Table of Contents

Preface – The importance of being called Ernesto

Andrea Bianchi, Vittorio Morato, and Giuseppe Spolaore

Many years ago, in the Nineties, we three editors of this book attended a philosophy of language class taught by Ernesto Napoli. Ernesto's style was unconventional, at least compared with the standards we were used to. He stood there, asking questions and making objections to our attempts at an answer. He could be intimidating at times, and not all of our fellow students seemed to appreciate his approach. But for those of us who got those classes right they have remained memorable. Ernesto was taking us, and even more philosophy, *seriously*. This made a world of difference.

Looking back on it, we do not think there was a plan behind this attitude. As anyone who has come into contact with him knows, Ernesto just loves to discuss things and argue. To him, this is the most natural and straightforward way to understand something. It comes as no surprise, then, that he took it to be the most natural way to teach, too. In a sense, it is part of Ernesto's intellectual style to literally wrestle with the ideas he stumbles upon. Passive acceptance is not an option. *Ditto* for partial acceptance, or compromise. Even acceptance under hypothesis is suspicious – an altogether too-frivolous game to qualify as serious philosophizing. Ernesto is just too *earnest* for all this.

Seriousness and earnestness are moral qualities. When we qualify Ernesto's approach to philosophy as serious and earnest, we do not mean that he never indulges in jokes or tricks. The point is, rather, that Ernesto

has set himself certain very strict methodological guidelines, which he understands as more or less moral principles. Here are some examples. Avoid choices of terminology that invite confusion, no matter how unproblematic they are taken to be within your community. Ensure that you have important, non-pragmatic reasons for dealing with the issue you are presently dealing with. Make the value of your contribution as independent as possible of the state of the surrounding debate. Do not worry about philosophical fashions.

That Ernesto's name and the adjective "earnest" share the same origin (*eornost*, serious) strikes us as a fortunate coincidence. In fact, this is one of the reasons why we chose to give this book its title. Another reason is that names and the relation of being called are a recurrent focus of Ernesto's interest. These two reasons are connected in a curious and ironic way. One may be tempted to conceive the etymology of a name as providing the name's meaning, understood as descriptive content. But that is precisely a view that Ernesto, an early advocate of direct reference, would most firmly reject. Moreover, and more generally, Ernesto never tired of warning us of the fascination with etymology, going against a popular trend in continental philosophy.

In the Seventies, Ernesto was part of a generation of young scholars who were striving to escape the strictures of the Italian philosophical atmosphere, which was dominated by a mixture of historicism, idealism and existentialism. They experienced the study of formal logic and the philosophy of language, disciplines almost exotic at the time, as liberating.

Ernesto, who was studying in Torino, was initially driven toward Wittgenstein, thanks to the influence of his supervisor, Mario Trinchero. Certainly, something like a Wittgensteinian style shaped, and continues to shape, Ernesto's overall intellectual attitude, even though he did not write much about Wittgenstein, and would probably be ready to criticize almost every aspect of the latter's philosophy.

After a period in Oxford, where he had the opportunity to meet some of the big names on the British analytic scene – legends abound about these encounters – Ernesto moved to Pisa, to the Scuola Normale Superiore. It was in the special atmosphere of the Scuola Normale that Ernesto flourished as a philosopher, together with a close group of remarkable colleagues (among whom the late Paolo Casalegno, a most beloved friend

and interlocutor). Judging by the many anecdotes he is always happy to recall, for Ernesto the years spent in Pisa were some of the most rewarding in his whole intellectual career.

After Pisa, came the long period in Padova (1988 – 1998), where, due to a series of lucky coincidences, but also to a bit of planning (mainly by Paolo Leonardi), a considerable group of philosophers started working together in a distinctively analytic style. The center of their activities was a weekly seminar which, over the years, hosted some of the most important philosophers worldwide (among others, David Kaplan, Joseph Almog, Nino Cocchiarella, George Bealer, Saul Kripke, Keith Donnellan, Terence Parsons, and George Wilson). On some of these occasions, we saw hyper-sophisticated, hyper-professionalised (and, sometimes, hyper-convoluted) philosophical projects fall apart under Ernesto's insistence over some very simple, profound and foundational issues.

The last stage of Ernesto's academic career was in Urbino, a beautiful as well as inaccessible city in the very center of Italy. The inaccessibility of Urbino was quite upsetting for Ernesto (who usually commuted, by car, from Padova), as was the fact that he was forced to teach sociolinguistics there, a discipline he has never been very fond of. In order to make his frequent car trips to Urbino less boring, he offered to give a ride to any colleague living somewhere along his way. We still find the scene of Ernesto discussing philosophy at high speed (he often speaks of his driving style as "quite lively") very funny to imagine.

Nowadays, in his *buen retiro* on the gentle hills of Vicenza, Ernesto is more active than ever. Finally free of academic duties, he keeps working on a wide range of strictly interconnected topics, as he has always done.

This volume brings together sixteen original essays, written both by well-established and emerging philosophers and linguists: colleagues, (former) students, and friends of Ernesto's. We have asked them to relate their contribution to Ernesto's work. The request was happily received and, as far as we can judge, broadly complied with. The contributions, which we have chosen to group into three thematic parts (*Reference, proper names, and definite descriptions*, *Truth and logical consequence*, and *Syntactic structure and logical analysis*), reflect both the variety and the underlying unity of Ernesto's interests.

We are sure that Ernesto's reactions will not be late in coming and he will have a lot of things to say on each of the essays in the book. Long

telephone calls, long meetings in front of a cup of coffee (strictly without sugar), or long, chance discussions at street corners will await us, as well as many other contributors. We certainly do look forward to them.

Ernesto Napoli's publications

1976

"La miseria della linguistica", *Annali della Scuola Normale Superiore di Pisa*, III, 6, pp. 279–327.

1978

"Linguistica: scienza empirica?", *Studi di grammatica italiana*, 7, pp. 57–79.

"A note on Quine's and Carnap's analyticity", *Annali della Scuola Normale Superiore di Pisa*, III, 8, pp. 249–255.

"A note on Quine's criterion of ontic commitment", *Annali della Scuola Normale Superiore di Pisa*, III, 8, pp. 685–693.

"Review of R. Harré, *Introduzione alla logica delle scienze*", *Annali della Scuola Normale Superiore di Pisa*, III, 8, pp. 1989–1996.

1980

"Codice e lingua, alcune osservazioni occasionali", *Studi di grammatica italiana*, 9, pp. 369–383.

1981

"All Kant's sons", *Studi di grammatica italiana*, 10, pp. 419–438.

"Review of W. V. Quine, *Logica e grammatica*", *Teoria*, 1, pp. 198–199.

"Review of D. Marconi (ed.)*La formalizzazione della dialettica* (with Enrico Moriconi), *Revue Internationale de Philosophie*, 136–137, pp. 385–392."

1982

"La contraddizione di Wittgenstein", *Teoria*, 2, pp. 57–84.

"Su un impiego dell'uso. Ovvero la ricognizione dummettiana del significato" (with Paolo Casalegno), *Annali della Scuola Normale Superiore di Pisa*, III, 12, pp. 749–770; reprinted in P. Casalegno, *Verità e significato. Scritti di filosofia del linguaggio*, Carocci, Roma, 2011, pp. 133–150.

1983

"Review of H. Putnam, *Reason, Truth and History*", *Lingua e stile*, 18, pp. 575–580.

"Review of W. V. Quine, *Theories and Things*", *Teoria*, 3, pp. 176–178.

1984

"Hilary Putnam: dal realismo metafisico al realismo critico", *Lingua e stile*, 19, pp. 451–474.

"On the vagaries of underdetermination", *Annali della Scuola Normale Superiore di Pisa*, III, 14, pp. 1441–1464.

1985

"Is vagueness a logical enigma?", *Erkenntnis*, 23, pp. 115–121.

"Priest's paradox", *Logique et Analyse*, 28, pp. 403–407.

1987

"Dimostrazione e significato" (with Enrico Moriconi), *Lingua e stile*, 22, pp. 153–178.

"Review of A. Bonomi, *Le immagini dei nomi*", *Teoria*, 7, pp. 178–181.

"Review of D. Marconi, *L'eredità di Wittgenstein*", *Teoria*, 7, pp. 183–185.

"Review of M. Bonfantini, *La semiosi e l'abduzione*", *L'Indice*, 10, p. 41.

1988

"Dummett's transcendence" (with Enrico Moriconi), *Philosophia*, 18, pp. 371–383.

1989

"La logica dei modelli mentali", *Teoria*, 9, pp. 145–160.

1990

"Probabilistic and dual aspect semantics", *Annali della Scuola Normale Superiore di Pisa*, III, 20, pp. 675–693.

1992

"Riferimento diretto", in M. Santambrogio (ed.), *Introduzione alla filosofia analitica del linguaggio*, Laterza, Roma-Bari, pp. 385–429.

1995

"(Direct) reference", *Journal of Philosohical Logic*, 24, pp. 321–339.

"On naming" (with Paolo Leonardi), in P. Leonardi and M. Santambrogio (eds.), *On Quine: New Essays*, Cambridge University Press, Cambridge, pp. 251–266.

1996

"Review of R Larson and G. Segal, *Knowldege of Meaning: An Introduction to Semantic Theory*", *Lingua e stile*, 31, pp. 600–603.

1997

"Names, indexicals, and identity statements", in M. Anduschus, A. Newen and W. Künne (eds.), *Direct Reference, Indexicality and Propositional Attitudes*, CSLI Publications, Stanford, pp. 185–211.

1998

"From discoursive anaphora to referentiality", in S. Gensini and D. Russo (eds.), *Annuario 1998 della Società di Filosofia del Linguaggio*, Roma, pp. 63–74.

"Reports", *Lingua e stile*, 33, pp. 501–514.

1999

"Come sinceramente dire quello che non si pensa", *Rivista di estetica*, 39, pp. 85–96.

2000

"Finti nomi", in G. Usberti (ed.), *Modi dell'oggettività*, Bompiani, Milano, pp. 197–220.

"Inferenze non monotòne", in P. Cherubini, P. Giaretta and A. Mazzocco (eds.), *Ragionamento: psicologia e logica*, Giunti, Firenze, pp. 69–73.

"Proper names, descriptions and quantifier phrases" (with Mario D'Angelo), in D. Marconi (ed.), *Knowledge and Meaning: Topics in Analytic Philosophy*, Mercurio, Vercelli, pp. 195–233.

2003

"Significato e ontologia", in C. Bianchi and A. Bottani (eds.), *Significato e ontologia*, Franco Angeli, Milano, pp. 145–160.

2004

"Poche parole" (with Andrea Bianchi), in M. Carrara and P. Giaretta (eds.), *Filosofia e logica*, Rubbettino, Soveria Mannelli, pp. 175–225.

2006

"Negation", *Grazer Philosophische Studien*, 72, pp. 233–252.

2008

"Citazione", in P. Cherubini, R. Rumiati and A. Tasso (eds.), *Rappresentazioni e congetture*, CLEUP, Padova, pp. 167–176.

2010

"Amarcord", in M. Carrara and V. Morato (eds.), *Verità. Annuario e Bollettino della Società Italiana di Filosofia Analitica*, Mimesis, Milano, pp. 375–378.

