-71.

Walton Kendall L., 1990, *Mimesis as Make-Believe: On the Foundation of the Representational Arts*. Cambridge (MA) / London, Harvard University Press.

Walton Kendall L., 1993, Metaphor and prop oriented make-believe, *European Journal of Philosophy* 1, 39-57.

Yablo, Stephen, 2000, Apriority & existence. In P. Boghossian and C. Peacocke (eds.) *New Essays on the A Priori*, Oxford: Oxford University Press, 197-228.

Yablo, Stephen, 2001, Go figure: a path through fictionalism. In P. A. French and H. K. Wettstein (eds.), *Midwest Studies in Philosophy* 25: *Figurative Language*, 72-102.

Yablo, Stephen, 2002, Abstract objects: a case study. *Nous* 36, supplementary volume 1, 220-240.

Yablo, Stephen, 2005, The myth of the seven. In M. Kalderon (ed.), *Fictionalism in Metaphysics*, New York, Oxford University Press, pp. 88-115.

Yablo, Stephen and Andre Gallois, 1998, Does ontology rest on a mistake?. *Proceedings of the Aristotelian Society*, supplementary volume 72, 229-262.

FOUNDATIONS OF SCIENCE

ANDRÉKA–MADARÁSZ–NÉMETI–SZÉKELY

On Logical Analysis of Relativity Theories

Abstract. The aim of this paper is to give an introduction to our axiomatic logical analysis of relativity theories.

1 INTRODUCTION

Our general aim is to build up relativity theories as theories in the sense of mathematical logic. So we axiomatize relativity theories within pure first-order logic (FOL) using simple, comprehensible and transparent basic assumptions (axioms). We strive to prove all the surprising predictions of relativity from a minimal number of convincing axioms. We eliminate tacit assumptions from relativity by replacing them with explicit axioms (in the spirit of the foundation of mathematics and Tarski's axiomatization of geometry). We also elaborate logical and conceptual analysis of our theories.

Logical axiomatization of physics, especially that of relativity theory, is not a new idea, among others, it goes back to such leading scientists as Hilbert, Reichenbach, Carnap, Gödel, and Tarski. Relativity theory was intimately connected to logic from the beginning, it was one of the central subjects of logical positivism. For a short survey on the broader literature, see, e.g., (Andréka *&* *al.* 2006). Our aims go beyond these approaches in that along with axiomatizing relativity theories we also analyze in detail their logical and conceptual structure and, in general, investigate them in various ways (using our logical framework as a starting point).

A novelty in our approach is that we try to keep the transition from special relativity to general relativity logically transparent and illuminating. We "derive" the axioms of general relativity from those of special relativity in two natural steps. First we extend our axiom system for special relativity with accelerated observers

(sec.7). Then we eliminate the distinguished status of inertial observers at the level of axioms (sec.8).

Some of the questions we study to clarify the logical structure of relativity theories are:

- What is believed and why?
- Which axioms are responsible for certain predictions?
- What happens if we discard some axioms?
- Can we change the axioms and at what price?

Our aims stated in the first paragraph reflect, partly, the fact that we axiomatize a physical theory. Namely, in physics the role of axioms (the role of statements that we assume without proofs) is more fundamental than in mathematics. Among others, this is why we aim to formulate simple, logically transparent and intuitively convincing axioms.Our goal is that on our approach, surprising or unusual predictions be theorems and not assumed as axioms. For example, the prediction "no faster than light motion ..." is a theorem on our approach and not an axiom, see Thm.5.1.

Getting rid of unnecessary axioms is especially important in a physical theory. When we check the applicability of a physical theory in a situation, we have to check whether the axioms of the theory hold or not. For this we often use empirical facts (outcomes of concrete experiments). However, these correspond to existentially quantified theorems[1] rather than to universally quantified statements—which the axioms usually are. Thus while we can easily disprove the axioms by referring to empirical facts, we can verify these axioms only to a certain degree. Some of the literature uses the term 'empirical fact' for universal generalization of an empirical fact elevated to the level of axioms, see, e.g., (Gömöri–Szabó 2010, §4), (Szabó 2009). We simply call these generalizations (empirical) axioms.

2 WHY RELATIVITY?

For one thing, Einstein's theory of relativity not just had but still has a great impact on many areas of science. It has also greatly affected several areas in the philosophy of science. Relativity theory has an impact even on our every day life, e.g., via GPS technology (which cannot work without relativity theory). Any theory with such an impact is also interesting from the point of view of axiomatic foundations and logical analysis.

Since spacetime is a similar geometrical object as space, axiomatization of relativity theories (or spacetime theories in general) is a natural continuation of

[1]We do not want to assume every experimental fact as an axiom. We only want them to be consequences of our theories.

the works of Euclid, Hilbert, Tarski and many others axiomatizing the geometry of space.

3 WHY AXIOMATIC METHOD?

There are many examples showing the benefits of using axiomatic method. For example, if we decompose relativity theories into little parts (axioms), we can check what happens to our theory if we drop, weaken or replace an axiom or we can take any prediction, such as the twin paradox, and check which axiom is and which is not needed to derive it. This kind of reverse thinking helps to answer the why-type questions. For details on answering why-type questions by the methodology of the present work, see (Andréka *& al.* 2002, 12–13.), (Székely 2010a).

The success story of axiomatic method in the foundations of mathematics also suggests that it is worth applying this method in the foundations of spacetime theories (Friedman 2004a), (Friedman 2004b). Let us note here that Euclid's axiomatic-deductive approach to geometry also made a great impression on the young Einstein, see (Herschbach 2008).

Among others, logical analysis makes relativity theory modular: we can change some axioms, and our logical machinery ensures that we can continue working in the modified theory. This modularity might come handy, e.g., when we want to unify general relativity and quantum theory to a theory of quantum gravity. For further reasons why to apply the axiomatic method to spacetime theories, see, e.g., (Andréka *& al.* 2006), (Andréka *& al.* 2002), (Guts 1982), (Schutz 1973), (Suppes 1968).

4 WHY FIRST-ORDER LOGIC?

We aim to provide a logical foundation for spacetime theories similar to the rather successful foundations of mathematics, which, for good reasons, was performed strictly within FOL. One of these reasons is that FOL helps to avoid tacit assumptions. Another is that FOL has a complete inference system while second-order logic (or higher-order logic) cannot have one.

Still another reason for choosing FOL is that it can be viewed as a fragment of natural language with unambiguous syntax and semantics. Being a *fragment of natural language* is useful in our project because one of our aims is to make relativity theory accessible to a broad audience. *Unambiguous syntax and semantics* are important, because they make it possible for the reader to always know what is stated and what is not stated by the axioms. Therefore they can use the axioms without being familiar with all the tacit assumptions and rules of thumb of physics (which one usually learns via many, many years of practice).

For further reasons why to stay within FOL when dealing with axiomatic foundations, see, e.g., (Andréka *& al.* 2002, §Appendix: Why FOL?), (Ax 1978), (Székely 2009, §11), (Väänänen 2001), (Wolenski 2004).

5 SPECIAL RELATIVITY

Before we present our axiom system let us go back to Einstein's original (logically non-formalized) postulates. Einstein based his special theory of relativity on two postulates, the principle of relativity and the light principle: "The laws by which the states of physical systems undergo change are not affected, whether these changes of state be referred to the one or the other of two systems of coordinates in uniform translatory motion." and "Any ray of light moves in the 'stationary' system of co-ordinates with the determined velocity c, whether the ray be emitted by a stationary or by a moving body.", see (Einstein 1905/1952).

The logical formulation of Einstein's principle of relativity is not an easy task since it is difficult to capture axiomatically what "the laws of nature" are in general. Nevertheless, the principle of relativity can be captured by our FOL approach, see (Andréka *& al.* 2002), (Madarász 2002, §2.8.3).

Instead of formulating the two original principles, we formulate the following consequence of theirs: "the speed of light signals is the same in every direction everywhere according to every inertial observer" (and not just according to the 'stationary' observer). Here we will base our axiomatization on this consequence and call it light axiom. We will soon see that the light axiom can be regarded as the key assumption of special relativity.

Since we want to axiomatize special relativity, we have to fix some formal language in which we will write up our axioms. Let us see the basic concepts (the "vocabulary" of the FOL language) we will use. We would like to speak about motion. So we need a basic concept of things that can move. We will call these object *bodies*.[2] The light axiom requires a distinguished type of bodies called *photons* or *light signals*.[3] We will represent motion as the changing of spatial location in time. Thus we will use reference frames for coordinatizing events (meetings of bodies). Time and space will be marked by *quantities*. The structure of quantities will be an *ordered field* in place of the field of real numbers.[4] For

[2]By bodies we mean anything which can move, e.g., test-particles, reference frames, electromagnetic waves, etc.

[3]Here we use light signals and photons as synonyms because it is not important here whether we think of them as particles or electromagnetic waves. The only thing that matters here is that they are "things that can move." So they are bodies in the sense of our FOL language.

[4]Using ordered fields in place of the field of real numbers increases the flexibility of the theory and reduces the amount of mathematical presuppositions. For further motivation in

simplicity, we will associate special bodies to reference frames. These special bodies will be called "observers." Observations will be formalized/represented by means of the *worldview relation*.

To formalize the ideas above, let us fix a natural number $d \geq 2$ for the dimension of spacetime. To axiomatize theories of the d-dimensional spacetime, we will use the following two-sorted FOL language:

$$\{ B, \mathsf{IOb}, \mathsf{Ph},\ Q, +, \cdot,\ \mathsf{W} \},$$

where B (bodies) and Q (quantities) are the two sorts,[5] IOb (inertial observers) and Ph (light signals or photons) are one-place relation symbols of sort B, $+$ and $\cdot$ are two-place function symbols of sort Q, and W (the worldview relation) is a $2 + d$-place relation symbol the first two arguments of which are of sort B and the rest are of sort Q.

Atomic formulas $\mathsf{IOb}(k)$ and $\mathsf{Ph}(p)$ are translated as "k *is an inertial observer*," and "p *is a photon*," respectively. To speak about coordinatization, we translate $\mathsf{W}(k, b, x_1, \ldots, x_{d-1}, t)$ as "*body* k *coordinatizes body* b *at space-time location* $\langle x_1, \ldots, x_{d-1}, t \rangle$," (i.e., at space location $\langle x, \ldots, x_{d-1} \rangle$ and at instant t). Sometimes we use the more picturesque expressions *sees* or *observes* for *coordinatizes*. However, these cases of "seeing" and "observing" have nothing to do with visual seeing or observing; they only mean associating coordinate points to bodies.

The above, together with statements of the form $x = y$ are the so-called *atomic formulas* of our FOL language, where x and y can be arbitrary variables of the same sort, or terms built up from variables of sort Q by using the two-place operations $\cdot$ and $+$. The *formulas* are built up from these atomic formulas by using the logical connectives *not* ($\neg$), *and* ($\wedge$), *or* ($\vee$), *implies* ($\rightarrow$), *if-and-only-if* ($\leftrightarrow$) and the quantifiers *exists* ($\exists$) and *for all* ($\forall$). For the precise definition of the syntax and semantics of FOL, see, e.g., (Chang–Keisler 1990, §1.3).

To meaningfully formulate the light axiom, we have to provide some algebraic structure for the quantities. Therefore, in our first axiom, we state some usual properties of addition $+$ and multiplication $\cdot$ true for real numbers.

AxFd: The quantity part $\langle Q, +, \cdot \rangle$ is a Euclidean field, i.e.,

- $\langle Q, +, \cdot \rangle$ is a field in the sense of abstract algebra,
- the relation $\leq$ defined by $x \leq y \overset{d}{\iff} \exists z\ x + z^2 = y$ is a linear ordering on Q, and
- Positive elements have square roots: $\forall x\ \exists y\ x = y^2 \vee -x = y^2$.

this direction, see, e.g., (Ax 1978). Similar remarks apply to our other flexibility-oriented decisions, e.g., to treat the dimension of spacetime as a variable.

[5] That our theory is two-sorted means only that there are two types of basic objects (bodies and quantities) as opposed to, e.g., set theory where there is only one type of basic objects (sets).

The field-axioms (see, e.g., (Chang–Keisler 1990, 40–41.)) say that $+$, $\cdot$ are associative and commutative, they have neutral elements 0, 1 and inverses $-$, $/$ respectively, with the exception that 0 does not have an inverse with respect to $\cdot$, as well as $\cdot$ is additive with respect to $+$. We will use 0, 1, $-$, $/$, $\sqrt{\ }$ as derived (i.e., defined) operation symbols.

AxFd is a "mathematical" axiom in spirit. However, it has physical (even empirical) relevance. Its physical relevance is that we can add and multiply the outcomes of our measurements and some basic rules apply to these operations. Physicists usually use all properties of the real numbers tacitly, without stating explicitly which property is assumed and why. The two properties of real numbers which are the most difficult to defend from an empirical point of view are the Archimedean property, see (Rosinger 2008), (Rosinger 2009, §3.1), and the supremum property,[6] see the remark after the introduction of axiom Cont on p.14.

Euclidean fields got their name after their role in Tarski's FOL axiomatization of Euclidean geometry (Tarski 1959). By AxFd we can reason about the Euclidean structure of a coordinate system the usual way, we can introduce Euclidean distance, speak about straight lines, etc. In particular, we will use the following notation for $\bar{x}, \bar{y} \in Q^n$ (i.e., $\bar{x}$ and $\bar{y}$ are n-tuples over Q) if $n \geq 1$:

$$|\bar{x}| \stackrel{d}{=} \sqrt{x_1^2 + \cdots + x_n^2}, \quad \text{and} \quad \bar{x} - \bar{y} \stackrel{d}{=} \langle x_1 - y_1, \ldots, x_n - y_n \rangle.$$

We will also use the following two notations:

$$\bar{x}_s \stackrel{d}{=} \langle x_1, \ldots, x_{d-1} \rangle \quad \text{and} \quad x_t \stackrel{d}{=} x_d$$

for the *space component* and the *time component* of $\bar{x} = \langle x_1, \ldots, x_d \rangle \in Q^d$, respectively.

Now let us see how the light axiom can be formalized in our FOL language.

AxPh: For any inertial observer, the speed of light is the same in every direction everywhere, and it is finite. Furthermore, it is possible to send out a light signal in any direction. Formally:

$$\forall m \ \exists c_m \ \forall \bar{x}\bar{y} \ \mathsf{IOb}(m) \rightarrow \\ \big(\exists p \ \mathsf{Ph}(p) \wedge \mathsf{W}(m,p,\bar{x}) \wedge \mathsf{W}(m,p,\bar{y})\big) \leftrightarrow |\bar{y}_s - \bar{x}_s| = c_m \cdot |y_t - x_t|.$$

Axiom AxPh has an immediate physical meaning. This axiom is not only implied by the two original principles of relativity, but it is well supported by experiments, such as the Michelson-Morley experiment. Moreover, it has been continuously tested ever since then. Nowadays it is tested by GPS technology.

[6]The supremum property (i.e., *every* nonempty and bounded *subset* of the real numbers has a least upper bound) implies the Archimedean property. So if we want to get ourselves free from the Archimedean property, we have to leave this property, too.

Axiom AxPh says that "It is *possible* for a photon to move from $\bar{x}$ to $\bar{y}$ iff ...". So, a notion of possibility plays a role here. In the present paper we work in an extensional framework, as is customary in geometry and in spacetime theory. However, it would be more natural to treat this "possibility phenomenon" in a modal logic framework, and this is more emphatically so for relativistic dynamics (Andréka *& al.* 2008). It would be interesting to explore the use of modal logic in our logical analysis of relativity theory. This investigation would be a nice unification of the works of Imre Ruzsa's school on modal logic and the works of our Tarskian spirited school on axiomatic foundations of relativity theory. Robin Hirsch's work can be considered as a first step along this road (Hirsch 2009).

Let us note that AxPh does not require that the speed of light be the same for every inertial observer or that it be nonzero. It requires only that the speed of light according to a fixed inertial observer be a quantity which does not depend on the direction or the location.

Why do we not require that the speed of light is nonzero? The main reason is that we are building our logical foundation of spacetime theories examining thoroughly each part of each axiom to see where and why we should assume them. Another (more technical) reason is that it will be more natural to include this assumption ($c_m \neq 0$) in our auxiliary axiom AxSm on page 8.

Our next axiom connects the worldviews of different inertial observers by saying that all observers observe the same "external" reality (the same set of events). Intuitively, by the event occurring for m at $\bar{x}$, we mean the set of bodies m observes at $\bar{x}$. Formally:

$$\mathsf{ev}_m(\bar{x}) \stackrel{d}{=} \{b : \mathsf{W}(m, b, \bar{x})\}.$$

AxEv: All inertial observers coordinatize the same set of events:

$$\forall mk \;\; \mathsf{IOb}(m) \wedge \mathsf{IOb}(k) \;\; \rightarrow \;\; \forall \bar{x} \;\; \exists \bar{y} \;\; \forall b \;\; \mathsf{W}(m, b, \bar{x}) \leftrightarrow \mathsf{W}(k, b, \bar{y}).$$

This axiom is very natural and tacitly assumed in the non-axiomatic approaches to special relativity, too.

Basically we are done. We have formalized the light axiom AxPh. We have introduced two supporting axioms (AxFd and AxEv) for the light axiom which are simple and natural; however, we cannot simply omit them without loosing some of the meaning of AxPh. The field axiom enables us to speak about distances, time differences, speeds, etc. The event axiom ensures that different inertial observers see the same events.

In principle, we do not need more axioms for analyzing/axiomatizing special relativity, but let us introduce two more simplifying ones. We could leave them out without loosing the essence of our theory, it is just that the formalizations of the theorems would become more complicated.

AxSf: Any inertial observer sees himself on the time axis:

$$\forall m \ \mathsf{IOb}(m) \rightarrow \ \big(\forall \bar{x} \ \mathsf{W}(m,m,\bar{x}) \leftrightarrow x_1 = 0 \wedge x_2 = 0 \wedge x_3 = 0\big).$$

The role of AxSf is nothing more than making it easier to speak about the motion of reference frames via the motion of their time axes. Identifying the motion of reference frames with the motion of their time axes is a standard simplification in the literature. AxSf is a way to formally capture this simplifying identification.

Our last axiom is a symmetry axiom saying that all inertial observers use the same units of measurements.

AxSm: Any two inertial observers agree about the spatial distance between two events if these two events are simultaneous for both of them; furthermore, the speed of light is 1:

$$\forall mk \ \mathsf{IOb}(m) \wedge \mathsf{IOb}(k) \rightarrow \forall \bar{x}\bar{y}\bar{x}'\bar{y}' \ \ x_t = y_t \wedge x'_t = y'_t \wedge$$
$$\mathsf{ev}_m(\bar{x}) = \mathsf{ev}_k(\bar{x}') \wedge \mathsf{ev}_m(\bar{y}) = \mathsf{ev}_k(\bar{y}') \rightarrow |\bar{x}_s - \bar{y}_s| = |\bar{x}'_s - \bar{y}'_s|, \text{ and}$$

$$\forall m \ \mathsf{IOb}(m) \ \ \rightarrow \ \ \exists p \ \mathsf{Ph}(p) \ \wedge \ \mathsf{W}(m,p,0,0,0,0) \ \wedge \ \mathsf{W}(m,p,1,0,0,1).$$

Let us see how AxSm states that "all inertial observers use the same units of measurements." That "the speed of light is 1" (besides that the speed of light is nonzero) means only that observers are using units measuring time distances compatible with the units measuring spatial distances, such as light years or light seconds. The first part of AxSm means that different observers use the same unit measuring spatial distances. This is so because if two events are simultaneous for both observers, they can measure their spatial distance and the outcome of their measurements are the same iff the two observers are using the same units to measure spatial distances.

Our axiom system for special relativity contains these 5 axioms only:

$$\mathsf{SpecRel} \stackrel{d}{=} \{\mathsf{AxFd}, \mathsf{AxPh}, \mathsf{AxEv}, \mathsf{AxSf}, \mathsf{AxSm}\}.$$

In an axiom system, the axioms are the "price" we pay, and the theorems are the "goods" we get for them. Therefore, we strive for putting only simple, transparent, easy-to-believe statements in our axiom systems. We want to get all the hard-to-believe predictions as theorems. For example, we prove from SpecRel that it is impossible for inertial observers to move faster than light relative to each other ("No FTL travel" for science fiction fans). In the following, $\vdash$ means logical derivability.

Theorem 5.1. (no faster than light inertial observers)

$$\mathsf{SpecRel} \vdash \forall mk\bar{x}\bar{y} \quad \mathsf{IOb}(m) \wedge \mathsf{IOb}(k) \wedge \mathsf{W}(m,k,\bar{x}) \wedge \mathsf{W}(m,k,\bar{y}) \wedge \bar{x} \neq \bar{y} \ \rightarrow \ |\bar{y}_s - \bar{x}_s| < |y_t - x_t|.$$

For a geometrical proof of Thm.5.1, see (Andréka *& al.* 2010).

In relativity theory we are often interested in comparing the worldviews of different observers. So we introduce the worldview transformation between observers m and k as the following binary relation:

$$\mathsf{w}_{mk}(\bar{x},\bar{y}) \ \stackrel{d}{\Longleftrightarrow} \ \mathsf{ev}_m(\bar{x}) = \mathsf{ev}_k(\bar{y}).$$

By Thm.5.2, the worldview transformations between inertial observers in the models of SpecRel are Poincaré transformations, i.e., transformations which preserve the so-called Minkowski-distance $(y_t - x_t)^2 - |\bar{y}_s - \bar{x}_s|^2$ of d-tuples $\bar{y}, \bar{x}$. For the definition, we refer to (d'Inverno 1992, 110.) or (Misner *& al.* 1973, 66–69.).

Theorem 5.2.

$$\mathsf{SpecRel} \vdash \forall m,k \ \ \mathsf{IOb}(m) \wedge \mathsf{IOb}(k) \ \rightarrow \ \mathsf{w}_{mk} \text{ is a Poincaré transformation.}$$

For the proof of Thm.5.2, see (Andreka *& al.* 2007, Thm.11.10, 640.) or (Székely 2009, Thm.3.2.2, 22.). By Thm.5.2, all predictions of special relativity, such as "moving clocks slow down," are provable from SpecRel. For details, see, e.g., (Andréka *& al.* 2006, §1), (Andreka *& al.* 2007, §2), (Andréka *& al.* 2002, §2.5).

6 LOGICAL ANALYSIS

Let us illustrate here by a simple example what we mean by logical analysis of a theory. In AxEv we have assumed that all observers see the same (possibly infinite) meetings of bodies. Let us try to weaken AxEv to an axiom assuming something similar but only for finite meetings of bodies. A natural candidate is one of the following finite approximations of AxEv:

$\underline{\mathsf{AxMeet_n}}$**:** All inertial observers see the same n-meetings of bodies:

$$\forall mkb_1 \ldots b_n\bar{x} \ \ \mathsf{IOb}(m) \wedge \mathsf{IOb}(k) \wedge \mathsf{W}(m,b_1,\bar{x}) \wedge \ldots \wedge \mathsf{W}(m,b_n,\bar{x}) \rightarrow \exists \bar{y} \ \ \mathsf{W}(k,b_1,\bar{y}) \wedge \ldots \wedge \mathsf{W}(k,b_n,\bar{y}).$$

For example, $\mathsf{AxMeet_1}$ means only that inertial observers see the same bodies. Let us also introduce axiom scheme Meet_ω as the collection of all the axioms $\mathsf{AxMeet_n}$. By Prop.6.1, $\mathsf{AxMeet_n}$ is strictly weaker assumption than $\mathsf{AxMeet_{n+1}}$ and AxEv is strictly stronger than all the axioms of Meet_ω together.