"Realtà, verità e linguaggio", in M. Carrara and V. Morato (eds.), *Verità. Annuario e Bollettino della Società Italiana di Filosofia Analitica*, Mimesis, Milano, pp. 295–302.

2011

"Reference fixing and the stiffness of reference. Or three (would be) puzzles concerning names", *Rivista di estetica*, 47, pp. 179–195.

2015

"Names as predicates?", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 211–224.

Part I
Reference, proper names, and definite descriptions

How Kripke Carnaps Mill*

Roberta Ballarin

> [I]t becomes us not to triumph over the great intellects [...]
> because we are now able to preserve ourselves from many errors
> into which they, perhaps inevitably, fell. The fire-teazer of a
> modern steam-engine produces by his exertions far greater effects
> than Milo of Crotona could, but he is not therefore a stronger man.
>
> J.S. Mill (1843: 79)[1]

In this paper I argue for an interpretation of *Naming and Necessity* (from now on NN) as a deeply divided text. On the one hand, NN breaks with the logically-centered philosophical tradition of the first half of the last century, is perhaps the single major contribution to the contemporary renaissance of metaphysics, and is also one of the initiators of the 1970's anti-Fregean movement in semantics. On the other hand, I will argue,

[1] All page references to Mill's 1843 *System* are from the 1973 edition.

NN is still very much influenced by Kripke's previous technical work in modal logic, and thus indirectly indebted to the great philosopher-logicians that paved its way. In this paper, I focus on Kripke's semantics, leaving metaphysics aside.[2] Also, I do not discuss the ways in which NN successfully delivers what it promises, namely, *prima facie* compelling intuition-based arguments against descriptivism for proper names – that part of NN is well known. I just aim to uncover the less recognized side of NN. In particular, I argue that somewhat surprisingly, NN is also influenced by Carnap's work in formal semantics. I say "surprisingly" because even the title of Kripke's monograph seems designed to suggest an opposition to Carnap's *Meaning and Necessity* (from now on MN) – an opposition on both the semantic and the metaphysical fronts. More-over, Carnap's key doctrines seem to be the exact opposite of Kripke's. Setting metaphysics aside, which Carnap tout court rejects and Kripke wholeheartedly embraces, even in semantics MN and NN appear to be irreconcilable. Carnap's MN constructs a systematic, formal counterpart of Frege's semantics. Kripke's semantic doctrines instead are meant to be neither systematic nor formal. Kripke aims to provide an informal but compelling picture of how names work, based on our intuitions as speakers of English. Moreover, Kripke rejects what he calls "the Frege-Russell view" of names as short for definite descriptions, and embraces an anti-Fregean, Millian view of proper names, extending it also to mass nouns for natural substances and to count nouns for natural kinds.

In what follows, I discuss what I take to be two 'anti-Kripkean' aspects of Kripke's semantics. First, I argue that Kripke's well-known thesis that names are rigid designators, a celebrated cornerstone of Kripke's anti-descriptivism, is a remnant of Kripke's earlier work in formal semantics. Next, I argue that Kripke's handling of Mill's view of common nouns obscures an important semantic distinction, as well as a metaphysical point. Kripke's mischaracterization of Mill's view is connected to Kripke's blind spot for the distinction between Frege's and Russell's versions of descriptivism. I also argue that these blind spots are the result of an unexpected Carnapian influence on Kripke's semantics.

---

[2] I argue for the metaphysical point in Ballarin 2004.

Rigid designators

Let me start with the case of rigid designators, where it is relatively easy to locate the contrast between Kripke's 'new' philosophical intuitions and his 'old' logical propensities. Since Kaplan made the point in "Demonstratives" (1989), we have all learned that the category of rigid designators is semantically promiscuous, in the sense that it mixes together semantically heterogeneous expressions, namely (i) proper names, which are directly referential and whose rigidity is a consequence of this semantic characteristic, and (ii) rigid descriptions like "the successor of eight" that are in fact descriptive and so semantically on a par with non-rigid descriptions like "the number of planets". In the case of descriptions, the rigidity is a semantically accidental byproduct of merely metaphysical, i.e. semantically irrelevant, facts, like the necessity that eight be succeeded by nine.[3]

It is also well known that Kripke unwaveringly endorses a Russellian theory of definite descriptions (recall his entrenched hostility to Donnellan's (1966) referential cases (Kripke 1977)), and firmly rejects a Fregean or Russellian descriptive theory of proper names by appealing to the intuitions of the speakers in the marketplace. So, why does he promulgate the notion of rigidity which mixes together names and descriptions and so much obscures their critical semantic distinction? A hint to the answer comes from Kripke's own pen:

> Some of the worst misinterpretations of rigidity would have had much less currency if the relevant philosophical discussions had been conducted in the context of a rigorous presentation in terms of 'possible world semantics'. (Kripke 1980: 15, fn. 16).

Indeed, in the formal possible world semantics the notion of rigidity emerges naturally. The individual variables and constants of a first-order language are assigned values at models. No additional interpretive step other than this model-relative valuation is required. Hence, the assignments are arbitrary and stipulative. No pre-semantic acts of naming (Kripke's 'baptisms'), as for vernacular proper names, nor semantic interpretations, as for definite descriptions, mediate between the formal

---

[3] On the shortcomings of rigidity see Almog 1986 and Marti 2003. Marti also argues that directly referential expressions need not be rigid.

individual terms – be they variables or constants – and their values. Hence, no distinctions in the mechanisms through which they secure their referents divide these terms up. The formal semantics of the primitive singular terms of first-order logic is unequivocal.[4] The singular terms of English instead are, according to Kripke, divided into (directly referential) non-descriptive names on the one hand, and definite descriptions on the other.[5]

When the language and proof-theory of non-modal first-order logic are extended to include the modal calculus, the formal semantics is correspondingly extended to possible world semantics for quantified modal logic. Now the models of first-order logic – or technical developments thereof, like points of evaluation accompanied by an interpretation function – become the worlds of the models of quantified modal logic.[6] The individual constants and variables of first-order logic are given model-relative one-step unmediated valuations, i.e., they are arbitrarily assigned values at models, and a term need not be assigned the same value across all models. Now that the models have become worlds in a modal structure, we can then call "rigid" the terms that are assigned a cross-world (that is, cross-first-order-model) *fixed* value in the structure (so rigidity is assignment relative). Formally, this stability of assignment is all there is to rigidity. One can stipulate that variables be assigned the same individual at all worlds, i.e., that they all be rigid (as in Kripke 1963), but this is a merely technical stipulation. Clearly, there is no distinction between, so to speak, two sorts of fixed-value variables: those that are rigid for semantic reasons and those that are rigid for extra-semantic reasons. This distinction makes no sense in a formal setting (not even for complex terms like "$(\imath x)Fx$" given that the extension of "F" at different worlds is just stipulated). Similarly, there is also no 'deep' semantic distinction, over and above a technical stipulation, between 'fixed-value' and 'non-fixed-value' variables, under an assignment.

---

[4] One can add non-primitive individual terms, for example by means of a description operator "$(\imath x)\ldots x \ldots$". However, the formal semantics of complex individual expressions like "$(\imath x)Fx$" is designed to reproduce that of the corresponding English terms, thus it does not throw any light on it.

[5] What is the difference between a stipulative assignment and a baptism? Baptisms are historical, communal events; assignments of values are logical functions.

[6] See Kaplan 1986, Section C, on models and worlds.

There is no 'hidden', ulterior semantic distinction between the terms whose valuation is rigid, i.e., stable across worlds, and those (if any) that are assigned different values in different worlds. The instability versus stability of assignment is *all there is* to the formal-semantic notion of rigidity.

It is then only natural that, as Kripke points out, no (formal) semantic obscurity is engendered by the rigidity of the assignments, and no confusion or misunderstanding is forthcoming. But the clarity is due exactly to the fact that in a formal setting there is nothing behind rigidity. This lack of 'hidden' reasons for the rigidity of the singular terms of a formal language makes the formal semantics clear, but it also makes it inadequate to explain the semantic behavior of English singular terms. If the formal mechanism that accounts for the rigidity of variables in possible world semantics cannot throw any light on the semantics of English proper names, we are still owed an explanation for the modal behavior of English terms, and in particular for their rigidity. Neither proper names nor definite descriptions are assigned values by a logical stipulation. Thus, their rigidity, or lack thereof, throws no light on how they acquire their referents and on the interpretation of the sentences in which they occur. That is why Kripke proceeds to present a theory of names in terms of 'baptisms' that explains exactly what the rigidity cannot explain: how the names of a natural language like English function, i.e., how they acquire their referents. We find then two alternative semantic pictures in NN. One model-theoretically driven in which the notion of rigidity naturally arises but all primitive singular terms are semantically on a par; the other grounded in intuitions about the truth values of interpreted vernacular sentences, intuitions that support radically different interpretations for names and descriptions.[7]

Kripke himself highlights the formal nature of the notion of rigidity, and the distinction between formal symbols and English names:

> In speaking of rigid designators, we are speaking of a possibility that certainly exists in a formal modal language. Logically, we as yet are committed to no thesis about the status of what we ordinarily call 'names' in natural language. (1980: 3–4)

---

[7] Be they intuitions about the actual truth value of a sentence like "Nixon might have lost" or about the truth value a non-modal sentence like "Nixon lost" would have in different circumstances.

Kripke uses the rigidity of names to argue that they are not descriptive and "that true identity statements between what we call 'names' in actual language are necessary" (1980: 4). Clearly, Kripke's emphasis on the rigidity of names is not part of a comprehensive Montagovian project of treating English as a formal language, that is, of providing a mathematical model-theoretic semantics for English.[8] Kripke is just exploiting a similarity between variables (the technical paradigm of singular terms) and English names to further his non-formal semantic agenda. Yet, in what follows I argue that Kripke's engagement with formal semantics – "Of course the work grew out of earlier formal work in the model theory of formal logic" (1980: 3) – is not always beneficial. Rigidity may very well help clarify some features of the natural interpretation of English names, but in some other cases the model-theoretic semantics hinders Kripke's philosophical project.

Frege and Mill on common names

Kripke endorses a Millian view of proper names (1980: 26–27), which he opposes to the Frege-Russell view. According to Mill, proper names are denotative and non-connotative. They "signify a subject only" (1843: 30) independently of its satisfying any attribute.[9] On the other hand, according to Mill general names are connotative. Mill's general names are those expressions that are "capable of being truly affirmed, in the same sense, of each of an indefinite number of things" (1843: 28). For Mill, general names, i.e. common nouns/names, e.g. "man", "stone" and "color", are semantically akin to adjectives like "white" and "virtuous". "White" applies to all the things that have whiteness. "Whiteness" is the abstract, purely denotative (I would like to say, directly referential) name of the attribute that the corresponding adjective connotes. A general name, like "man", connotes a set of attributes and applies to those things that

---

[8] As in Montague 1970.