Proposition 6.1.

$$\mathsf{AxEv} \vdash \mathsf{AxMeet}_{n+1} \vdash \mathsf{AxMeet}_n \quad (1)$$

$$\mathsf{AxMeet}_n \nvdash \mathsf{AxMeet}_{n+1} \quad (2)$$

$$\mathsf{Meet}_\omega \nvdash \mathsf{AxEv} \quad (3)$$

Proof. Item (1) follows easily by the formulations of the axioms.

To prove Item (2), we are going to construct a model of AxMeet_n in which AxMeet_{n+1} is not valid. Let $Q = \{0, 1, \ldots, n\}$, $B = \{b_i : i \leq n\}$. Let all the bodies be inertial observers. Let b_0 see all the bodies in $\langle 0, \ldots, 0\rangle$ and none of them in any other coordinate points, i.e., let $\mathsf{W}(b_0, b_i, \bar{x})$ hold iff $\bar{x} = \langle 0, \ldots, 0\rangle$; and for all $k \neq 0$ let b_k see all the bodies but b_i at coordinate points $\langle i, \ldots, i\rangle$ for all $i \leq n$, i.e., let $\mathsf{W}(b_k, b_i, \bar{x})$ hold iff $\bar{x} = \langle j, \ldots, j\rangle$ and $i \neq j$. In this model, all inertial observers see all the possible n-meetings. So AxMeet_n is valid in this model. However, the only inertial observer who sees the $n+1$-meeting $\{b_0, \ldots, b_n\}$ is b_0. So AxMeet_{n+1} is not valid in this model.

We are going to prove Item (3) by a similar model construction. The only difference is that now Q will be infinite. For simplicity, let Q be the set of natural numbers. Let all the other parts of the model be defined in the same way. Now all the inertial observers see all the possible n-meetings of the bodies for all natural numbers n. So AxMeet_n is valid in this model for all natural number n. Hence Meet_ω is valid in this model. However, only b_0 sees the event $\{b_1, b_2, \ldots\}$. So AxEv is not valid in this model. □

Now we will use that there are no stationary (i.e., motionless) light signals. So let us formalize this statement.

Ax(c ≠ 0): Inertial observers do not see stationary light signals.

$$\forall mp\bar{x}\bar{y} \quad \mathsf{IOb}(m) \wedge \mathsf{Ph}(p) \wedge \mathsf{W}(m, p, \bar{x}) \wedge \mathsf{W}(m, p, \bar{y}) \wedge x_t \neq y_t \rightarrow \bar{x}_s \neq \bar{y}_s.$$

Proposition 6.2.

$$\mathsf{AxMeet}_3, \mathsf{AxFd}, \mathsf{AxPh}, \mathsf{Ax}(\mathsf{c} \neq 0) \vdash \mathsf{AxEv} \quad (4)$$

$$\mathsf{AxMeet}_2, \mathsf{AxFd}, \mathsf{AxPh}, \mathsf{Ax}(\mathsf{c} \neq 0) \nvdash \mathsf{AxEv} \quad (5)$$

$$\mathsf{Meet}_\omega, \mathsf{AxFd}, \mathsf{AxPh} \nvdash \mathsf{AxEv} \quad (6)$$

Proof. First let us make some general observations. By AxFd, there is no nondegenerate triangle in Q^d whose sides are of slope c. This is clear if $c = 0$; and in the case $c \neq 0$, this can be shown by contradiction using the fact that the vertical projection of a triangle of this kind is a triangle whose one side is the sum of the other two sides. Therefore, AxFd and AxPh together imply that any

inertial observer m sees the events in which a particular photon participates on a line of slope c_m.

By AxFd, AxPh and Ax($\mathsf{c} \neq 0$), every inertial observer m sees different meetings of photons at different coordinate points. This is so since (by AxFd) for every pair of points there is a line of slope $c_m \neq 0$ containing only one of the points. Hence, by AxPh, there is a photon seen by m only at one of the two coordinate points.

Let us now prove Item (4). Let m and k be inertial observers and let $\bar{x}$ be a coordinate point. To prove AxEv, we have to find a coordinate point $\bar{x}'$ such that $\mathsf{ev}_m(\bar{x}) = \mathsf{ev}_k(\bar{x}')$. To find this $\bar{x}'$, let $\bar{y} = \langle x_1 + c_m, x_2, \ldots, x_{d-1}, x_t + 1\rangle$, $\bar{z} = \langle x_1 - c_m, x_2, \ldots, x_{d-1}, x_t + 1\rangle$ and $\bar{w} = \langle x_1, \ldots, x_{d-1}, x_t + 2\rangle$, see Fig.1.

By AxPh, there are photons p_1, p_2 and p_3 such that $p_1, p_2 \in \mathsf{ev}_m(\bar{x})$, $p_2, p_3 \in \mathsf{ev}_m(\bar{y})$, $p_1 \in \mathsf{ev}_m(\bar{z})$ and $p_3 \in \mathsf{ev}_m(\bar{w})$. Since m sees every photon on a line of slope c_m, he sees the meeting of p_1 and p_2 only at $\bar{x}$ and does not see the meeting of p_1 and p_3.

Since AxMeet_3 implies AxMeet_2, k sees the same meetings of pairs of photons. So there is a $\bar{x}'$ where k sees p_1 and p_2 meet. $\bar{x}'$ is the only point where k sees both p_1 and p_2. This is so because k sees different meetings of photons at different points but sees the same 3-meetings as m. So if there were another point, say $\bar{x}''$, where k sees p_1 and p_2, there were photons $p' \in ev_k(\bar{x}')$ and $p'' \in \mathsf{ev}_k(\bar{x}'')$ such that $p' \notin ev_k(\bar{x}'')$, $p'' \notin \mathsf{ev}_k(\bar{x}')$ and k does not see the meeting of p' and p''. By axiom AxMeet_3 m has to see the meetings $\{p_1, p_2, p'\}$ and $\{p_1, p_2, p''\}$. The only point where m can see these meetings is $\bar{x}$ since $\bar{x}$ the only point where m sees p_1 and p_2 meet. Therefore m sees the meeting of p' and p'' at $\bar{x}$. Thus, by AxMeet_3, k also has to see the meeting of p' and p'', but k does not see it. Hence $\bar{x}'$ is the only point where k sees both p_1 and p_2.

Let b be a body such that $\mathsf{W}(m, b, \bar{x})$. By AxMeet_3, k has to see the meeting of p_1, p_2 and b. This point has to be $\bar{x}'$ since the only point where p_1 and p_2 meet is $\bar{x}'$. Since b was an arbitrary body, we have $\mathsf{ev}_m(\bar{x}) \subseteq \mathsf{ev}_k(\bar{x}')$. The same argument shows that $\mathsf{ev}_k(\bar{x}') \subseteq \mathsf{ev}_m(\bar{x})$. So $\mathsf{ev}_m(\bar{x}) = \mathsf{ev}_k(\bar{x}')$ as desired.

We are going to prove Item (5), by constructing a model. Let $\langle Q, +, \cdot\rangle$ be the field of real numbers. Let us denote the set of natural numbers by ω. Let $B = \{m, k\} \cup \{b_i : i \in \omega\} \cup \{p : p \text{ is a line of slope } 1\}$. Let m and k be all the inertial observers and let the lines of slope 1 be all the photons. Let m and k see the photon p at coordinate point $\bar{x}$ iff $\bar{x} \in p$. Let m see all the bodies b_i at $\bar{x}$ iff $x_t = 0$. Let k see all the bodies $b_0, \ldots, b_n, \ldots$ but b_i at $\bar{x}$ iff $x_t = i$ (i.e., iff $\bar{x}$ is in the horizontal hyperplane $\{\bar{y} \in Q^d : y_t = i\}$).[7] It is straightforward from this construction that axioms AxFd, AxPh and Ax($\mathsf{c} \neq 0$) are valid in this model. Since every line of slope 1 intersects every horizontal hyperplane, m and

[7]If $d = 2$, vertical lines can be used instead of horizontal hyperplanes, which gives a counterexample with bodies having more natural properties.

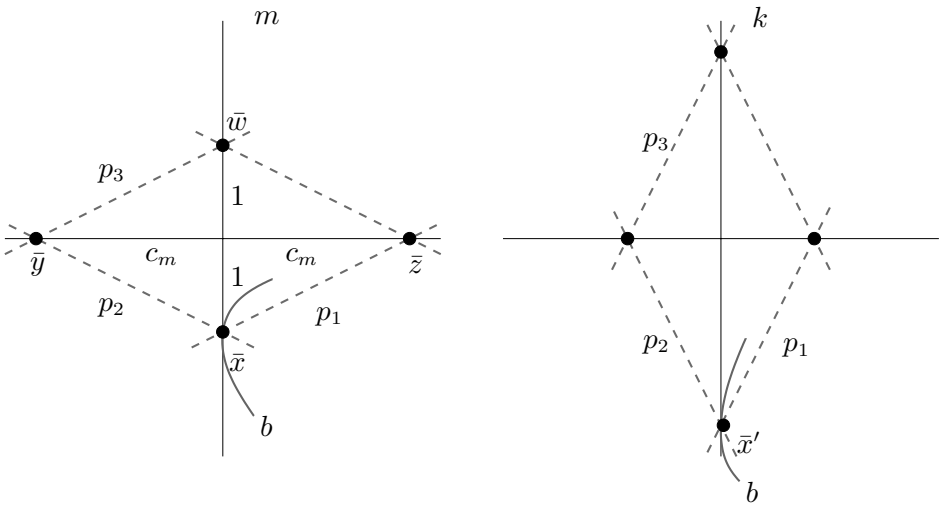

Figure 1.

k see the same 2-meetings of bodies. Hence AxMeet_2 is also valid in this model. However, the only inertial observer who sees the meeting $\{b_i : i \in \omega\}$ is m. So AxEv is not valid in this model.

We prove Item (6) by a similar construction. The only difference is that now the set of bodies is $B = \{m, k\} \cup \{b_i : i \in \omega\} \cup \{p : p \text{ is a vertical line}\}$; and the photons are the vertical lines. It is straightforward from the construction that axioms AxFd, AxPh are valid in this model ($c = 0$). Since every vertical line intersects every horizontal hyperplane, m and k see the same n-meetings of bodies. Hence Meet_ω is also valid in this model. However, only m sees the meeting $\{b_i : i \in \omega\}$. So AxEv is not valid in this model. □

Prop.6.2 shows that a price to weaken axiom AxEv to AxMeet_3 is to assume that there are no stationary light signals. Since AxSm contains this assumption, we can simply replace AxEv with AxMeet_3 in SpecRel. A natural continuation of this investigation can be a search for assumptions that allow us to weaken AxMeet_3 to AxMeet_2. A possible candidate is that bodies move along straight lines and the dimension d is at least 3. The proof of Item (5) shows that assuming only that bodies move along straight lines is not enough, if $d = 2$.

We have several similar investigations on the logical connections of axioms and predictions, see, e.g., (Andréka *& al.* 2008), (Székely 2009, §5) on dynamics, (Madarász *& al.* 2006), (Székely 2009, §4,§7), (Székely 2010b) on twin paradox, (Andréka *& al.* 2002) on kinematics, time-dilation and length-contraction, twin paradox, etc.

7 ACCELERATED OBSERVERS

In SpecRel we restricted our attention to inertial observers. It is a natural idea to generalize the theory by including accelerated observers as well. It is explained in the classic textbook (Misner *& al.* 1973, 163–165.) that the study of accelerated observers is a natural first step (from special relativity) towards general relativity.