[9] According to Marti (2003), the ultimate significance of Mill's thesis is that the referring relation (between a name and its denotation) is semantically speaking basic. I think she is right about this. In what follows, I argue that according to Mill the relation between a term and its connotation is also semantically basic.

satisfy all those attributes. Kripke claims that Frege and Russell agree with Mill about common names, but disagree on proper names: "Frege and Russell [seem] to hold that Mill was wrong about singular names, but right about general names" (1980: 127). He instead agrees with Mill about proper names, but disagrees about (at least some) common names. Kripke concedes that some general terms, and he mentions adjectives like "yellow" and "fat", perhaps express properties (attributes) in virtue of which they apply to particular things, but claims that common names for natural kinds like "cow" and "tiger" and mass nouns for natural substances like "gold" and "water" are really non-connotative proper names of species and substances.

In this paper I do not discuss Russell's view. Perhaps Kripke is right about Russell, but is he right about Frege? Does Frege agree with Mill on common names? We can read Kripke's claim weakly, as stating that Frege disagrees with nothing that Mill positively claims about common names. I do not want to question this weaker claim, though it too is not literally true. However, on a stronger, rather natural, understanding of Kripke's remark, he seems to be saying that Frege's theory just is the same as Mill's, i.e., that Frege would not regard Mill's semantics for general names as incomplete. This last claim, I argue, is wrong.

The key feature of Mill's view of predicates, i.e., predicative expressions in general – including common nouns and adjectives – is that they denote an indefinite number of individuals in virtue of these individuals satisfying some attributes, i.e., bearing some properties.[10] Frege's semantics for predicates, however, is more complex. In my view, Frege's semantics for predicates includes two distinct key theses. First, predicates, like names, express a sense and have a referent. The referent of a predicate, as for a name, is determined by the predicate's sense. This two tiered-character is probably the fundamental and surely the most advertised general feature of Frege's semantics.[11] But Frege also holds, about predicates in particular, that they stand for (refer to) what he

---

[10] The individuals need not be concrete. For example, "color" is truly predicated of the colors (white, blue, red and so on) which are for Mill abstract attributes.

[11] "On Sense and Reference" (1892a), titled "On *Sinn* and *Bedeutung*" in Beaney 1997, is the paper of Frege most obviously concerned with semantic matters. Its focus is exactly on the need of senses over and above referents.

calls "concepts". Concepts are essentially functional entities.[12] Nonetheless, Frege's concepts are part of the furniture of what is standardly regarded as the real world, inasmuch as they belong to the same ontological realm as objects.[13] I think it is safe to think of Frege's concepts as ontologically akin to Mill's attributes, insofar as Mill's attributes are real features of the individuals that possess them – as opposed to ideas or mental concepts which are not to be found *in* the objects they represent.

We may say that for Frege both names and predicates are descriptive – if senses are – given that their referents are mediated by the senses. Thus, the semantic relation that ties a name or predicate to its referent is not semantically basic, it must be analyzed into two distinct components. This common feature should not obscure the *semantic* diversity of names and predicates. What I want to suggest is that for Frege the distinction between objects and concepts is not only ontological: it plays also a key semantic role. Thanks to their saturated or unsaturated characters, objects and concepts are able to play two distinct semantic roles. Frege's names are fundamentally referential expressions, though not direct, in the sense that they stand for a saturated entity. In my interpretation, Frege's requirement that the referent of a name be saturated is Frege's counterpart of Mill's thesis that proper names are *just* denotative expressions. In this sense, Frege's names, like Mill's, are essentially referential, non-predicative expressions. Predicates instead stand for concepts, thus are not referential in this specific sense. Predicates are fundamentally predicative expressions, thus their referents are essentially functional, unsaturated entities. This thesis corresponds to Mill's claim that general names are connotative. Both Frege and Mill think that the predicative role of predicates consists in their being true of an

---

[12] See Frege 1891 and Frege 1892.

[13] Andrea Bianchi has suggested that concepts belong with senses to the third realm of Frege's "Thought" (1918). I disagree. I believe that for Frege concepts are part of what he calls the external reality. In my reading of "Thought", Frege introduces the third realm as the realm of judgments, in contrast to ideas, in order to emphasize the non-psychological nature of logic. Thus, even granting the abstract and non-subjective nature of concepts, concepts do not belong to the third realm insofar as they are not parts of judgments. See Ricketts 1986 on the realm of thought as the realm of the *objective*, not as an ontological category, leave alone an ontological niche for all non-subjective abstracta. Moreover, I am also not sure that Frege's concepts are as ontologically independent of objects as Bianchi takes them to be.

indefinite number of things in virtue of the fact that these things satisfy the features semantically associated with the predicate.

The thesis I am putting forward is that the referential vs. predicative distinction is orthogonal to the direct vs. indirect distinction. Kripke thinks that Frege and Mill disagree on names and agree on predicates (common names) because he thinks that Mill's connotations correspond to Fregean senses (or vice versa). In my view instead they disagree on both names and predicates about directness versus indirectness, but they agree on both names and predicates as far as referentiality and predicativity are concerned. We may say that for Frege names indirectly refer, and predicates indirectly predicate.[14] For Mill instead names directly refer (denote, in his terminology) and predicates directly predicate (connote, in his terminology). This is the case because Mill's connotations play the semantic role of Frege's concepts, not of the senses.[15]

Once we understand that for Frege the saturated-unsaturated distinction plays a semantic role, we are in a position to make sense of Frege's cryptic claim about the concept horse. Frege insists that the concept horse is not a concept because the expression "the concept horse" is a name and can therefore only signify an object. In the same vein, he writes,

> [a] similar thing happens when we say as regards the sentence 'This rose is red': the grammatical predicate 'is red' belongs to the subject 'this rose'. Here the words 'the grammatical predicate "is red"' are not a grammatical predicate but a subject. By the very act of explicitly calling it a predicate, we deprive it of this property. (1892: 185)

The very last sentence in this passage clearly indicates that the grammatical category of an expression determines the nature of its referent, not vice versa. My proposal is to regard this perplexing thesis as fundamentally semantical, not ontological. The grammatical category of an expression does determine the kind of referent it can have, but it does

---

[14] Fara (2015: 62) claims that the referential vs. predicative distinction is orthogonal to the descriptive vs. directly referential distinction. I fully agree with her on this. Fara's main thesis is that names are predicates, thus she disagrees with both Mill and Frege. Notice that the thesis that the mediation works by description is additional to the claim that there is a mediation.

[15] Of course, Frege and Mill disagree on descriptions. For Frege they are name-like referential expressions, for Mill they are predicate-like predicative expressions.

not make its referent be the kind of thing that it is. I believe, though I cannot defend the claim here, that Fregean concepts and objects have a genuine ontological independence, namely that they do not supervene on grammar. In other words, the semantic roles they play do not exhaust their natures. At the very least, my interpretation of the passage above is compatible with this view of objects and concepts. In my reading, the idea that an ontological distinction must reflect the grammatical partition is part and parcel of Frege's semantics. I read the passage above as claiming not so much that the nature of an object (concept) is exhausted by its being the referent of a name (predicate), but instead that to be a name (a referring expression) is to stand for an object and to be a predicate (a predicative expression) is to stand for a concept. That is why I say that it is part of Frege's *semantic* theory that predicates refer to unsaturated entities (semantics, not metaphysics, depends on syntax). Similarly, an expression cannot be a name unless it stands for a saturated, i.e., non-predicative, entity. "The concept horse" is a grammatical name, thus its referent cannot be predicative. It is easier, though by no means uncontroversial, to concede that only properties can be the referents of predicates. But it does seem wrong that we should not be able to name a property, to make it the subject of discourse. Nonetheless, I think we can make sense of Frege's point, if we embrace my proposal. Frege's concepts play a semantic role, insofar as the semantic function of predication is cashed in through them. Similarly for objects and naming. But if to name is to stand for an object, it does follow that we cannot name concepts. So, both Mill and Frege believe that a semantic distinction must correspond to the grammatical distinction between names and predicates. Mill cashes in this semantic distinction in terms of two distinct basic semantic relations. Names denote, predicates connote. Frege instead has only one semantic relation between words and their referents. Thus, for him to name is to be semantically related to a saturated entity and to predicate is to be semantically related to an unsaturated one. Notice that in this perspective Frege assimilates descriptions to names, not names to descriptions. Both are referential, non-predicative expressions.

How can we then make sense of Frege's remark that "[b]y the very act of explicitly calling it a predicate, we deprive it of this property"? In my view, this passage suggests not so much that naming a concept turns *it* into an object, but rather that concepts have a double semantic role.

A concept, being unsaturated, literally embodies predication. But a concept has also the potential to provide an object when it is, so to speak, in a non-predicative position. It does so by automatically leading to its extension. It is in this sense that, contrary to the standard interpretation, I regard the relation of a Fregean concept to its extension as semantically charged. What I mean is the following. If you want to speak of the concept horse, and say "the concept horse", the moment you manage to refer to that concept, the concept cannot play its semantic predicative role given that "the concept horse" is a grammatical subject, not a predicate. "The concept horse" cannot refer to a concept *per se* because concepts are by their nature predicative (unsaturated) and only a predicate can stand for a predicative entity. Luckily concepts are not only unsaturated, they also have extensions: this gives them an additional semantic potential. (Really, to be unsaturated is to have an extension.) When subjects of discourse, concepts are semantically programmed to 'transform' into objects. Which objects? Their extensions of course! Thus, in subject position concepts inevitably lead to their extensions. Concepts have a double semantic role: in predicate position they are predicative, in subject position (where they cannot stand) they denote their extensions.[16] An unexpected bonus of this interpretation is that now Frege need not be wrong about descriptions (he does not really think that they are name-like simply referential expressions), he just did not bother to state his view completely. Descriptions, unlike names, have a predica-

---

[16] You may have noticed that this is Russell's 1905 Grey's Elegy problem. In 1903 Russell views denoting concepts as essentially semantically charged entities. Like Frege's concepts, Russell's denoting concepts hypostatize a semantic function (denoting in Russell's case, predicating in Frege's). This is no place to expand on this, but I believe that in 1903 Russell distinguished three basic semantic roles: referring, predicating, and denoting, and took denoting concepts to embody the denoting semantic function, while for some reason he did not think he needed an embodiment for reference and predication. When he became aware of the Grey's Elegy problem – which I take to be Frege's 'the concept horse' problem – Russell renounced denoting concepts. Kaplan (see for example 2005: 942) emphasizes that Russell's 1905 notion of denotation is non-semantical. But this is standardly (though not by Kaplan) taken to mean that in 1905 Russell has relinquished the sense-like theory he held in 1903. Russell himself wrongly thought that Frege's senses played the role of his denoting concepts. Instead, if I am right, in 1905 Russell renounces the hypostatization of the semantic function of denoting which in Frege's system is played by Frege's concepts if by anything at all.

tive kernel, but grammar places them in referential position. Thus, their predicative function is turned off, which is to say that the concept which the predicative part of the description stands for cannot be the referent of the description. Nonetheless, that concept still plays a semantic role insofar as *its* extension is the referent of the description. In subject position concepts seamlessly switch to their denotation (the sleight is so fast that Frege has no time to notice it!). I call this the concept-horse Fregean theory of descriptions. Basically, my proposal is that in subject position Frege's concepts be understood as Russellian denoting concepts, thus preserving for them a semantic role (denoting) despite their inability to be the referents of descriptions. This Russellian interpretation of Frege is inspired by the affinity between Frege's remarks on the concept horse and Russell's 1905 argument against denoting concepts.[17]

Setting aside the controversial thesis I am controversially attributing to Frege,[18] that concepts hypostatize predication and perhaps even denotation, the basic kernel of Frege's second semantic tenet is that it is the grammatical function of an expression, and not the sense that it expresses, that determines whether it is a name or a predicate. This last thesis at least stands a better chance of being uncontroversial.