We have not introduced the concept of observers as a basic one because it can be defined as follows: an *observer* is nothing other than a body who "observes" (coordinatizes) some other bodies somewhere, this property can be captured by the following formula of our language:

$$\mathsf{Ob}(m) \iff^{d} \exists b\bar{x}\ \mathsf{W}(m,b,\bar{x}).$$

Our key axiom about accelerated observers is the following:

AxCmv: At each moment of his life, every accelerated observer sees (coordinatizes) the nearby world for a short while in the same way as an inertial observer does.

For formulation of AxCmv in our FOL language, see (Madarász *& al.* 2006), (Székely 2009) or (Andréka *& al.* 2010).

Axiom AxCmv ties the behavior of accelerated observers to those of inertial ones. Justification of this axiom is given by experiments. We call two observers *co-moving* at an event if they "see the nearby world for a short while in the same way" at the event. By this notion AxCmv says that at each event of an observer's life, he has a co-moving inertial observer. We can think of a dropped spacepod as a co-moving inertial observer of an accelerated spaceship (at the event of dropping). Or, if a spaceship switches off its engines, it will move on as a co-moving inertial spaceship would.

Our next two axioms ensure that the worldviews of accelerated observers are big enough. They are generalized versions of the corresponding axioms for inertial observers, but now postulated for all observers.

AxEv$^-$: If m sees k in an event, then k cannot deny it:

$$\forall m,k \in \mathsf{Ob}\ \mathsf{W}(m,k,\bar{x}) \to \exists \bar{y}\ \mathsf{ev}_m(\bar{x}) = \mathsf{ev}_k(\bar{y}).$$

AxSf$^-$: Any observer sees himself in an interval of the time axis:

$$\forall m \in \mathsf{Ob}\ \forall \bar{x}\ \mathsf{W}(m,m,\bar{x}) \ \to\ x_1 = x_2 = x_3 = 0 \quad \text{and}$$
$$\forall \bar{x}\bar{y}\quad \mathsf{W}(m,m,\bar{y}) \wedge \mathsf{W}(m,m,\bar{x}) \ \to\ \forall t\ \ x_t < t < y_t \to \mathsf{W}(m,m,0,0,0,t).$$

Our last two axioms will ensure that the worldlines of accelerated observers are "tame" enough, e.g., they have velocities at each moment. In SpecRel,

the worldview transformations between inertial observers are affine maps, the next axiom will state that the worldview transformations between accelerated observers are approximately affine, wherever they are defined.

AxDf: The worldview transformations have linear approximations at each point of their domain (i.e., they are differentiable).

For a precise formalization of AxDf, see, e.g., (Andréka *& al.* 2010).

We note that AxDf implies that the worldview transformations are functions with open domains. However, if the numberline has gaps, still there can be crazy motions. Our last assumption is an axiom scheme supplementing AxDf by excluding these gaps.

Cont: Every definable, bounded and nonempty subset of Q has a supremum (i.e., least upper bound).

In Cont "definable" means "definable in the language of AccRel, parametrically." For a precise formulation of Cont, see (Madarász *& al.* 2006, 692.) or (Székely 2009, §10.1). Cont is a "mathematical axiom" in spirit. It is Tarski's FOL version of Hilbert's continuity axiom in his axiomatization of geometry, see (Goldblatt 2004, 61–162.), fitted to the language of AccRel. When Q is the field of real numbers, Cont is automatically true.

That Cont requires the existence of supremum only for sets definable in the language of AccRel instead of every set, is important not only because by this trick we can keep our theory within FOL (which is crucial in a foundational work), but also because it makes this postulate closer to the the physical/empirical level. The latter is true because Cont does not speak about "any fancy subset" of the quantities, just those "physically meaningful" sets which can be defined in the language of our (physical) theory.

Adding this 5 axioms to SpecRel, we get an axiom system for accelerated observers:

$$\mathsf{AccRel} \stackrel{d}{=} \mathsf{SpecRel} \cup \{\mathsf{AxCmv}, \mathsf{AxEv}^-, \mathsf{AxSf}^-, \mathsf{AxDf}\} \cup \mathsf{Cont}.$$

As an example we show that the so-called *twin paradox* can be naturally formulated and analyzed logically in AccRel. Our axiomatic approach also makes it possible to analyze the details of the twin paradox (e.g., who sees what, when) with the clarity of logic, see (Andréka *& al.* 2002, 139–150.) for part of such an analysis.

According to the twin paradox, if a twin makes a journey into space (accelerates), he will return to find that he has aged less than his twin brother who stayed at home (did not accelerate). We formulate the twin paradox in our FOL language as follows.

TwP: Every inertial observer m measures at least as much time as any other observer k between any two events e_1 and e_2 in which they meet;

and they measure the same time iff they have encountered the very same events between e_1 and e_2:

$$\forall m \in \mathsf{IOb}\ \forall k \in \mathsf{Ob}\ \forall \bar{x}\bar{x}'\bar{y}\bar{y}'\ \ x_t < y_t \wedge x'_t < y'_t \wedge$$
$$m, k \in \mathsf{ev}_m(\bar{x}) = \mathsf{ev}_k(\bar{x}') \wedge m, k \in \mathsf{ev}_m(\bar{y}) = \mathsf{ev}_k(\bar{y}')\ \rightarrow\ y'_t - x'_t \leq y_t - x_t$$
$$\wedge \big(y'_t - x'_t = y_t - x_t \leftrightarrow enc_m(\bar{x}, \bar{y}) = enc_k(\bar{y}', \bar{y}')\big),$$

where $enc_m(\bar{x}, \bar{y}) = \{\mathsf{ev}_m(\bar{z}) : \mathsf{W}(m, m, \bar{z}) \wedge x_t \leq z_t \leq y_t\}$.

Theorem 7.1.

$$\mathsf{AccRel} \vdash \mathsf{TwP} \tag{7}$$
$$\mathsf{AccRel} - \mathsf{AxDf} \vdash \mathsf{TwP} \tag{8}$$
$$\mathsf{AccRel} - \mathsf{Cont} \nvdash \mathsf{TwP} \tag{9}$$
$$\mathsf{Th}(\mathbb{R}) \cup \mathsf{AccRel} - \mathsf{Cont} \nvdash \mathsf{TwP} \tag{10}$$

For the proof of Thm.7.1, see (Madarász *& al.* 2006) or (Székely 2009, §7).

Item (10) of Thm.7.1 states that Cont cannot be replaced with the whole FOL theory of real numbers in AccRel if we do not want to loose TwP from its consequences.

Our theory AccRel is also strong enough to predict the gravitational time-dilation effect of general relativity via Einstein's equivalence principle, see (Madarász *& al.* 2007), (Székely 2009).

8 GENERAL RELATIVITY

Our theory of accelerated observers AccRel speaks about two kinds of observers, inertial and accelerated ones. Some axioms are postulated for inertial observers only, some apply to all observers. We get an axiom system GenRel for general relativity by stating the axioms of AccRel in a generalized form in which they are postulated for all observers, inertial and accelerated ones equally. In other words, we will change all axioms of AccRel in the same spirit as AxSf^- and AxEv^- were obtained from AxSf and AxEv, respectively. This kind of change $\mathsf{AccRel} \mapsto \mathsf{GenRel}$ can be regarded as a "democratic revolution" with the slogan "all observers should be equivalent, the same laws should apply to all of them." Here "law" translates as "axiom." This idea originates with Einstein (see his book (Einstein 1921/2006, Part II, ch.18)).

For simplicity, we will use an equivalent version of the symmetry axiom AxSm (see (Andréka *& al.* 2002, Thm.2.8.17(ii), 138.) or (Székely 2009, Thm.3.1.4, 21.)), and we will require the speed of photons to be 1 in AxPh^- (as opposed to requiring it in AxSm^-).

AxPh$^-$: The velocity of photons an observer "meets" is 1 when they meet, and it is possible to send out a photon in each direction where the observer stands.

AxSm$^-$: Meeting observers see each other's clocks slow down with the same rate.

For a precise formulation of these axioms, see (Andréka *& al.* 2010), (Székely 2009).

We introduce an axiom system for general relativity as the collection of the following axioms:

$$\mathsf{GenRel} \stackrel{d}{=} \{\mathsf{AxFd}, \mathsf{AxPh}^-, \mathsf{AxEv}^-, \mathsf{AxSf}^-, \mathsf{AxSm}^-, \mathsf{AxDf}\} \cup \mathsf{Cont}.$$

Axiom system GenRel contains basically the same axioms as SpecRel, the difference is that they are assumed only locally but for all the observers.

Thm.8.1 below states that the models of GenRel are exactly the spacetimes of usual general relativity. For the notion of a Lorentzian manifold we refer to (d'Inverno 1992, 55.), (Misner *& al.* 1973, 241.) and (Andreka *& al.* 2007, sec.3.2).

Theorem 8.1 (Completeness theorem)**.** GenRel *is complete with respect to its standard models, i.e., with respect to Lorentzian Manifolds over real closed fields.*

This theorem can be regarded as a completeness theorem in the following sense. Let us consider Lorentzian manifolds as intended models of GenRel. How can we do that? We give a method for constructing a model of GenRel from each Lorentzian manifold; and conversely, we show that each model of GenRel is obtained this way from a Lorentzian manifold. After this is elaborated, we have defined what we mean by a formula φ in the language of GenRel being valid in a Lorentzian manifold. Then completeness means that for any formula φ in the language of GenRel, we have $\mathsf{GenRel} \vdash \varphi$ iff φ is valid in all Lorentzian manifolds over real closed fields. This is completely analogous to the way in which Minkowskian spacetimes were regarded as intended models of SpecRel in the completeness theorem of SpecRel, see (Andreka *& al.* 2007, Thm.11.28, 681.) and (Madarász 2002, §4).

We call the worldline of an observer *timelike geodesic*, if each of its points has a neighborhood within which this observer "maximizes measured time (wristwatch time)" between any two encountered events. For formalization of this concept in our FOL language, see, e.g., (Andréka *& al.* 2010).

According to the definition above, if there are only a few observers, then it is not a big deal that a worldline is a time-like geodesic (it is easy to be maximal if there are only a few to be compared to). To generate a real competition for the rank of having a timelike geodesic worldline, we postulate the existence of many observers by the following axiom scheme of comprehension.

Compr: For any parametrically definable timelike curve in any observers worldview, there is another observer whose worldline is the range of this curve.

A precise formulation of Compr can be obtained from that of its variant in (Andreka *& al.* 2007, 679.).

An axiom schema Compr guarantees that our definition of a geodesic coincides with that in the literature on Lorentzian manifolds. Therefore we also introduce the following theory:

$$\mathsf{GenRel}^{+} \stackrel{d}{=} \mathsf{GenRel} \cup \mathsf{Compr}.$$

So in our theory GenRel^{+}, our concept of timelike geodesic coincides with the standard concept in the literature on general relativity. All the other key concepts of general relativity, such as curvature or Riemannian tensor field, are definable from timelike geodesics. Therefore we can treat all these concepts (including the concept of metric tensor field) in our theory GenRel^{+} in a natural way.

In general relativity, Einstein's field equations (EFE) provide the connection between the geometry of spacetime and the energy-matter distribution (given by the energy-momentum tensor field). Since in GenRel^{+} all the geometric concepts of spacetime are definable, we can use Einstein's equation as a definition of the energy-momentum tensor, see, e.g., (Benda 2008) or (d'Inverno 1992, §13.1, 169.), or we can extend the language of GenRel^{+} with the concept of energy-momentum tensor and assume Einstein's equations as axioms. As long as we do not assume anything more of the energy-momentum tensor than its connection to the geometry described by Einstein's equations, there is no real difference in these two approaches. In both approaches, we can add extra conditions about the energy-momentum tensor to our theory, e.g., the dominant energy condition or, e.g., that the spacetimes are vacuum solutions.

9 CAN PHYSICS GIVE FEEDBACK TO LOGIC?

There is observational evidence suggesting that in our physical universe there exist regions supporting potential non-Turing computations. Namely, it is possible to design a physical device in relativistic spacetime which can compute a non-Turing computable task, e.g., which can decide whether ZF set theory is consistent. This empirical evidence is making the theory of hypercomputation more interesting and gives new challenges to the physical Church Thesis, see, e.g., (Andréka *& al.* 2009).