---

[17] Can this really be Frege's ultimate view of descriptions? Honestly, I doubt it. It is not clear that the *Bedeutung* of a description is the extension of the corresponding concept. "The queen of England in 2015" *bedeutet* Elizabeth, not the Fregean extension of the concept *queen of England in 2015*, which according to Frege is a logical object; and "the king of England in 2015" refers to nothing according to Frege, not to an empty extension. In fact, Frege never claims that "the concept horse" refers to the extension of the concept horse, only that it refers to an object. Mine of course is a very Russellian interpretation of Frege. Yet, despite the fact that Frege's ultimate view of descriptions is that they are, like names, referential expressions, I do think that Frege envisioned denoting as a third semantic function, distinct from both referring and predicating.

[18] Ricketts (1986: 84) claims that for Frege ontological categories are supervenient on logical ones. Moreover, Ricketts takes this to mean that "[t]he dichotomy between objects and concepts comes into its own not as regards proper names and first-level predicates but rather in introducing and distinguishing first-level generality (over objects) from second-level generality (over concepts)" (1986: 87). My point is instead that the dichotomy between objects and concepts *does* come into its own as regards proper names and first-level predicates. Moreover, I do not think that ontological categories are supervenient on grammatical ones, only that the ontological saturated-unsaturated distinction plays a semantic role too. Thanks to Ori Simchen who referred me to Ricketts 1986.

What about Mill? According to Mill, proper names like "John" and "Mary" denote a singular individual. The mark of predicates instead is that they denote an indefinite number of objects. Predicates denote, literally name, the individuals of which they are true:

> As a proper name is said to be the name of the one individual which it is predicated of, so [...] a connotative name ought to be considered a name of all the various individuals which it is predicable of or in other words *denote*, and not of what it connotes. (1843: 36)

I have already pointed out that ontologically Mill's attributes are similar to Frege's concepts: they are features of things in the real world. In what follows I defend the claim that Mill's notion of connotation plays the semantic role of Frege's concepts, not of Frege's senses. Mill's claim that predicates have both a connotation and a denotation corresponds to Frege's semantic tenet that the referent of a predicate is a predicative entity, not to the view that predicates express senses.

According to Mill, though both singular and general names are said to name individuals, a proper name and a predicate relate to their respective bearers in a different way. The name "Secretariat" relates to that famous horse in a way that Mill does not specify much, except to point out that it is not mediated. The predicate "horse" instead is a name of that same animal in virtue of its bearing the attributes that the word connotes. The relation that holds between "Secretariat" and Secretariat is different from the (composite) relation that holds between "horse" and Secretariat. When Mill says that both "Socrates" and "man" denote/name/are predicable of Socrates, he his concealing this important distinction.

Mill seems to be struggling with the following problem: he has only two words when he needs three. Let us follow Mill in calling "denotation" whatever semantic relation holds between a proper name and its bearer, regardless of whether it is mediated by a Fregean sense or unmediated as Mill takes it to be. That exact same relation holds, according to Mill, between the name of an attribute, like "whiteness", and the attribute it stands for, whiteness in this case. Now, we call whiteness "whiteness", but we do not call whiteness "white" (we call white things "white"). Thus, the semantic relation that holds between the term "white" and whiteness is not the same semantic relation that holds between the name "whiteness" and whiteness. Thus, Mill picks the term "connote" instead

of "denote" to speak of the way in which "white" relates to whiteness. It is the fact that it is connoted, rather than denoted, that makes an attribute apt for predication. Mill is explicit about this in a long footnote (1843: 40–41) where he explains why he chooses to say that a general name connotes an attributes and denotes a class, instead of saying that it denotes the attribute and connotes the class as previous writers, including his father James Mill, did. Mill needs to mark out the grammatical distinction between referring and predicating, and he uses "connote" for this purpose. For Frege, predication consists in the unsaturated character of the referent. For Mill, the metaphysics of attributes is irrelevant to semantics. Whether denoted or connoted an attribute is the kind of entity that has an extension, but this metaphysical feature of attributes plays no semantic role in Mill's system. The difference between "whiteness" and "white" consists in their bearing two distinct basic semantic relations to the same entity (not in standing in the same relation to two different entities).

Unfortunately, Mill employs only two words: "connotation" and "denotation". If he uses them to distinguish the "white"-whiteness relation from the "whiteness"-whiteness relation, he is left with no verbal resources to distinguish the relation that holds between "whiteness"-whiteness and "Secretariat"-Secretariat, on the one hand, and the relation that holds between "white"-Socrates and "brown"-Secretariat, on the other hand. He calls all of them denotation. Ideally, he should have used a new term for the composite relation that holds between a predicate and the objects it is true of. Despite the terminological blunder, Mill is absolutely clear on the fact that predicates do not *just* name individuals, nor classes of individuals (1843: 90–97).

When connoted, attributes are predicated. They also determine the class of things of which the predicate is true. This is exactly the role played by Frege's concepts. (The sense of a predicate instead determines the concept, not its extension). So, Mill's view of predicates is not that they express some sort of Fregean senses in virtue of which they relate to attributes. Rather, they stand in a *direct* semantic relation to attributes, a relation he calls "connotation". Connotation is not denotation, but like denotation it directly relates a term to an entity. Both objects and attributes can be denoted (named), but only attributes can be connoted (predicated).

Should we say that for Mill, as for Frege, attributes in and of themselves play the semantic role of predication? I do not think so. If they were intrinsically predicative entities they would not be able to withstand denotation. Once denoted, they would push forward to their extensions. If I am right about this, the relation that holds between a Millian attribute and the class of things it applies to is not semantical. Attributes may very well be unsaturated, property-like entities, from a metaphysical point of view, and as such they may be apt to be predicated of individuals, but the semantic role of predication is played out by the connotation relation.

In Mill's semantic system nothing plays the role of senses. The main semantic value of an expression is always unmediated. A predicate, adjective or general name, connotes an attribute or a cluster of attributes as directly as a singular name denotes its denotation. An attribute is employed predicatively, so to speak, when it is connoted rather than denoted.

So what is Kripke disagreeing with when he disagrees with Frege and Mill on predicates? What does he mean when he says that "[c]ertainly 'cow' and 'tiger' are *not* short for the conjunction of properties a dictionary would take to define them"? (1980: 128) The problem as I see it is that Kripke's term "property" plays too many roles. It covers Fregean senses, as well as Fregean concepts and Millian attributes. First, talk of properties hides the distinction between indirect reference and predication, two distinct semantic functions. When he says that some nouns (like "gold" and "tiger") do not express properties, Kripke is not only rejecting Frege's senses, and with them the entire conception that semantics should track cognition (senses after all form thoughts), but also the Frege-Mill idea that these are really predicative expressions. Kripke takes "gold" and "tiger" to be singular names, not general, predicative terms. These are two distinct theses.

Consider the sentence "Secretariat is a horse". According to Frege's first semantic tenet, this sentence expresses a thought made up of the senses of its parts. The sense of "horse" may very well be descriptive, but insofar as its referent is a concept/attribute/property, the description is a description of the property itself, not of the objects that have that property. This is not to deny that a description of a property *may* in some cases indirectly describe the things that bear the property, but the

sheer theory that predicates express senses does not require this. Surely, Kripke disagrees with the idea that "horse" expresses a sense, but so does Mill.

What about Frege's second semantic tenet? Is "horse" a predicative term? On this too Kripke disagrees. This means that for Mill, who agrees with Frege that common names are predicates, the truth conditions of "Secretariat is a horse" are something like this:

(MTC) "Secretariat is a horse" is true just in case the referent of "Secretariat" bears the attributes connoted by "horse".[19]

For Kripke instead:

(KTC) "Secretariat is a horse" is true just in case the referent of "Secretariat" is a member of the referent of "horse".

This is still a semantic disagreement. Sure, if "horse" is the name of a species, the species must exist (or have existed) to be its referent. But one may have merely semantic reasons for preferring (MTC), for example parity of semantic analysis with those cases where the common name is not the name of a natural kind (like "bachelor" and "knife").[20] As Kripke himself points out, one may even take *being a cow* trivially as a property (corresponding to the kind) and thus (I add) employ it rather than the kind in semantics, for whatever semantic reasons she has.

Yet, when Kripke rejects Mill's definitional properties, he is also making a metaphysical point. When he says that an animal may not look like a tiger and still be a tiger, or look like a tiger and not be one (1980: 137), Kripke is claiming not only that the word "tiger" is not defined as short for a list of properties, but also that in real nature to be a tiger is to be a member of a certain biological species and not to bear some properties.

---

[19] Of course, the referent of "Secretariat" bears the attributes connoted by "horse" if and only if it is (one of the things) denoted by "horse". From a semantical point of view however bearing the attributes is basic, that is, Secretariat is denoted by "horse" because it bears the relevant attributes. In other words, "horse" denotes what it denotes in virtue of connoting what it connotes, not vice versa.

[20] I am not saying that Kripke could not instead try to extend his semantic analysis of natural kind terms to other common names. In any case, a uniform semantics for expressions that belong to the same grammatical category (like "horse" and "knife") seems desirable. Grammar, not metaphysics, should inform semantics.

Thus, Kripke is also rejecting Mill's view on the metaphysics of kinds. In fact, Mill holds that kinds are logical constructions out of attributes. According to Mill, kinds have a real existence in nature (because attributes do). But when he talks of the distinction between *natural* kinds, on the one hand, and the collection of more disparate things, e.g. all white things, on the other, he says:

> There is no impropriety in saying that, of these two classifications, the one answers to a much more radical distinction in the things themselves, than the other does. And if anyone even chooses to say that the one classification is made by nature, the other by us for our convenience, he will be right; provided he means no more than this: Where a certain apparent difference between things [...] answers to we know not what number of other differences, pervading not only their known properties, but properties yet undiscovered, it is not optional but imperative to recognize this difference as a foundation of a specific distinction [...] the necessity of taking notice of this depends on the importance or unimportance of the particular qualities in which the difference happens to consist. (1843: 123)

According to Mill, there is no real essential/accidental distinction and logicians keep talking of kinds but simply as collections of attributes. We see then that Kripke's rejection of Mill's view that 'general' names like "cow", "horse" and "tiger" express properties, packs together three distinct theses. First, these names express no senses. Second, they are really singular names, i.e., they are referential non-predicative terms. Third, their referents are not clusters of properties but object-like things, i.e. biological species. Why is Kripke running together these three points?

Before I venture to answer this question, let me consider briefly Kripke's famous Frege-Russell view. According to Kripke's Frege-Russell view, which he rejects, "[t]o every name [...] there corresponds a cluster of properties" and the referent of the name is the unique object, if any, that satisfies most, or a weighted most, of them. (1980: 64 and 71). The Frege-Russell view is neither Frege's nor Russell's. According to Frege names express senses and refer to individuals. Senses, we have already seen, are not properties. According to Russell, names are disguised descriptions and descriptions do not have referents, in the sense that the denotation of a description is irrelevant to the semantic analysis of the sentences that include the description.[21] I do not want to recapitulate

---

[21] I am speaking of the Russell of "On Denoting" (1905), the one that Kripke is criticizing.

here the important differences between Frege and Russell.[22] Nor do I
want to recapitulate Russell's theory of descriptions.[23] I also think that,
despite their crucial differences, Frege's and Russell's views do share
some important general features, those that Kripke hints at when he
says that they associate names with clusters of properties. Kripke need
not enter into the specifics to reject both views at once. His arguments
indeed apply both to Frege's and Russell's versions of the Frege-Russell
view. Nonetheless, I do not think that it is just convenience that induces
Kripke to overlook the distinctions between Frege and Russell. The dis-
tinction between Frege and Russell that he is neglecting is fundamentally
the same he overlooks when he deals with Mill on common names. The
'properties' he is talking about can be either Fregean senses or Russel-
lian attributes. Kripke rejects Fregean senses for proper names and for
definite descriptions too. This is once again the endorsement of a di-
rect reference relation. But when it comes to Russell, Kripke disagrees
only on names. Names are not descriptions. On descriptions however he
agrees. This is the case, because Russell's analysis of descriptions does
not postulate any indirect semantic relation. The terms that make up
a description are directly referential, if referential at all. Russell's 1905
denoting relation is not semantical. It is equivalent to Mill's relation be-
tween attributes and their bearers. In the case of descriptions, Kripke
has no metaphysical ax to grind, but the confusion is the same. Kripke
does not care to distinguish between senses and worldly attributes. And
when it comes to attributes, he is careless about distinguishing their
employment in semantics from the metaphysical doctrine that natural
kinds are collections of attributes. Why is he blind to these distinctions
(at least when rejecting those theses)?

Carnap and possible world semantics

In this section I briefly speculate on what I take to be the source of the
confusion. The culprit once again seems to be possible world semantics.

---

[22] See Kaplan 1975 and Bach 1996 on this.
[23] See Kaplan 1969 and 2005.

Conceptually, the first step is Church's development of an intensional logic to formalize Frege's theory of senses:

> While we preserve what we believe to be the important features of the theory of Frege, we do make certain changes to which he would probably not agree. [...] Another is the abandonment of Frege's notion of a function (including propositional functions) as something *ungesättig*, in favor of a notion according to which the name of a function may be treated in the same manner as any other name [...] (But it is even possible that Frege might accept this latter change, on the basis of the understanding that what we call a function is the same thing which he calls *Werthverlauf einer Funktion*.) (Church 1951: 4)

Frege's concepts are the first victims of the rigorous principle of systematization. Recall that, if my analysis is right, Frege's concepts are not just an ontological caprice. If they were, that is if they played no semantic role, then it would be perfectly fine to ignore them when the goal is to formalize Frege's semantics. However, Frege distinguished the semantic role of senses – meanings as parts of thoughts and reference-providers – from the semantic role of concepts: predication. Using concepts for this purpose may not have been his finest idea, but it is a mark of Frege's genius that he recognized the distinction.[24] When Church gives us a logical reconstruction of Fregean semantics, the concepts disappear (though Church calls senses "concepts").

If we try to reconstruct Frege's semantics from Church's formalization, which choices do we have? With one less distinction at our disposal, we may give up predicativity. But who wants to do that? Language cannot be merely referential. There must be predicative expressions to hold our sentences (not to mention our thoughts) together – to make them sentences as opposed to lists of words – and to enable us to say that it is *a feature of* Secretariat that it is brown. So, the natural way out has been to make senses do all the predicative work. Now Church's senses do the work of Mill's connotations and naturally of Frege's senses too. We lose track of the distinction and we start to think that Mill's connotations are Frege's senses. Even worse, all senses soon become predicative. Now it looks as if Fregean names are predicates too. With the collapse of the sense-attribute distinction, the direct-indirect and the referential-predicative distinctions become one. No wonder Kripke does not track

---

[24] This is not to say that Church wasn't a genius.

them. No wonder Mill becomes Frege and Frege becomes Russell. Incidentally, Church's project is purely logical, but notice that some may welcome, indeed have welcomed, the collapse on philosophical grounds. Church's intensional logic naturally suggests a sober nominalistic metaphysics (the world is made up of individuals and collections thereof) accompanied by a grandiose view of the mind, as the one and only 'predicative force' whose meanings cement the universe.[25]

As temporal logic aims to capture in a rigorous formal way the most general (logical) truths about time and their structural relations, similarly Church's intensional logic aims to capture in a rigorous formal way the most general (logical) truths about senses and their structural relations. The senses themselves are not reconstructed as logical entities. In fact, the logic is given with the intended interpretation in mind. The system is consistent with the non-intended interpretation that takes a sense to be the same as the object it refers too. Church notices that this result ends up paying some logical dividends, yet from an interpretive point of view he regards it as a shortcoming (1951: 22–23).

From the formal point of view, we may regard Carnap as taking a step beyond Church. In MN, Carnap provides a model-theoretic semantics for an intensional language. Carnap's model theory is a form of possible world semantics, with state-descriptions playing the role of the worlds. The state-descriptions correspond to the models of first-order logic. Terms are assigned values at worlds in a standard way: sentences are assigned truth values, singular terms are assigned objects, and predicates classes of objects. By definition, two terms of the same syntactic category have the same extension if and only if they are equivalent, and the same intension if and only if they are L-equivalent (1956: 23). And two terms are L-equivalent just in case they are assigned the same values at all worlds. Thus, the intension of a term is standardly identified with the function that assigns to it values at worlds. That is, an intension is a function from worlds to extensions. Carnap is ontologically neutral concerning the existence of intensions, insofar as the metalanguage does not need to refer to them (1956: chapter 4). This however is a result of Carnap's general views about ontology, and not a form of skepticism con-

---

[25] I am aware that these are very sketchy remarks.

cerning intensions in particular.[26] Nonetheless, less ontologically wary logicians happily endorse intensions. Intensions are mathematical functions and Carnap is happy to say that in his methodological system intensions correspond to Fregean senses.

There is no doubt that Carnap's semantic methodology (possible world semantics) has advantages over Frege's. Nonetheless, it has disadvantages too. We have already discussed the case of rigid designators. The same intension corresponds to a name like "9" and a description like "the sum of 3 and 6", but it is not clear that they have the same meaning. Let us return to the case of predicates. Carnap reserves the term "property" for the intensions of monadic predicates ("predicators" in his terminology), which he is willing to call "concepts" too. He is aware that the informal notion of property takes them to be attributes to be found in things, as he puts it "something physical that things have" (1956: 20). Nonetheless, in his methodology we need not take any such stand on the nature of properties. Indeed, we are not even committed to their existence at all.

We see then that in Carnap's formal semantics, the same entities (intensions) play the role of both Fregean senses and Millian attributes. And indeed Quine's anti-Carnap rejection of intensions is a rejections of meanings, attributes, and relations all at once (see Quine 1960: 206–211).

Considering that Kripke's own work in possible world semantics is a technical development of Carnap's, the discovery that Kripke's notion of property, like Carnap's, obscures the distinction between senses and attributes turns out to be not surprising at all. In possible world semantics, we lose not only the distinction between predication and sense-mediation (a loss from which I believe we haven't recovered yet). This semantic distinction may well be of no particular interest to Kripke. But possible world semantics obscures also the distinction between semantical and metaphysical matters. For example, a sentence true in all worlds can be seen as true in virtue of its meaning, i.e. true independently of the vicissitudes of the world (similarly to the way in which names are rigid), or as true in all circumstances (like a *de facto* rigid description). Similarly, possible world semantics cannot keep track of the distinction between the semantics of kind terms and the metaphysics of kinds. Suppose the

---

[26] See Carnap's "Empiricism, Semantics, and Ontology" reprinted as a supplement of Carnap 1956.

extension of the word "tiger" in each world is the class of things that satisfy the set of attributes XYZ in that world. Possible world semantics is silent on whether this is the case because the word "tiger" is short for "XYZ" (a semantic thesis), or because it (rigidly) names a natural kind which may be (as Mill thinks) nothing over and above the cluster of attributes XYZ or instead (as Kripke thinks) a particular biological species whose members are essentially XYZ. For better or for worse, Carnap does not care about these distinctions. He endorses possible world semantics exactly because of its clarity and neutrality, and because it eliminates what he may well have regarded as philosophical pseudo-problems. But Kripke is no Carnap, and possible world semantics does not serve well his philosophical agenda.

References

Almog, J. (1986), "Naming without necessity", *Journal of Philosophy*, 83, pp. 210–242.

Bach, K. (1996), "Comparing Frege and Russell", MS, `http://userwww.sfsu.edu/kbach/FregeRus.html`.

Ballarin, R. (2004), "The interpretation of necessity and the necessity of interpretation", *Journal of Philosophy*, 101, pp. 609–638.

Beaney, M. (ed.) (1997), *The Frege Reader*, Blackwell, Oxford.

Carnap, R. (1956), *Meaning and Necessity*, University of Chicago Press, Chicago, 2nd edition.

Church, A. (1951), "A formulation of the logic of sense and denotation", in P. Henle, H.M. Kallen and S.K. Langer (eds.), *Structure Method and Meaning*, The Liberal Arts Press, New York, pp. 3–24.

Donnellan, K. (1966), "Reference and definite description", *Philosophical Review*, 75, pp. 281–304.

Fara, D.G. (2015), "Names are predicates", *Philosophical Review*, 124, pp. 59–117.

Frege, G. (1891), "Function and concept", in Beaney 1997, pp. 130–148.

——— (1892), "On concept and object", in Beaney 1997, pp. 181–193.

——— (1892a), "On *Sinn* and *Bedeutung*", in Beaney 1997, pp. 151–171.

——— (1918), "Thought", in Beaney 1997, pp. 130–148.

Kaplan, D. (1969), "What is Russell's theory of descriptions", in P. Yourgrau (ed.), *Physics, Logic and History*, Plenum, New York, pp. 277–295.

——— (1975), "How to Russell a Frege-Church", *Journal of Philosophy*, 72, pp. 716–729.

——— (1986), "Opacity", in L.E. Hahn and P.A. Schlipp (eds.), *The Philosophy of W.V. Quine*, Open Court, Chicago, pp. 229–289.

——— (1989), "Demonstratives", in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford University Press, New York, pp. 481–563.

——— (2005), "Reading 'On denoting' on its centenary", *Mind*, 114, pp. 933–1003.

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

——— (1977), "Speaker's reference and semantic reference", *Midwest Studies in Philosophy*, 2, pp. 255–276.

——— (1980), *Naming and Necessity*, Blackwell, Oxford.

Marti, G. (2003), "The question of rigidity in new theories of reference", *Noûs*, 37, pp. 161–179.

Mill, J.S. (1843), *A System of Logic,* Part 1, John W. Parker, London, reprinted in *Collected Works of John Stuart Mill*, J.M. Robson (ed.), Volume VII, University of Toronto Press, Toronto 1973.