These new challenges do more than simply providing a further connection between logic and spacetime theories; they also motivate the need for logical understanding of spacetime theories.

10 CONCLUDING REMARKS

We have axiomatized both special and general relativity in FOL. Moreover, via our theory AccRel, we have axiomatized general relativity so that each of its axioms can be traced back to its roots in the axioms of special relativity. Axiomatization is not our final goal. It is merely an important first step toward logical and conceptual analysis. We are only at the beginning of our ambitious project.*

REFERENCES

Andréka, H., J. X. Madarász, and I. Németi, with contributions from A. Andai, G. Sági, I. Sain and Cs. Tőke, 2002, *On the Logical Structure of Relativity Theories*. Research report. Budapest, Alfréd Rényi Institute of Mathematics. `http://www.renyi.hu/pub/algebraic-logic/Contents.html`.

Andréka, H., J. X. Madarász, and I. Németi, 2006, Logical axiomatizations of space-time. Samples from the literature. In A. Prékopa, & al. (eds.) *Non-Euclidean Geometries*. Berlin, Springer, 155–185.

Andréka, H., J. X. Madarász, and I. Németi, 2007, Logic of space-time and relativity theory. In M. Aiello, & al. (eds.), *Handbook of Spatial Logics*. Berlin, Springer, 607–711.

Andréka, H., J. X. Madarász, I. Németi, and G. Székely, 2008, Axiomatizing relativistic dynamics without conservation postulates. *Studia Logica* 89, 163–186.

Andréka, H., I. Németi, and P. Németi, 2009, General relativistic hypercomputing and foundation of mathematics. *Nat. Comp.* 8, 499–516.

Andréka, H., J. X. Madarász, I. Németi, and G. Székely, 2010, A logic road from special relativity to general relativity. *Synthese*, submitted.

Ax, J., 1978, The elementary foundations of spacetime. *Found. Phys.* 8, 507–546.

Benda, T., 2008, A formal construction of the spacetime manifold. *J. Philos Logic* 37, 441–478.

Chang, C. C., and H. J. Keisler, 1990, *Model theory*. Amsterdam, North-Holland.

d'Inverno, R., 1992, *Introducing Einstein's relativity*. Oxford, Oxford Univ. Press.

Einstein, A., 1905/1952, Zur Elektrodynamik bewegter Körper. *Annalen der Physik* 17, 891–921. English translation in A. Einstein, *The principle of Relativity*. Mineola (NY), Dover.

Einstein, A., 1921/2006, *Relativity. The Special and the General Theory*. London, Penguin Classics. Translated by W. Lawson.

Friedman, H., 2004a, On foundational thinking 1. Posting in FOM (Foundations of Mathematics) Archives `www.cs.nyu.edu` (Jan. 20, 2004).

Friedman, H., 2004b, On foundations of special relativistic kinematics 1. Posting in FOM (Foundations of Mathematics) Archives `www.cs.nyu.edu` (Jan. 21, 2004).

Goldblatt, R., 2004, *Orthogonality and spacetime geometry*. Berlin, Springer.

Gömöri, M., and L. E. Szabó, 2010, Is the relativity principle consistent with electrodynamics? Towards a logico-empiricist reconstruction of a physical theory. arXiv:0912.4388v3.

Guts, A. K., 1982, The axiomatic theory of relativity. *Russ. Math. Surv.* 37, 41–89.

Herschbach, D., 2008, Einstein as a student. In P. L. Galison & al. (eds.) *Einstein for the 21st century*. Princeton, Princeton Univ. Press, 217–238.

Hirsch, R., 2009, Relativity and modal logic. *Hungarian Philosophical Review*, this issue.

*This research is supported by the Hungarian Scientific Research Fund for basic research grant No. T81188, as well as by a Bolyai grant for J. X. Madarász.

Madarász, J. X., 2002, *Logic and Relativity: in the Light of Definability Theory*. PhD thesis, Budapest, Eötvös Loránd Univ.

Madarász, J. X., I. Németi, and G. Székely, 2006, Twin paradox and the logical foundation of relativity theory. *Found. Phys.* 36, 681–714.

Madarász, J. X., I. Németi, and G. Székely, 2007, First-order logic foundation of relativity theories. In D. Gabbay, & al. (eds.), *Mathematical Problems from Applied Logic II*. Berlin, Springer, 217–252.

Misner, C. W., K. S. Thorne, and J. A. Wheeler, 1973, *Gravitation*. New York, W. H. Freeman and Co.

Rosinger, E. E., 2008, *Two Essays on the Archimedean versus Non-Archimedean Debate*. arXiv:0809.4509v3.

Rosinger, E. E., 2009, *Special Relativity in Reduced Power Algebras*. arXiv:0903.0296v1.

Schutz, J. W., 1973, *Foundations of Special Relativity: Kinematic Axioms for Minkowski Space-Time*. Berlin, Springer.

Suppes, P., 1968, The desirability of formalization in science. *J. Philos.* 27, 651–664.

Szabó, L. E., 2009, Empirical Foundation of Space and Time. In M. Suárez, & al. (eds.), EPSA07: *Launch of the European Philosophy of Science Association*. Berlin, Springer.

Székely, G., 2009, *First-Order Logic Investigation of Relativity Theory with an Emphasis on Accelerated Observers*. PhD thesis, Budapest, Eötvös Loránd Univ.

Székely, G., 2010a, On why-questions in physics. In F. Stadler & al., (ed.), *Wiener Kreis und Ungarn*. Berlin, Springer, to appear.

Székely, G., 2010b, A geometrical characterization of the twin paradox and its variants. *Studia Logica*, online-first.

Tarski, A., 1959, What is elementary geometry? In L. Henkin, & al. (eds.) *The Axiomatic Method. With Special Reference to Geometry and Physics*. Amsterdam, North-Holland, 16–29.

Väänänen, J., 2001, Second-order logic and foundations of mathematics. *Bull. Symb. Log.* 7, 504–520.

Wolenski, J., 2004, First-order logic: (philosophical) pro and contra. In V. F. Hendricks & al. (eds.), *First-Order Logic Revisited*. Berlin, Logos, 369–398.

Robin Hirsch

Modal Logic and Relativity

Abstract. We argue that modal logic is the natural logic to use to reason about relativity theory. We define a complete modal axiomatisation of the kinematics of special relativity theory.

Relativity Theory, in its most general sense, rejects the notion of absolute space. In Relativity Theory statements such as "this rod is one meter long" are frowned upon, rather we should say "this rod is one meter long when measured in this frame of reference". When devising a logic to reason about relativity theory, it is therefore natural to adopt a modal logic where every statement has an implicit 'point of view'. In that sense, a modal logic might be more true to the subject matter of relativity theory. We do not claim that such an approach will in itself provide new technical results in relativity theory, but apparent paradoxes and other conceptual difficulties with relativity theory might be more easily avoided in a modal setting. A second motivation for modal logic is that the complexity of reasoning in a modal logic can often be lower than with first-order logic. Thirdly, the current article is only an initial step and we believe that modal logic should be used to reason about general relativity where it is even more important to work in a local framework. In this article we will consider how a modal logic for special relativity theory might be devised.

A fundamental concept in relativity theory is that of the observation. For Einstein, and his group of followers in the *Vienna Circle*, the precise nature of an observation was of great importance. In Einstein's original paper on Special Relativity (Einstein 1905), he takes care to clarify the meaning of words like "time", "simultaneous", "length", etc. by replacing them by statements concerning observers and observations. Later, *Logical Positivists* formulated the *verification principle*, which stated that a proposition could be held to be true to the extent that it could be tested by experiments and observations. The critical thing about observations in relativity theory is that what is observed depends

not only on the event but on the observer too. Later, we will define a modal logic with a Kripke semantics in which each observer is a Kripke world.

One intriguing feature of observers is that they act both as subjects and as objects—they see, but they can also be seen. In Einstein's original paper, he refers to an observer as "the man at the railway-carriage window", suggesting a point-like body moving through space. In most presentations of relativity theory, an observer is a point-like body with its own world line in 4D spacetime. This tells us that when we see an observer, he looks one-dimensional. We see the observer's time axis, but we do not see his space axes. However, from his point of view, an observer can see various events taking place at various spacetime points distributed throughout the four dimensions of spacetime. Indeed, when we think of an observer as a subject, it is better not to think of an individual point-like body, but a whole team of colleagues arranged in a grid in three-dimensional space, not moving relative to each other over time, who send messages to each other (or perhaps to a central control centre) about their immediate observations. But as we mentioned, only one member of the team of observers is directly visible to observers from other reference frames. These two aspects of observers are related in (Andréka *& al.* 2010) by an axiom that requires that all observers see themselves at the space origin of their reference frame, at all times, i.e. their world line is the time axis.

Here we consider two different applications of modal logic to special relativity theory: the first approach has to do with modal frame definability, and the second uses model definability. To take a simple and perhaps more familiar example of frame definability, the logic **S4** with axioms $\{\Box(p \to q) \to (\Box p \to \Box q),\ \Box p \to p,\ \Box p \to \Box\Box p\}$ and inference rules modus ponens and necessitation defines the set of all validities over Kripke frames whose accessibility relation is reflexive and transitive, that is, **S4** *defines the validities* of the class of reflexive transitive frames. Furthermore, if these axioms are valid over any Kripke frame (W, R), then R is transitive and reflexive over W, i.e. **S4** *defines the class* of reflexive transitive frames. For model definability, if we take the class of reflexive transitive frames as given, the formula $\Box(p \to \Diamond\neg p)$, which is not valid over any frame, *defines those models* where p is never 'eventually true'.

For special relativity, the frame definability approach has already been considered. A Kripke frame may be defined whose Kripke worlds are the points of a four-dimensional Minkowski spacetime and where the accessibility relation is "can send a signal to" and the problem is to define a modal logic that derives all modal formulas valid over frames with this accessibility relation. If we require that signals travel at less than the speed of light and that a signal may be sent from a spacetime point to itself (reflexivity of the accessibility relation) then the set of modal validities is **S4.2** (Goldblatt 1980; Shehtman 1983), the logic obtained by adding the axiom $\Diamond\Box p \to \Box\Diamond p$ to the axioms of **S4**. However, **S4.2** does not define this class of four-dimensional models, as the same set of

validities hold over Minkowski spacetime of dimension $2, 3, \ldots$, over Galilean models or even over models with discrete time. **S4.2** actually defines the class of all reflexive, transitive and *directed* frames. If you consider an irreflexive accessibility relation "can send a signal to a different point" then modal logic can be more discriminating, however, the purpose of all this is to axiomatise the validities of the given accessibility relation, it does not express the lengths of rods or how they may be transformed when viewed by another observer.

The main focus of the current article is about model definability. We seek a modal logic which is able to express the kinds of observations we might want to make in four-dimensional spacetime, and it should also be able to express the kinds of observations we might expect other observers to make. It can be considered as part of a line of research which seeks to provide simple logical axioms from which the main theorems of relativity theory can be derived. Many such logics have been devised, we make no attempt to survey them here, see (Andréka *& al.* 2006) for an extensive survey. In general, previous attempts to provide a logic for relativity adopt a first-order, or in many cases second-order (e.g. the axiom of continuity in (Schutz 1997)) logic. Consequently they adopt Tarskian semantics. Global variables may be used to denote bodies, field values etc. Instead of the Newtonian statement "Body b is at (x, y, z) at time t", the dependence of an observation on the observer is expressed by an observation predicate W, so that $W(o, b, x, y, z, t)$ expresses "observer o sees body b at (x, y, z, t)".