Montague, R. (1970), "English as a formal language", in B. Visentini (ed.), *Linguaggi nella società e nella tecnica*, Edizioni di Comunità, Milano, pp. 118–221.

Quine, W.O. (1960), *Word and Object*, MIT Press, Cambridge, MA.

Ricketts, T.G. (1986), "Objectivity and objecthood: Frege's metaphysics of judgment", in L. Haaparanta and J. Hintikka (eds.), *Frege Synthesized*, Reidel, Dordrecht, pp. 65–95.

Russell, B. (1903), *Principles of Mathematics*, Cambridge University Press, Cambridge, reprinted by Norton, New York 1964.

——— (1905), "On denoting", *Mind*, 14, pp. 479–493.

Reference without denotation
Semantics vs. model theory, natural vs. formal language*

Joseph Almog

1. This note is about contrasting two relations – reference vs. denotation – that have been thought to operate at the fundamental semantic level. Variants of the latter, as introduced half a century later by Carnap, are called "extension" and "designation". I am interested in understanding how the two kinds of relations arise in the setting of two 'sciences' of *value-association* with expressions – semantics vs. model theory – and two kinds of *languages* investigated by the two sciences – natural and formal.

2. In pursing this contrast – indeed, these contrasts – I pick up a theme dear to Ernesto Napoli.[1] I will not quite put things in his own words. Nonetheless, I believe I am after a sheaf of themes he noticed:

A.   Reference and denotation are two logically distinct relations.
B.   Proper nouns in natural language refer and do not denote.

[1] See Napoli 1995.

C.   More generally, simple nouns (common, deictic pro-nouns) in nat-
     ural language refer and do not denote.
D.   More generally yet, single words in natural language refer and do
     not denote.
E.   Compound phrases in ordinary English ("King of France (in 1905;
     in 1788)", "author of *Principia Mathematica*", "rich and happy",
     "walked and talked", "John and Mary") do not *simply* refer. But
     they also do not primarily denote (let alone have senses and deno-
     tations). These compositions *combine* referents of the ingredients.
F.   Wherever reference takes place – and reference, though not denota-
     tion, takes *place* (in historical space and time) – it is 'direct' both
     in the intuitive sense of this word and in the technical sense in-
     troduced by modern semantics; there cannot be in-direct reference,
     just as there cannot be direct denotation.
G.   Denotation is not a (fundamental) semantic relation; it applies in
     a post-semantic stage.

3. I believe Napoli, in his own way, saw all of A–G theses and judged
that they make an organic 'ensemble'. Earlier, I believe that, in essence,
Bertrand Russell proposed A–G and the late Keith Donnellan came close
to embracing them for proper nouns and some other singular devices.[2]

4. When I say that Russell originated the essentials of this methodology
I mean – and at this time I focus on the large scale picture, not on the
local technical mechanics – that: at the level of *fundamental* expressions
– *genuinely* (and not merely *seemingly*) single words – the fundamental
semantical relation is *reference.* When we come next to genuine linguistic
compounds, Russell's method is not to proceed as is popular in many
modern circles, viz. to call upon a hybrid of Russell's reference at the
fundamental level and *denotation* (also, *extension/designation*) higher-
up, thus a second relation based on truth-evaluations, be they in 'the'
real world, in an index of possible world and moment of time, in a pair
of indices of context and possible world (time), in a triple of indices of a

---

[2] As pointed out by Bianchi, Donnellan developed a non-referential account of com-
mon nouns for kinds, though he notes that the word "bachelor" can be taken to refer
directly (and referentially!) to bachelorhood. In the work of both Kaplan and Kripke,
there are intimations that all single word common nouns are referential (in my sense).
Thesis D is stated in Kaplan's 1971 unpublished lectures in Princeton.

model (structure), a context and a world, etc. When the method of truth (at an index) is generalized it makes *all* of semantics occur at this evaluational level – designations/denotations are computed for the ingredient expressions by evaluation at the index and the designation/denotation – true/false – for the full sentence is computed again by way of the 'local' pertinent facts.

5. Russell denies that 'denotation' plays the fundamental semantic role. Suppose "John" and "happy" are genuinely single words. Then each – provided it is not altogether empty – is a *referring* expression. But now, in explaining why the compounds, e.g. "Happy and Rich", "John and Mary", "John is happy", "John (and Mary) is (are) happy and rich" etc., are *significant units* in the language, we do not abruptly change gears and shift to the method of truth and its denotation-assignment, we do not look at what 'satisfies' these compound expressions or what they are *true-of.* Not at all, says Russell. Like the fundamental units, e.g. "John" and "happy", semantics loads *composed units* with semantic values – composed from the ingredient-referents – prior to assessing what is true. Semantics precedes truth.

6. This may exclude semantics from assigning at the fundamental single word level extensions/denotations/designations and their modal or model theoretic generalization – functions from indices to extensions. But it may seem to still allow the fundamental semantic values – for either the ingredients or the composed items – to be 'primitive' and pre-truth units like Frege's senses.

7. This again is where Russell denies any drift from referentialism. The denial is in two stages. At the fundamental level, the semantic significations cannot be 'new' primitive entities that are not world bound and associated – in historical space and time – with the expressions by a natural-historical process connecting the entity with the expression. It is *referents* – say the planet Neptune – that are so associable to the name. Assuming "Neptune" to be a genuine single word, it cannot be loaded with a sense, only a world-bound referent.

8. And now enters Russell's second stage, wherein he is concerned with compound expressions: no Fregean 'penthouse' construction of senses-determining-denotations is available for the semantics of the compound units because their semantical referents are *compositions* out of the *ref-*

erents of the ingredient units. What is the *correct* composition of a compound like "the queen of England" or "every queen of England" or just "queen of England" or "happy and rich" or "John and Mary" is a matter of further investigation. But the composed locution has only the referents of the ingredients to operate on. Perhaps we end up with composed-referents (e.g. a sequence of the ingredient referents) or perhaps we give quite a different composition, viz. a de-composition of "the queen of England" to quantifiers, attribute-symbols etc. Either way, no new second level semantical units – not induced by the ingredient-referents – are posited. We have the referents the world has provided for the ingredient locutions and modes of composition and that is all. Semantics 'higher up' viz. for the (seemingly) composed units is a *compositional-referential-semantics*. Jump starting the semantics with alleged counterexamples based on first blush implicit presuppositions about modes of composition is not a healthy methodology. One ends up positing in a rush higher basic semantic values, without attention to the genuine grammar of the construction. Thus the temptation to pronounce that "It is necessary that George the IV believes that Scott is Scott (Walter, the author of Waverley)" is a counterexample to a referential (not denotational!) semantics should be tempered. We return to such apparent counterexamples below.

9. Quite apart of Napoli's or Russell's reasons for holding theses A–G, I see the theses as a natural organic body of theses and I'd like to provide a framework that makes them true. In pursing these issues, I will not wax 'historical' with quotes from Russell etc. If the theoretical distinctions I am after will be clear to the reader, he will see, so I hope, that Russell basically had it all. And I conjecture Napoli, in his way, appreciated versions of the theses. We cannot ask Russell whether he did but it'd be interesting to know whether Napoli did.

Intermezzo – Natural vs. formal language, semantics vs. model theory

10. The discussion below introduces one key factor that I do not think Russell was concerned with: the distinction between natural-historical

language (e.g. English) and formal languages (e.g. the language of first order (modal) logic). If anything, Russell thought that certain devices of formal languages, individual constants and individual variables, provide the model for singular 'direct' reference to individuals by English proper nouns and indexical pronouns. And this, I think is not quite right.

11. Though I will speak of a (given) formal language as a *language* and its semantics as a *semantics*, in the end, this is a mere stopgap to facilitate communication. As will emerge, I do not merely deny that proper names of natural language such as "Nixon" and deictic pronouns such as "he" are to be represented by/reducible to individual constants and variables of formal languages. Stronger yet, I submit that a formal language is not really (each) a *language*. And the model theory(-ies) we give such a language is (are) not a *semantics*, at least not in the sense natural-historical ordinary English is a *language* and has an *essentially historical* semantics.

12. The language of e.g. classical first order logic is sure enough a certain kind of algebra of strings and it may be correlated – by model theory – with 'values' by assigning its strings 'values' and computing for its sentence-level strings validity and model-theoretic consequence relations. The case of the language of sentential logic should always be our paradigm – the two-valued tables and the functions on these values display a fine model theory but any reading of a *semantics* of English words like "or" and "if, then" (and "and"!) is premature. The (standard) model theory of the formal language of sentential logic is just that – a model theory (a two-valued algebra).

13. We may well say our common use in philosophical logic of the terms "language" and "semantics" to cover both phenomena – the natural-historical and the as-if-transcendental and ahistorical – is as with Hilary Putnam's word "jade" where, chemically, two very different kinds hide behind a somewhat similar appearance. Indeed our case is even more dramatically disjoint than Putnam's. The two Putnam jades are *both* intra-natural history. In our case, the analogy is between *natural* and *hyper-natural* kinds.

14. Speaking of Putnam, it was he – rather than Russell – who anticipated this second main theme of the present essay in his important – but not often noticed – piece "Is Semantics Possible?" (1970).

The title says it all. Putnam observes that classical formal 'seman-
tic' studies (by the late sixties) pursued things very much as in the
model theory of formal languages, first or higher order. The focus of the
model theory is not the 'atomic' level of individual constants, monadic
predicates, relations etc. but rather it is the (sentential) operators that
correspond to compositions of complex (iterated) structures. At the level
of the 'lexicon', we are given a very schematic assignment, in a model
(on an assignment) of individuals and sets, the extensions/denotations.
Putnam observes that if this is the method with which we are going to
pursue natural language, we may as well agree with Quine, and abandon
semantics (of English!) as a natural-historical *science.*

The key, says Putnam, is to invert the abstract focus of the model
theory of formal languages. In natural language studies, it is the *lexicon*
where all the (historical!) action is whereas the compositional 'connec-
tives' are rather secondary. Putnam then goes to exhibit such a lexical-
semantics investigation, his famous account of count and mass nouns
such as "elm(s)", "beech(es)", "water" and "gold". It is precisely this
focus – the lexicon of English – that we adopt below and with an eye
to providing a *natural-historical semantics*, not a model theory that ab-
stracts away from the natural historical.

15. So, I do not assume that *lexica* of English such as "Aristotle" and
"clever" are 'represented' or 'symbolized' by formal language locutions
such as "a" and "F"; I do not assume that the sentence "Aristotle was
(is) clever", even ignoring tense (!) – is represented by "Fa". We may *dis-
cover*, in due course, that there are revelatory correspondences between
the *two* cases. But we cannot (pre-)suppose, as most modern model the-
oretized versions of semantics of English do, that just as a matter of
course we *symbolize* "Aristotle was clever" by "Fa", as if this was short-
hand script of a secretary who was in a hurry to take dictation. "Fa"
is no shorthand for English. It is a string generated by certain rules we
stipulate.

16. Let us concentrate on the simplest case pertinent to our singular
reference/denotation concern, the sentence "Fa" or open-sentence "Fx"
– both are not historical units. The syntactic shapes in question – "F",
"a", "x" – are given their identity independently of history and surely
independently of being loaded with particular historical entities as values.
We write down constructed rules – *rules of denotation* – stipulating, say,

that "c" will stand for the number 0 or, on another *application*, for the man Richard Nixon; "F" is to stand for the property of being (a/some) republican (a variant: the set of republicans); and on a given assignment, we assign Richard Nixon to "x".

17. Let it be noted that the varying models or assignments are not what philosophy of language became very focused on in the late fifties – modal operators and their 'interpretation' by means of possible worlds, or counterfactual alternatives to the real world. Models are prior altogether to the considerations of modal questions: models – in the sense of *interpretation-providers* – precede modals. They are also prior to questions of contextual use – who is speaking, who is pointed at and what is the time of use. It is one thing to let the shape "I" signify the agent of the context, as it does in standard English, and not a cry of pain, which it does in other languages, analogous to "ouch" in standard English. Once we have assigned this interpretation, we may ask, who is the speaker who uses "I" in this paper? The answer is the man, JA, though when Napoli writes his own paper it is he, EN, that's picked up by his use of "I". We may add to the standard formal language of non-modal non-indexical predicate logic – in the manner indicated initially by David Kaplan and Hans Kamp – a context and a possible world coordinate to represent contextual facts and modal variations in the facts relative to which we evaluate sentences (and other compounds) for truth or denotation/extension/designation.

18. It is said that the double life of models is confusing. But there is no intrinsic mystery here. If we have two individual constants "c" and "d", there is a sense that if "$c = d$" is true at all, we want it to be fixedly – modally-necessarily – true across the worlds of one modal structure, and at the same time, another sense – one of varying *interpretations* – on which "$c = d$" must be such that it turns false on other 'interpretations', in distinction from "$c = c$". "$c = c$" turns out (necessarily) true in every model/structure, even if in different models, it expresses different necessary truths. Very well then, say that "$c = d$" is evaluated relative to a model(structure) $\mathcal{M}$, and possible world w. Let $\mathcal{M}(c) = \mathcal{M}(d)$, say it is this one Roman orator, Cicero (Tully). We can go on to demand that in all other possible worlds w' of $\mathcal{M}$, the sentence – as interpreted by $\mathcal{M}$ – is fixedly ('necessarily') true. Furthermore, on an alternative model/structure $\mathcal{M}'$, $\mathcal{M}'(c)$ may well be different from $\mathcal{M}'(d)$, and then,

at all possible worlds in that alternative *modal* structure, when so interpreted (as by $\mathcal{M}'$) "$c = d$" is to come out fixedly (necessarily) false. The same apparatus can be used for the free variable case of "$x = y$", when the assignment is segregated from the modal/possible world parameters.

19. What matters in the present context is that the input for this calculation is the extension/denotation/designation of the ingredient locutions, such as our constants "c" and "d", relative to the said parameters. The model theory is exclusively computing extensions/designations of compound units from those of ingredient units. None of this involves the historical relation of reference, which is not between formal symbols (formulas) and models but between – users (real people!), natural language historically 'loaded words' (on which more in a minute) and real objects.

20. And so it has become for us natural to compare three frameworks – the model theory of pertinent formal languages, the assimilation offered by certain denotational model theories for 'analog' natural language names (and descriptions) and the semantical account by reference-and-no-denotation that I see emerging in Russell and Donnellan. I believe both Kripke and Kaplan have made vivid to us in their 'intuitive' remarks about historical relations of reference a similar referential framework. But their encoding within the denotational/designational model theory is blurring just what needs to be kept apart – reference vs. denotation.

21. Let me say a word – many more elucidatory words will follow below – about the 'picture' I have of this relation of *re-ferring*. In a word, it is all encapsulated in the etymology of "referring", as it comes from Latin and is present in contemporary Italian and French. To *re-ferrer* in Latin alludes to returning to the original cause of something ferried-carried to you. And so I think of referring as proceeding as follows – an initial loading of a natural language historical product, a word, takes place, whereby the word, say "(John) Locke" is loaded with an individual, and is originated as a name for, a carrier of (as if pasted to its very orthographic appearance) that man.

22. The variations abound – see "Madagascar" below – and are interesting to play with but the passion for clever cases should not distract us too early. At bottom is a simple story – in natural history, at a particular moment and corner of space, a word gets loaded with an object and

it is thereafter carrying it and allows those who receive it to 'perceive' what is carried, in the key sense that does not involve *imaging* the entity carried.

23. There are many similarities to the case of the photograph, where imaging is involved, but is not key to *what* object the photograph ferries. The photo is also a loaded item and not one to be confused with the *photograph-apparent*, the surface look, that may carry-ferry simultaneously many look-alike different individuals. What is key at this early stage is to take notice of the historical nature of this product – word, photograph – and its *determination* to be what it is by the origination process that brings it into historical existence. As with other such products – trees, men, ships etc. – later, the produced entity may alter its apparent form considerably and its apparent form may come to be associated with another entity (as well). When I receive and thus am in a position to perceive the word, I perceive with it the individual it ferried. I now re-fer (back) to it.

24. This was an initial gloss of our third framework, the referring relation framework. I spoke of three frameworks to be compared. The first two are compared by Kripke himself in his 1980 introduction and in Lecture I of "Naming and Necessity" (Kripke 1972, henceforth NN), where he observes that he first noticed certain possibilities about constants in formal (modal) languages and their model theory and he then asked himself how they extend to proper names (descriptions) in natural language.

Let us encapsulate in a few theses the 'standard' profile of a constant in the model theory of the pertinent formal languages:

(F1) Constant "c" – on a given interpretation – is not ambireferential – one *constant*, one denotation.

(F2) "c" is schematic and allows many interpretations.

(F3) Supposing "c" on a given interpretation to designate entity x, "c" *might not* have designated x and might have designated some other entity y; it is not of "c"'s *identity* as a locution that it designates x.

(F4) Suppose entity x is designated by constant "c" (on a given interpretation). x *might not* have been designated by "c".

(F5) "c" *denotes* c is a semantical truth, analytic and *a priori* (if not modally necessary).

(F6) We may have constants whose denotation is fixed by the rule: "c" denotes whatever is the unique (The) F; this rule of denotation is known *a priori* even if it is modally contingent.

(F7) "$c = d$" if true (on an interpretation) is necessarily so, even if not *a priori* known.

(F8) No genuinely empty constants are allowed.

25. In the introduction of NN, Kripke considers the historical proper name "Aristotle" and later he considers empty names such as "Vulcan". Regarding such historical natural language names, it is said – in the attempt to model them after the constants just discussed – that:

(NMF1) "Aristotle" is *ambireferential(-designational)* but this is not a fundamental theoretical problem for the semantics.

(NMF2) "Aristotle" may be disambiguated so as to have for each disambiguation a different linguistic function.

(NMF3) Suppose "Aristotle" on this one disambiguation denotes/designates the Greek philosopher x. Even on this disambiguation, "Aristotle" *might not* have designated x and might have designated some other entity y, say Plato; it is not of "Aristotle"'s *identity* as a locution that it designates x.

(NMF4) Suppose entity x is designated by "Aristotle" (on a given disambiguation). x *might not* have been designated by "Aristotle" at all.

(NMF5) *"Aristotle" designates Aristotle* is a semantical truth, analytic and *a priori* (if not modally necessary).

(NMF6) We may have names whose denotation/designation is fixed by the rule: "Neptune" denotes whatever is the unique (The) Uranus perturber; this rule of denotation is known *a priori* by the stipulator even if it is modally contingent.

(NMF7) "Cicero = Tully" if true (on an interpretation) is necessarily so, even if not *a priori* known.

(NMF8) English has empty names, e.g. "Vulcan", and there is no possible but not actual entity they name.

26. Before we linger on the significance of the differences between the formal model theory for constants and the analog for proper names advanced by the introduction of Kripke's NN, let me put before us the third

profile, of the referentialist – and no denotation – semanticist, he/she who denies the modeling after formal language constants:

(NH1) "Aristotle" is *ambireferential*: ambireferentiality is inevitable (necessary!) for natural language names.

(NH2) "Aristotle" is a single *name-apparent*; for the two described uses (viz. to refer to the philosopher and to refer to Onassis), we have two different historically determined loaded names.

(NH3) Supposing "Aristotle" on a given historical use to re-fer to entity x, "Aristotle" on that use *might not* have not-referred to x. It could not have designated – on that use – some other entity y; it is of "Aristotle"'s *identity* as historically-determined loaded name that it refers to x (or as with "Vulcan" that it fails to refer altogether).

(NH4) Suppose entity x is referred on this use by "Aristotle". Still, x *might not* have been referred to by "Aristotle".

(NH5) *"Aristotle" refers to Aristotle* is a historical-semantical truth, necessary and *a posteriori*.

(NH6) No 'rules of denotation' may stipulate in abstraction that a historically determined name refers. Only a natural-historical process can initiate and determine reference.

(NH7) "Cicero = Tully" if true on a given use, is necessarily so, and it must be *a posteriori*; "Cicero = Cicero" is not *a priori*.

(NH8) Genuinely empty names exist and are inevitable (necessary!). Such historically determined names are necessarily empty though the name-apparent may well refer on a different use.

27. Comparing the three profiles, it seems the only item fully agreed upon is the ('*de re*') modal fact that a given entity, say the famed Greek philosopher Aristotle, might not been so called (called "Aristotle"), indeed he may have been so revered that no one would be allowed to call him by any name.

28. The entity might not have the name it actually bears. But the converse direction is in intense dispute – the formal model theoretic methodology and following it the designation theory applied to natural language assert that a given word (here: name) of natural language could be the very word (name) it is and not designate/denote the entity it actually does. This is obviously at the heart of the model theoretic approach to a formal language, at least for the nonlogical symbols.

29. The key to the nature of expressions in the formal language, as understood in the Hilbert tradition, is that the identity of the expression is given by apparent and transparent symbols; the model theory, in turn, 'works on' in its recursion clauses on that apparent level. We start with a symbol – given by its shape – and try to connect it to a value (by a given model). It is obvious that the connection set is fragile because the identity of the symbol is given *independently* of the posterior connection to the object. Furthermore, the pertinent object settled on, say x, is due to *satisfaction* of rules of denotation that are the *makers* of the connection. Thus we open ourselves to the 'vulnerability' that, as long as some conditions are satisfied (e.g. an underlying axiomatic system is seen to be satisfied by the model), the denoting by the schematic constant or predicate or relation locution could be altered while the formal symbol – with its identity *already* given independently – is staying the same symbol. Another possibility that is opened because of the independence of the identity of the symbol from the posterior connection is that a rule of denotation can *stipulate* for a given symbol a new denotation – say the planet Neptune – just because it is this entity that satisfies the determining rule.

30. The vexing problem of *determination of reference* (I would say of *denotation*) that Kripke selects (NN, pp. 28-30) as the prime problem of a theory wishing to improve on Frege and Russell is a problem with a model theoretic flavor, if it assumes that the name is *given independently* of an already loaded history, as if one would ask of a certain surface appearance of a person – what makes that appearance to be *of* Barack Obama? In our case when we ask of the as if historically immaculate shape "Nixon" what determines its 'referent' (denotation), we are abstracting, as in the model theory case, from the obvious determiner – the man Nixon who started-up this loaded name.