In a modal logic in which the role of Kripke world is taken by an inertial frame of reference, the dependence on the observer will be suppressed. The modal operator $\Diamond$ will permit us to transfer from one intertial frame to another. We will take as given a frame in which the accessibility relation is the universal relation. But what language should be adopted to describe the observations made within a single reference frame? In order to keep the presentation fairly general, here we use a two sorted first-order logic with one sort for bodies (B) and the other for quantities (Q). Variables of sort Q will be used to record coordinate values. We may use subscripts b, q for constants, functions and predicates to indicate their sort. The predicate See_{q^4b} requires four coordinates and one body and tells us that this body is observed at the four coordinate values. We avoid a number of difficulties that sometimes arise with modal first-order languages by requiring that the first-order variables have the same domain at each world of a structure.

The Language

Variables:	$b, c, \ldots$ (type B), $x, y, z, \ldots$ (type Q)
Constants/functions:	$0, 1, +, \times$ (type Q)
Predicates:	$Ph_b, Obs_b, \leq_{qq}, \mathsf{See}_{q^4b}$
Formulas:	$\phi ::= Atom \mid \neg\phi \mid (\phi_1 \vee \phi_2) \mid \exists var\phi \mid \Diamond\phi$

The following abbreviations will be useful later.

$$\mathsf{vel}(b) = (v_0, v_1, v_2) \quad \text{means} \quad \exists \overline{x}\ \forall \overline{y} \\ [\mathsf{See}(\overline{y}b) \leftrightarrow \exists \lambda\ \overline{y} = \overline{x} + \lambda \times (v_0, v_1, v_2, 1)]$$

$$|(x_0, x_1, x_2)| \quad \text{means} \quad \sqrt{x_0^2 + x_1^2 + x_2^2}$$

Structure As promised, our semantics will be based on Kripke-like structures, where the Kripke worlds are inertial frames of reference. Thus a structure will have the form

$$(W, \beta, F, I, W \times W)$$

where W is the set of Kripke worlds, β is the set of bodies, F is the set of quantities, I interprets variables as elements of β or F (depending on the sort of the variable), $0, 1, +, \times, \leq$ as functions/predicates on F, and I satisfies $I(Ph) \cup I(Obs) \subseteq \beta$, $I(\mathsf{See}) \subseteq W \times F^4 \times \beta$. As we mentioned before, the sets β, F are the same for all worlds (constant domains). Further, all constants, functions and predicates in our language are rigidly designated, except for $\mathsf{See}_{q^4 b}$ which is expected to vary from one world to another. Given such a structure, we may evaluate formulas in the obvious way. Let $\mathcal{S} = (W, \beta, F, I, W \times W)$ and $w \in W$.

$$\begin{aligned}
\mathcal{S} \models Obs(b) &\iff I(b) \in I(Obs) \\
\mathcal{S} \models t \leq s &\iff (I(t), I(s)) \in I(\leq) \\
\mathcal{S}, w \models \mathsf{See}(x, y, z, t, b) &\iff (w, I(x, y, z, t, b)) \in I(\mathsf{See}) \\
\mathcal{S}, w \models \exists x \phi &\iff (W, \beta, F, I', W \times W), w \models \phi \\
&\qquad \text{(some } I' \text{ that agrees with } I \text{ except perhaps on } x) \\
\mathcal{S}, w \models \Diamond \phi &\iff \mathcal{S}, v \models \phi \text{ (some } v \in W)
\end{aligned}$$

Axioms for defining class of structures $\Box$-$\forall$-closure of:

(1) $(F, 0, 1, +, \times, \leq)$ is a Real Closed Field, i.e. an ordered field where every non-negative element has a square root (Euclidean) and every polynomial of odd degree has a root.

In many presentations of special relativity, the ordered field is only required to be Euclidean and this suffices for most of our results. Here we assume that the field is a Real Closed Field in order to obtain a decidability result. Note that all Real Closed Fields are elementarily equivalent to the real numbers (Tarski 1951).

(2) Observers are inertial: $Obs(b) \rightarrow \exists v_0 v_1 v_2 (\mathsf{vel}(b) = (v_0, v_1, v_2))$

(3) Speed of light is constant: $Ph(b) \to |\mathsf{vel}(b)| = 1$

(4) There is an observer on every 'slow line', there is a photon on every 'fast line'.

$$|(v_0, v_1, v_2)| < 1 \to \exists b(Obs(b) \wedge \mathsf{See}(\overline{x}, t, b) \wedge \mathsf{vel}(b) = (v_0, v_1, v_2))$$
$$|(v_0, v_1, v_2)| = 1 \to \exists b(Ph(b) \wedge \mathsf{See}(\overline{x}, t, b) \wedge \mathsf{vel}(b) = (v_0, v_1, v_2))$$

In view of the last axiom, for each observer and for all $x, y, z, t \in F$, there are bodies b_1, b_2, b_3 moving on distinct lines that meet uniquely (pairwise and jointly) at (x, y, z, t), according to that observer. The triple $e = (b_1, b_2, b_3)$ is called an event, we may write $\mathsf{See}(x, y, z, t, e)$ instead of $\bigwedge_{i=1,2,3} \mathsf{See}(x, y, z, t, b_i)$.

(5) All observers see the same events:

$$\mathsf{See}(x, y, z, t, e) \to \Box\exists x, y, z, t\ \mathsf{See}(x, y, z, t, e).$$

(6) Symmetry. Let e_0, e_1, e'_0, e'_1 be events.

$$\bigwedge_{i=0,1} (\mathsf{See}(0, 0, 0, i, e_i) \wedge \mathsf{See}(x'_i, y'_i, z'_i, t'_i, e'_i)) \to$$
$$\Box(\bigwedge_{i=0,1} (\mathsf{See}(0, 0, 0, i, e'_i) \wedge \mathsf{See}(x_i, y_i, z_i, t_i, e_i)) \to (t'_1 - t'_0 = t_1 - t_0))$$

(7) Isotropy: $(\bigwedge_{i,j<4}(|\overline{x}_i - \overline{x}_j| = |\overline{x}'_i - \overline{x}'_j|) \wedge \bigwedge_{i<4} \mathsf{See}(\overline{x}_i, t, b_i)) \to \Diamond(\bigwedge_{i<4} \mathsf{See}(\overline{x}'_i, t', b_i))$

where $\overline{x}_i$ is a triple of three spatial coordinates, for $i < 4$.

The symmetry axiom implies that any two observers see each other's clocks slow at the same rate. This usefully rules out a situation where one observer measures in seconds while another observer measures in years, it also rules out the situation where one observer measures time going forward while the other measures time going backwards. The isotropy axiom relates to the difference between the one dimensional appearance of observers and the four dimensions that an observer sees. Recall that when we see an observer, we see only his time axis, we do not see the orientation of his spatial axes. The isotropy axiom states that if I can see four bodies at $\overline{x}_0, \ldots, \overline{x}_3$ at time t, and if the spatial Euclidean distances between the $\overline{x}_i$ are identical to the distances between the $\overline{x}'_i$, then another observer can see the same four bodies at $\overline{x}'_0, \ldots, \overline{x}'_3$ at time t'. I can transform myself to the second observer by performing an isometry of the spatial coordinates followed by a time translation. Let Ax be the set of six axioms, just defined.

Having defined the semantics of our language and the axioms for our logic, we now briefly evaluate our system by three criteria: how expressive is the language? are the axioms complete over an appropriate class of structures? what is the complexity of the satisfiability problem for formulas in our language,

over Minkowski structures? As far as expressive power is concerned, it seems that this language is capable of expressing most of the technical statements you find in a textbook on special relativity. For example, given a velocity vector $\overline{v} = (v_0, v_1, v_2)$ and any formula ψ we may write $\Diamond_{\overline{v}}\psi$ for $\exists b(\mathsf{vel}(b) = \overline{v} \wedge \Diamond(\mathsf{vel}(b) = 0 \wedge \psi))$, which means "there is a frame moving with velocity $\overline{v}$ and ψ is true in that frame". Our language should be able to express the purely kinematic properties of special relativity. However, our language can only express kinematic statements, we are not able to express properties relating to mass, energy or electric charge, for example.

Next, we assess the deductive strength of our axioms.

Lemma 1. *Let* $\mathcal{S} = (W, \beta, F, I, W \times W) \models$ *Axioms 1--6 and let* $w, v \in W$. *There is a Poincaré map* $p : F^4 \to F^4$ *(an isometry with respect to Minkowski distance in* F^4*) such that*

$$\mathcal{S}, w \models \mathsf{See}(x, y, z, t, b) \iff \mathcal{S}, v \models \mathsf{See}(p(x, y, z, t)) \tag{1}$$

PROOF:

A map $p : F^4 \to F^4$ satisfying (1) is uniquely defined, since $\mathcal{S} \models 4, 5$. We have to show that p is a Poincaré map. By axioms 2 and 3, for any $b \in I(Obs) \cup I(Ph)$, the set $\{(x, y, z, t) : \mathcal{S}, w \models \mathsf{See}(x, y, z, t, b)\}$ is a line of F^4 and by axiom 4 each light line of F^4 is the trace of a photon and each slow line is the trace of an observer. Thus p maps lines to lines and maps light lines to light lines. By the Alexandrov-Zeeman theorem, p is a Poincaré transformation followed by a dilation and a field automorphism induced transformation. By axiom 6, the dilation and field induced transformations must be the identity transformations. □

The isotropy axiom, axiom 7, is needed to complete the proof of the next lemma.

Lemma 2. *Let* $\mathcal{S} \models Ax$, $\mathcal{S}' \models Ax$ *be two models of our axioms, where* $\mathcal{S} = (W, \beta, F, I, W \times W)$ *and* $\mathcal{S}' = (W', \beta', I', F', W' \times W')$. *There are maps* $i : W \to W'$, $j : \beta \to \beta'$ *and* $k : F \to F'$ *such that* k *is an ordered field embedding, and for all* $w \in W$, $b \in \beta$,

$$\mathcal{S}, w \models \mathsf{See}(x, y, z, t, b) \iff \mathcal{S}', i(w) \models \mathsf{See}(k((x), k(y), k(z), k(t), j(b))$$

Next, we define a standard model $\mathcal{M}$. The set of worlds of $\mathcal{M}$ is the set of Poincaré transformations of $\mathbb{R}^4$. If p is a Poincaré transformation and l is a line then we write $p(l)$ for the image of l under p, note that $p(l)$ is always itself a line. The set of bodies of $\mathcal{M}$ is L, the set of lines of $\mathbb{R}^4$ of gradient at most one, observers are the slow lines of gradient strictly less than one and photons are the lines of gradient one. The interpretation I_M in the standard model is given by

$$(p, x, y, z, t, b) \in I_M(\mathsf{See}) \iff (x, y, z, t) \in p(b)$$

Theorem 1.

(1) *$\mathcal{M}$ is a model of Ax.*
(2) *For any formula ϕ of our language we have*

$$\mathcal{M} \models \phi \iff Ax \vdash \phi$$

The first part can easily be proved, simply by verifying that each of our axioms holds in $\mathcal{M}$. The second part follows from lemma 2. Thus our axioms are complete for the validities of the standard model.

We now consider the decidability of the following decision problem: Is modal formula ϕ satisfiable over models of Ax? If we allow arbitrary bodies to exist in our models, then this problem is undecidable. We may reduce the tiling problem to this satisfiability problem by considering each tile as a body that may be observed at positions with integer coordinates and where the adjacencies are expressed by universally quantified modal formulas. This undecidability arises purely from the first-order part of our language and can be proved without using the modalities. However, if we restrict to bodies whose paths are described by polynomial equations, then the problem becomes decidable. Since we have a universal modality, an arbitrary modal formula ϕ may be converted to a disjunction of clauses of the form $\Box\psi \wedge \bigwedge_i \Diamond\psi_i$, where ψ, ψ_i are non-modal. Such a clause is satisfiable iff $\psi \wedge \psi_i$ is satisfiable, for each i. By our assumption about the trajectories of bodies, such a non-modal formula may be translated into a first-order formula in a language with a binary predicate $\leq$ and functions $+, \times$. Tarski showed (Tarski 1951; Canny 1988), by elimination of quantifiers, that the satisfiability problem for this language over real closed fields is decidable, although the complexity is rather high (at least double exponential). By imposing restrictions on the use of the first-order connectives in our language, satisfiability problems with lower complexities may be obtained.

Questions

- As an alternative to the current exposition, consider a modal logic for special relativity theory where an observer sees a body not merely as a line in four-dimensional space but as a line with a spatial orientation, so at an instance we see a point and three spatial unit vectors.
- Can we define models of special relativity without variables? One approach would be to use a propositional modal logic with **S5** modalities for moving in the directions of each of the spatial unit vectors and a temporal modality for moving in time, along with the already included modality for changing reference frame. It is known that this class of frames cannot be finitely axiomatised and the equational theory is undecidable (Hirsch *& al.* 2002), but here we are more interested in model definability.

- Can we use a similar framework to define a modal logic of general relativity? A Kripke world could be considered as an open neighbourhood in a coordinate system and the whole Kripke frame would correspond to a manifold. An extra modality to transfer to an adjacent neighbourhood would be needed. See (Shapirovski and Shehtman 2005) for useful results on modal logics of regions in Minkowski spacetime.

REFERENCES

Andréka, H., Madarász, J. and Németi, I., 2010, *On the Logical Structure of Relativity Theories (Making Relativity Modular, Changeable, and Easy)*. Dordrecht, Kluwer. Provisionally accepted to appear with Kluwer. Draft available at `http:www.math-inst.hu/pub/algebraic-logic/contents.html`.

Andréka, H., Madarász, J. X. and Németi, I., 2006, Logical axiomatizations of space-time. Samples from the literature. In A. Prékopa, et al. (eds.) *Non-Euclidean Geometries*. Berlin, Springer, 155–185.

Canny, J., 1988, Some algebraic and geometric computations in pspace. In *Proc. of the 20th SCM symposium on theory of computing*, 460–467.

Einstein, A., 1905, Zur Elektrodynamik bewegter Körper. *Annalen der Physik 17*. 891–921. English translation available at `http://www.fourmilab.ch/etexts/einstein/specrel/www/`.

Goldblatt, R., 1980, Diodorean modality in Minkowski space-time. *Studia Logica* 39, 219–236.

Hirsch, R., Hodkinson, I., and Kurucz, A., 2002, On modal logics between $K \times K \times K$ and $S5 \times S5 \times S5$. *J. Symbolic Logic* 67, 221–234.

Schutz, J., 1997, *Independent Axioms for Minkowski Space-Time*. Pitman Research Notes in Mathematics. New York, Longman.

Shehtman, V., 1983, Modal logics of domains on the real plane. *Studia Logica* 64, 63–88.

Shapirovski, I. and Shehtman, V., 2005, Modal logics of regions and Minkowski spacetime. *Journal of Logic and Computation* 15, 559–574.

Tarski, A., 1951, *A Decision Method for Elementary Algebra and Geometry*. Berkeley, Univ. California Press.

Máté Szabó

On Field's Nominalization of Physical Theories

Abstract. In his book *Science Without Numbers* Hartry Field argues that we can "nominalize" our physical theories, that is we can reformulate them in such a way that (1) the new version preserves the attractivity of the theory, and (2) the nominalized theory does not contain quantification over mathematical entities. I reconsider Field's nominalization procedure for a toy physical theory formulated in a first order language, in order to make a clear distinction between the following three steps: (1) the physical theory in terms of empirical observations; (2) the standard physical theory, which contains quantification over mathematical entities, as usual; (3) the nominalized version of the theory without any reference to mathematical entities. Having Field's nominalization procedure reconstructed, it will be clear that from a formalist point of view there is no difference between the original and the nominalized versions of the theory. It is because the only difference would come from the different "meanings" of the variables over which the quantification is running. The formalist philosophy of mathematics, however, denies that the variables have meanings at all. Finally, I consider further arguments for and against the indispensability of mathematical objects.

1 INTRODUCTION

One of the most important questions in the philosophy of mathematics is the ontological status of mathematical entities. In the late 1970s, Quine and Putnam suggested an argument for the existence of mathematical entities. The argument is based on the idea that mathematics is not only applicable but in fact *indispensable* for the empirical sciences; and if that is the case, then mathematical entities are as indispensable for our best ontological picture of the world as electrons and other physical entities, to the existence of which physicists are committed.

In its most explicit form the argument reads as follows:

Indispensability argument

(P1) We ought to have ontological commitment to all and only the entities that are indispensable to our best scientific theories.
(P2) Mathematical entities are indispensable to our best scientific theories.

(C) We ought to have ontological commitment to mathematical entities.

The argument has attracted a great deal of attention. Many Platonists regard it as the best available argument for the existence of mathematical entities. The opponents of the argument object, first of all, to the first premise; while the second premise is considered uncontroversial (Colyvan 2004).

What does the first premise exactly mean? First of all, it definitely presupposes a kind of naturalism. For naturalism claims that we have to have ontological commitment to all and only the entities that exist according to our best scientific theories. According to Quine—see (Quine 1961), (Quine 1981)—if the language of the scientific theory quantifies over some entities which are, at the same time, indispensable, then we ought to have ontological commitment to those entities. It is therefore necessary to clarify the proper meaning of "indispensability".

It would be quite straightforward to interpret dispensability as eliminability. An entity is eliminable from a theory if there is another theory which is empirically equivalent to the original one, but does not quantify over the entity in question. In this case, however, every non-observable entity would be dispensable, due to the well-known Craig theorem; see (Craig 1953). In other words, we would have to reject the existence of anything but the directly observable entities. In order to avoid such a radical conclusion, many suggest, we need to prescribe some further requirements for the new theory. These requirements are usually the following: clarity, simplicity, unificatory power, generality and fecundity ((Burgess 1983), (Maddy 2005)). These requirements altogether are called *attractivity*. So, an entity is dispensable if it is eliminable and the theory we obtain by its elimination remains attractive. In any event, the notion of attracivity is quite ambiguous; and it is hard to believe that the most fundamental ontological questions depend on such unclear and sociologically relative notions.

On the other hand, the second premise was considered as an evident one. Hartry Field was the first who claimed that the second premise is, in fact, false; mathematical entities are not indispensable to our best scientific theories. Field adopted Quine's linguistic criterion that a scientific theory asserts the existence of an entity by quantifying over the entity in question. He also accepted the attractivity requirements. But he showed that a physical theory can always be "nominalized", by which he means that it can be reconstructed such that (1) the new theory does not contain quantification over mathematical entities, however, (2) remains attractive.

2 THE NOMINALIZATION PROCEDURE

In his *Science Without Numbers*(Field 1980) Field showed, as an example, how a fragment of Newtonian gravitational theory can be nominalized. In what follows, I will reconstruct Field's nominalization procedure in the case of an even simpler "toy" physical theory. My purpose is not only to demonstrate the nominalization steps on a perhaps more clear-cut example, but also to lay the emphasis on different points. First, within a physical theory, we will make a clear separation of the formal system and the semantics. Second, we will keep it clear that the equivalence of physical theories is understood as *empirical* equivalence (Fig. 1).

The nominalization procedure consists of the following three steps:

(1) We have a body of physical facts, in terms of empirical observations.
(2) We have the usual platonistic physical theory describing the observable phenomena in question—containing quantification over mathematical entities. The platonistic theory will consist of a formal system L, and a semantics S.
(3) We construct a new theory which is capable of describing the same phenomena, but without quantification over mathematical entities. The nominalized theory will consist of a formal system L', and a semantics S'. We will show the equivalence of these theories on the level of observable phenomena.

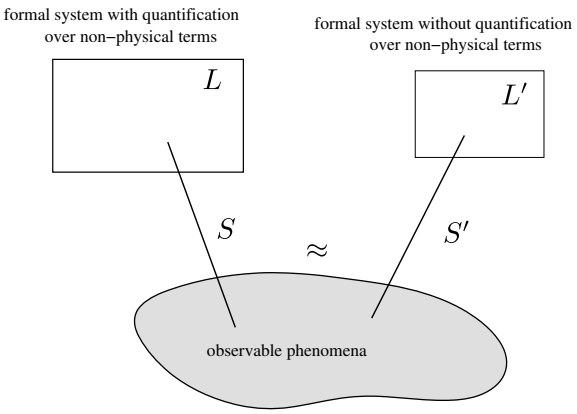

Figure 1. The platonistic physical theory and the nominalized physical theory should be equivalent on the empirical level.

3 THE "TOY" PHYSICAL THEORY

The "toy" physical theory is about a few empirically observable regularities related with the spatial relations of properties of the material points/molecules of a sheet of paper. Thus the only physical entities are the molecules of the paper and the only measuring equipment will be a scale-free ruler. We will examine—empirically—the following two properties of the molecules:

Betweenness We say that molecule γ is between molecules α and β if whenever the ruler fits to α and β then it also fits to γ and the mark on the ruler corresponding to γ falls between the marks corresponding to α and β.

Congruence We will say that a pair of molecules α, β is congruent to a pair of molecules γ, δ if whenever we mark the ruler at α, β, the same marks will also fit to γ, δ.

With our scale-free ruler we can observe the following empirical facts about the molecules of the paper:

(E1) If molecules α and β are congruent to molecules γ and δ, and γ and δ are congruent to molecules ε and ζ, then α and β are congruent to ε and ζ.

(E2) If we consider three molecules fitting to the ruler, then there is exactly one that lies between the other two.

4 THE USUAL PLATONISTIC PHYSICAL THEORY OF THE PAPER

We present a "platonistic" physical theory (L, S) which describes the empirical facts of the paper. The formal system of the physical theory will be $L = (\mathbb{R}^2, \Gamma, \Lambda)$, where

$\Gamma(a, b, c)$ is a relation between three points of $\mathbb{R}^2$ (six real numbers):

$$\Gamma(a,b,c) \Longleftrightarrow \sqrt{(a_1-b_1)^2+(a_2-b_2)^2}+\sqrt{(c_1-b_1)^2+(c_2-b_2)^2}$$

$$=\sqrt{(a_1-c_1)^2+(a_2-c_2)^2}$$

$\Lambda(a, b, c, d)$ is a relation between four points of $\mathbb{R}^2$ (eight real numbers):

$\Lambda(a,b,c,d) \Longleftrightarrow$

$$\sqrt{(a_1-b_1)^2+(a_2-b_2)^2}=\sqrt{(c_1-d_1)^2+(c_2-d_2)^2}$$

The semantics S is defined as follows: First, to every molecule we assign an element of $\mathbb{R}^2$: α corresponds to $a = (a_1, a_2) \in \mathbb{R}^2$, β to $b = (b_1, b_2) \in \mathbb{R}^2$, and so on. Second, relation Γ corresponds to the *Betweenness* and Λ corresponds to the *Congruence* of the molecules. All this representation is carefully constructed so that the physical theory (L, S), that is, the formal system $(\mathbb{R}^2, \Gamma, \Lambda)$ with the

above semantics provides a proper description of our empirical knowledge about the paper. It means that if $\Gamma(a, b, c)$ is true for $a, b, c \in \mathbb{R}^2$ then it is true for the corresponding molecules α, β and γ that molecule β is between α and γ. Similarly, if $\Lambda(a, b, c, d)$ is true for $a, b, c, d \in \mathbb{R}^2$ then it is true that molecules α, β are congruent to molecules γ, δ.

For example, empirical facts (E1) and (E2) are obviously represented in the theory (L, S). Moreover, (L, S) has predictive power. For instance, in $(\mathbb{R}^2, \Gamma, \Lambda)$ we can prove the following theorem (Fig. 2):

Theorem 1.

$$\forall a \forall b \forall g \forall d \forall e \forall z \exists o\ \Gamma(a, d, b) \wedge \Gamma(b, e, g) \wedge \Gamma(g, z, a)$$
$$\wedge \Lambda(a, d, d, b) \wedge \Lambda(b, e, e, g) \wedge \Lambda(g, z, z, a)$$
$$\rightarrow \Gamma(a, o, e) \wedge \Gamma(b, o, z) \wedge \Gamma(g, o, d)$$

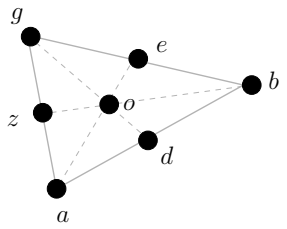

Figure 2. The centroid of a triangle always exists.

With the above semantics, we arrive at the following *hypothesis* about the molecules of the paper (Fig. 3):

Hypothesis If molecules $\alpha, \beta, \gamma, \delta, \varepsilon$ and ζ satisfy that δ is between α and β, ε is between β and γ, and ζ is between γ and α, furthermore, α, δ are congruent to δ, β, and β, ε are congruent to ε, γ, and γ, ζ are congruent to ζ, α, then we can always find a molecule ω such that it is between α and ε, and it is between β and ζ and it is between γ and δ.

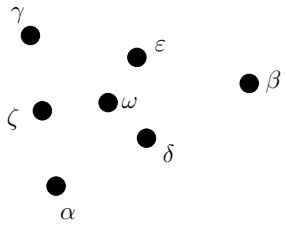

Figure 3. According to the semantics of the theory, **Theorem 1.** (Fig. 2) is a statement about the molecules of the paper.

This hypothesis can be verified empirically by means of the scale-free ruler; and we will find that the hypothesis is true.

In Field's terminology, (L, S) is a *platonistic* theory: It contains quantification over mathematical entities, namely, over real numbers, since Γ and Λ are relations between real numbers.

5 THE NOMINALIZED THEORY

Now we will construct another physical theory which can equally well describe the same observable phenomena but without quantification over mathematical entities. This will consist of another formal system L' with another semantics S'.

The formal language L' will be a first order formal system. L' will contain individuum variables $A, B, C, \ldots$ and a three-argument predicate symbol Bet and a four-argument predicate symbol $Cong$. Beyond the logical axioms of PC(=) (predicate calculus with identity), we will need the following "physical" axioms:[1]

T1: $\forall A \forall B\ Cong(A, B, B, A)$

T2: $\forall A \forall B \forall C\ Cong(A, B, C, C) \rightarrow A = B$

T3: $\forall A \forall B \forall C \forall D \forall E \forall F\ Cong(A, B, C, D) \wedge Cong(C, D, E, F)$
$\rightarrow Cong(A, B, E, F)$

T4: $\forall A \forall B\ Bet(A, B, A) \rightarrow A = B$

T5: $\forall A \forall B \forall C \forall D \forall E\ Bet(A, D, C) \wedge Bet(B, E, C))$
$\rightarrow \exists F\ (Bet(D, F, B) \wedge Bet(E, F, A)$

T6: $\exists E \forall A \forall B\ \phi(A) \wedge \psi(B) \rightarrow Bet(E, A, B)$
$\rightarrow \exists F \forall A \forall B\ \phi(A) \wedge \psi(B) \rightarrow Bet(A, F, B)$
where ϕ and ψ are two arbitrary formulas of the language, containing no free instances either E or F. Let there also be no free instances of A in $\psi(B)$ or of B in $\phi(A)$.

T7: $\exists A \exists B \exists C\ \neg Bet(A, B, C) \wedge \neg Bet(B, C, A) \wedge \neg Bet(C, A, B)$

T8: $\forall A \forall B \forall C \forall D \forall E\ Cong(A, D, A, E) \wedge Cong(B, D, B, E)$
$\wedge Cong(C, D, C, E) \wedge \neg D = E$
$\rightarrow Bet(A, B, C) \vee Bet(B, C, A) \vee Bet(C, A, B)$

T9: $\forall A \forall B \forall C \forall D \forall E\ Bet(A, D, E) \wedge Bet(B, D, C) \wedge \neg A = D$
$\rightarrow \exists F \exists G\ Bet(A, B, F) \wedge Bet(A, C, G) \wedge Bet(G, E, F)$

T10: $\forall A \forall B \forall C \forall D \forall E \forall F \forall G \forall H\ \neg A = B \wedge Bet(A, B, C) \wedge Bet(E, F, G)$
$\wedge Cong(A, B, E, F) \wedge Cong(B, C, F, G) \wedge Cong(A, D, E, H)$
$\wedge Cong(B, D, F, H) \rightarrow Cong(C, D, G, H)$

T11: $\forall A \forall B \forall C \forall D \exists E\ Bet(D, A, E) \wedge Cong(A, E, B, C)$

[1]The reader may recognize that these are nothing but the well known Tarski–Givant (1999) axioms of Euclidean geometry. But it must be emphasized that this fact is irrelevant.

The S' semantics of the theory is defined as follows. The individuum variables $A, B, C, \ldots$ will refer to the molecules of the paper. The predicate symbol Bet corresponds to the *Betweenness*, and the predicate symbol $Cong$ corresponds to the *Congruence*.

The physical theory (L', S') with the above semantics provides a proper description of our empirical knowledge about the paper. For example, empirical facts (E1) and (E2) are obviously represented by theorems in (L', S'). This theory equally well describes our empirical knowledge about the paper. It also has the same predictive power. For instance, in L' we can prove the following theorem Fig. 4:

Theorem 1′.

$$\forall A \forall B \forall G \forall D \forall E \forall Z \exists O \; Bet(A, D, B)$$
$$\wedge Bet(B, E, G) \wedge Bet(G, Z, A) \wedge Cong(A, D, D, B)$$
$$\wedge Cong(B, E, E, G) \wedge Cong(G, Z, Z, A)$$
$$\rightarrow Bet(A, O, E) \wedge Bet(B, O, Z) \wedge Bet(G, O, D)$$

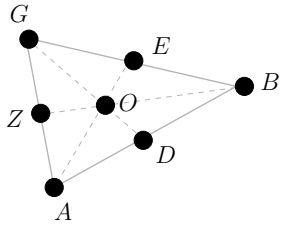

Figure 4. The centroid of a triangle always exists.

With the above semantics, this leads us to the following hypothesis about the molecules of the paper (Fig. 5):

Hypothesis If molecules $\alpha, \beta, \gamma, \delta, \varepsilon$ and ζ satisfy that δ is between α and β, ε is between β and γ, and ζ is between γ and α, furthermore, α, δ are congruent to δ, β, and β, ε are congruent to ε, γ, and γ, ζ are congruent to ζ, α, then we can always find a molecule ω such that it is between α and ε, and it is between β and ζ and it is between γ and δ.

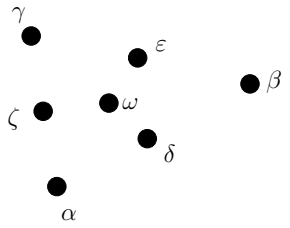

Figure 5. According to the semantics of the theory, Theorem 1′. (Fig. 4) is a statement about the molecules of the paper.

This hypothesis can be verified empirically by means of the scale-free ruler; and we will find that the hypothesis is true.

6 CONCLUDING REMARKS

As we can see, quantification over the mathematical entities can be eliminated from a physical theory. For example, in this "toy" physical theory we eliminated quantification over the points of $\mathbb{R}^2$ and quantification over the real numbers. This does not mean, however, that we have really purified our physical theory from Platonic objects: Although we eliminated *quantification* over mathematical entities, we did not eliminate the *mathematical structures* themselves. We still need the structure defined by axioms T1–T11. It thus seems unavoidable to draw the conclusion that we ought to have ontological commitment to formal systems as abstract entities; and this is sufficient for the structuralist version of Platonism (Shapiro 1997).

It must be noted that both the Quine–Putnam argument and Field's criticism are based on the tacit assumption that the terms and statements of mathematics have *meanings*, and the only question is the ontological status of the entities that mathematics refers to. According to the formalist philosophy of mathematics, however, this assumption is unacceptable, *ab ovo*.

As we have seen in our example, both the *platonistic* and the *nominalized* versions of the physical theory have the same structure: a *meaningless formal system* + a *partial semantics* pointing only to physical, moreover, observable things. From this point of view, it does not matter whether or not the formal system in question contains quantification over certain variables. Formal systems are obviously indispensable from both platonistic and nominalized physical theories, in spite of the fact that they are meaningless. The only question is: What is the ontological status of formal systems? And still, one can answer this question from a structuralist–Platonist position, drawing support from the Quine–Putnam indispensability argument. Or, one can consider an entirely different account

for formal systems, which completely intact from the indispensability argument (see for example (Szabó 2003) for a physicalist ontology of formal systems).*

REFERENCES

Burgess, John, 1983, Why I am not a nominalist. *Notre Dame J. Formal Logic* 24(1), 93-105.

Colyvan, Mark, 2004, Indispensability arguments in the philosophy of mathematics. *The Stanford Encyclopedia of Philosophy* (Fall 2004 Edition)

Craig, William, 1953, On axiomatizability within a system. *The Journal Of Symbolic Logic* 18, 30-32.

Field, Hartry H., *Science Without Numbers*. Basil Blackwell, Oxford.

Maddy, Penelope, 2005, Three forms of naturalism. in Shapiro, S. (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford University Press.

Quine, Willard V. O., 1961, On what there is. In *From a Logical Point of View*, 2nd ed. Harvard University Press.

Quine, Willard V. O., Things and their places in theories. In *Theories and Things*. The Belknap Press of Harvard University Press.

Shapiro, Stewart, 1997, *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press, Oxford.

Szabó, László E., 2003, Formal systems as physical objects: A physicalist account of mathematical truth, *International Studies in the Philosophy of Science* 17, 117-125.

Tarski, Alfred and Givant, Steven, 1999, Tarski's system of geometry. *Bulletin of Symbolic Logic* 5, 175-214.

*I would like to thank László E. Szabó for the conversations and insights which led up to this paper. I would also like to thank Réka Bence and Kristóf Szabó for their help.

Contributors

HAJNAL ANDRÉKA • Rényi Institute of Mathematics, Budapest
ANNA BROŻEK • Department of Logical Semiotics, Warsaw University
FERENC CSABA • Institute of Philosophy, Eötvös Loránd University, Budapest
GÁBOR FORRAI • Department of Philosophy, University of Miskolc
ZOLTÁN GENDLER SZABÓ • Department of Philosophy, Yale University, New Haven
ROBIN HIRSCH • Department of Computer Science, University College London
LÁSZLÓ KÁLMÁN • Department of Theoretical Linguistics, Eötvös Loránd University, Budapest / Hungarian Academy of Sciences
EDWARD KANTERIAN • Trinity College / Jesus College, Oxford
GYULA KLÍMA • Department of Philosophy, Fordham University, New York
ANDRÁS KORNAI • Media Research Center, Budapest University of Technology and Economics
JUDIT X. MADARÁSZ • Rényi Institute of Mathematics, Budapest
ANDRÁS MÁTÉ • Department of Logic, Eötvös Loránd University, Budapest
PÉTER MEKIS • Department of Logic, Eötvös Loránd University, Budapest
TAMÁS MIHÁLYDEÁK • Department of Computer Science, University of Debrecen
NENAD MISCEVIC • Department of Philosophy, Central European University, Budapest / Department of Philosophy, University of Maribor
ISTVÁN NÉMETI • Rényi Institute of Mathematics, Budapest
EDI PAVLOVIĆ • Faculty of Philosophy, University of Rijeka
JIŘÍ RACLAVSKÝ • Department of Philosophy, Masaryk University, Brno
MÁTÉ SZABÓ • Department of Philosophy and History of Science, Budapest University of Technology and Economics
ANNA SZABOLCSI • Department of Linguistics, New York University
GERGELY SZÉKELY • Rényi Institute of Mathematics, Budapest
MÁRTA UJVÁRI • Department of Philosophy, Corvinus University, Budapest
ZSÓFIA ZVOLENSZKY • Department of Logic, Eötvös Loránd University, Budapest