31. The very question supposes we have a gap between name – that we have 'here' before our mind – and the far away object. This confuses an epistemic question – how do I know of this shape *whose* shape it is? with a metaphysical question of process of determination – this shape (shadow etc.) of Obama is of ... Obama because Obama engendered it. In the model theoretic case there is a genuine metaphysical issue because, in itself, the symbol is not determined with a denotation. Not so with a natural language name.

32. In contrast, the relation of *reference* is determined at the outset – *historical* outset – and there is no determination problem to be solved. When the name-apparent "Aristotle" floats in front of my eyes in a philosophy textbook, in fact it is the appearance of a historically determined name loaded 2300 years ago with that ancient Greek man. I may not be able to tell (to 'determine' in this *epistemic* sense) with which entity the historically transmitted name is . . . *determined*. But it – the loaded name – is determined, *on ne peut plus*.

33. The historical name is not different from any other historically generated entity. When I first lay my eyes on Mary Kate Olsen, I may not be able to 'determine' whether it is Mary Kate or Ashley – her identical twin sister. Who is the person who has appeared? But nothing in the identity of the person who appeared is un-determined – assuming it *is* Mary Kate, her identity has been determined at early stages of conception, her origination process did it all. All that is left for me now – upon perceiving the gal – is to track which *determined* person is impacting my perceptual system.

34. And there is more in this vein. No doubt if I travel far and away, say to Madagascar where I lay my eyes upon a person with the/a Mary-Kate look, I may well take myself to continue previous usage. So when I and my group of local friends call the local person "Mary Kate Olsen" I take myself to just go on with previous use. In fact, I have been now using this name-apparent "Mary Kate Olsen" to introduce a new historically-determined name for Ashley Olsen, who is touring Madagascar. If my group is socially influential, at least in those areas of the world, and assuming the 'old' Mary Kate never shows up, this new name for Ashley – "Mary Kate Olsen" – will take over. If, as luck would have it – life is always stranger than art with Hollywood artists – back home, Mary Kate decides to change her name to "Princess" (for gender-parity's sake, following the singer called "Prince") and not be called anymore "Mary Kate", soon enough only . . . Ashley will be called "Mary Kate Olsen".

35. The reader will have noticed this is what happened, in essence, with "Madagascar" famous tell-tale. We do not have one name that *shifted* its denotation. Rather we have one *name-apparent – one* that was *co-initiated* with the *loaded name* "Madagascar", originally used for Mozambique. Then, by accident rather than by premeditation, we went through a new *origination* of a new loaded name, with the same appear-

ance, now for the island. In all, there is no mystery. The re-ferring I do now with "Madagascar" is determined not by the shape of the word but by *which entity* was loaded into the signal I have received – most likely it is the island. But if old texts were unraveled of travelers preceding Marco Polo, travelers who got the local use correctly and referred to what the word they received was then loaded with – Mozambique – I'd have in my repertoire one apparent name but, in tow, two *loaded* names, exactly as with "Aristotle".

36. There is more in this vein, the vein of questioning our 'control' over names in the natural-historical language. First, as already mentioned about (NH6), it is not a stipulated 'in the head' rule of denotation that makes a name bear the reference it does. What makes the historical entity, the loaded name carry the entity Neptune, is not the description Le Verrier uses later to communicate to his fellows what entity he is in contact with. What is key is the actual-contact relation between the planet and Le Verrier's perceptual system as he loads the name with that entity that he has been impacted by. I would have liked to say – by the entity he is *acquainted* with, if only Russell did not taint this common vernacular verb, making it now mean in philosophical circles a very demanding immune to error relation that I bear to nobody, not even my own body.

37. So, denotation may – indeed *must* – be fixed by rule-satisfaction and we hope for a unique satisfier. Referring in contrast is not induced by satisfaction of rules. Referring is a worldly relation of *re*-ferring back to an entity that has been ferried to you. The signals bouncing off Uranus – just like the footprints of the burglar in the garden – have ferried to me the original cause – Neptune (or that burglar – Frank Jones). It is to that ferried item that I now re-fer using a word loaded with it.

38. So much for 'fixing' or 'determining' the denotation but not the referent which is determined at the outset. Next, we lingered in (NH7) on a related issue – all identities of the form "Cicero = Tully" are, if true, necessarily so and yet none are *a priori*, not even "Cicero = Cicero". If I was – to use Donnellan's deft phrase, an *omniscient observer of history*, they would all be *a priori* for me. But I am not. And given my cognitive constitution, I am *necessarily* not.

39. It is in this mundane fact – in both senses of "mundane" – that lies the dissolution of various versions of what we call now *Frege's puzzle*. The informativeness of both "Cicero = Cicero" and "Cicero = Tully" for me, at this late point in history, is not due to an *over*-loading of the names on the sides of "=" with reference-transcending 'presentations'. Many do not have these legendary 'in the head' discerning presentations. And even if I did – following a repeated reading of Quine, I do have two distinct associated descriptions for "Cicero" and "Tully" – these descriptions are not the *determiners* of the informativeness of the original identity sentence.

40. When I ponder "Cicero = Tully (Cicero)", I am not contemplating *another* sentence which is not an *identity* but a sentence telling me that one and only one thing is both the famous author of *De Fato* and the denouncer of Catiline. The source of the informativeness of "Cicero = Tully" is that very sentence and no other. I cannot settle the truth value of this sentence on the basis of information I do have, 'in the head'. This sentence is thus informative for me at this moment in history precisely because I am not omniscient about the history of the two names and what they ferry to me. Notice that if my cognitive basis was different, if I had 'tunnel vision' unto the past, I could – just by looking at the name-apparent – figure out the question. But as things stand, on the evidence internally available, this sentence I am looking at – taken as involving apparent names – could well produce a necessary falsehood. And so perception of the surface of the page of the appearance "Cicero = Tully (Cicero)" leaves me in the dark.

41. The effect of the historicity of the names on historically limited cognizers reaches further. It becomes clear why on the referentialist semantics there are and *must* be empty names, though the denotationalist does his best to *eliminate* them either by making – by stipulation – *all* names have a denotation (classical logic) or let names stand for possible objects. Indeed suppose the *identity* of the name does not depend on the referent loaded into it by an actual historical process. Rather, as in the alleged Neptune-case the name's identity is molded by the rule of denotation fixing it with a denotation. Then, it might connote a blueprint for a possible object.

42. The referentialist reverses the flow diagram. We never really 'control', as Le Verrier allegedly did, the use of the name. Both the name

and the object come to us from history and this introduces 'vulnerabilities'. When we reverse the flow diagram, when we don't start with the name-apparent but start with the history bringing us the name, we see that the name apparent "Vulcan" comes to us from history just as the loaded name "Neptune" does. There is no telling which has been loaded *in* history. But one is loaded and the other is not, this much is a fact of their respective origination and thus of their identity. Different kinds of names – loaded and not – *were thus made* to be by two different kinds of historical processes. Vulcan, on my use of it, is necessarily empty. But I might receive tomorrow that apparent name from a historical chain that has loaded it with a referent all right. That referent would be built in of necessity to the identity of that other use – loaded – of "Vulcan".

43. These facts explain also how we *sort* the identity of historically determined empty names. Donnellan pointed out that "Père Noël est genereux" goes, in standard translation, over to "Father Christmas is generous", not to "Robin Hood is generous".[3] Why? Obviously, we have here no common referent or hyper referential common 'meaning' ('sense'). We do have a common origination for the translates and a distinct one for "Robin Hood". Like remarks applies to translating the English common noun "unicorn" by "licorne" and not "dragon", even though both of these French common nouns are equally empty. It is thus a truth of semantics – inasmuch as bilingual translations are – that "licorne" translates "unicorn" – and this is a historically generated and thus determined truth, not one of matching abstractly 'senses' or (empty) 'denotations'.

44. There is more coming to us from this 'historical pipeline'. We lingered for a moment on the alleged *theorem – "Neptune" refers to Neptune.* This much was said by the model theoretic designationalist to be (i) not necessary, (ii) known *a priori* and (iii) true by semantical rules of the language. On the referentialist account, we should not think of provability and theorems in this context. The profile of such claims are those of natural-historical truths. We can say *"Neptune" refers to Neptune* is a truth of semantics – for they are – but this need not connote a sheaf of *epistemically transparent* features – semantics lives on historical relations of ferrying up in history and re-ferring back.

---

[3] See Donnellan 1974.

45. If *"Neptune" refer to Neptune* is *a priori*, one would expect by existential generalization on the object of the verb "refers to" to conclude (with appropriate attention to tensing): there once was an object, Neptune, and it is the one referred to by "Neptune". This much would be . . . *a priori*! But of course, it is not *a priori* that Neptune once was (anymore than it now is). The fact that Neptune once was and originated the loaded name "Neptune" is a historical and not *a priori* fact.

46. But if *"Neptune" refers to Neptune* is a fact, it is a *necessary* fact. It is necessary because what this product name is – its very *identity* – is due to the process originating it. The apparent contingency is relating only to the *appearance* "N-e-p-t-u-n-e", which, sure enough, might have been involved in different uploadings, e.g. I could have called my dog "Neptune".

47. We may sum the fundamental difference between a denotational/designational model theory method and referential historical-semantics by attending to the questions: can a proper name denote/designate? Can a definite description refer?

48. Saul Kripke suggests in fn. 3 of NN and throughout "Speaker's Reference and Semantic Reference" (1977) that he uses a single scheme for any singular term *"t"* – *the referent of "t" is t*. This scheme is according to Kripke a standard principle in logical semantics (see fn. 3). He uses this principle to introduce the notion of *semantic reference*. He then separates semantic reference from another notion, attributed to Donnellan, *speaker reference*, which on Kripke's drawing of it applies to both names and descriptions. Let us attend to this distinction.

49. If I see a man – who happens to be Jones – raking the tree's leaves and I say "Smith is raking the leaves", my semantic reference is to *Smith* because by the anointed principle, the referent of "Smith" is Smith. According to Kripke, Donnellan makes Jones the referent but Donnellan should be corrected – Jones is merely the *speaker's* referent because I-the speaker have a mistaken belief about who the man by the tree is.

I will not linger on what Donnellan had in mind here and speak mainly for myself. I will only say that "the Donnellan phenomenon" Kripke alludes to is misconstrued by Kripke. It does not turn on having *false beliefs* by the speaker and thus the very idea of 'speaker reference' – as defined – misses its mark (viz. Donnellan). For Donnellan, the *referential*

use (and no speaker reference) by me when using the expression "the man with the martini" is as *semantic referent*-bound a use as there could ever be. I explain.

50. First, I do not refer to Frank in front of me in the party by means of any beliefs – true or false – viz. beliefs whose satisfier will *turn out* to be my referent. It is *incoming* perception from and of Frank and my reception of it that makes Frank the pertinent referent – Frank has been ferried to me by light bouncing from his body and now that I have him in mind – now that I am *related* to the source of that ferrying – I re-fer back to it. No beliefs forge the reference relation. Secondly, in the case described, if I believe anything at all, I may actually strongly *dis*-believe that this man, Frank, in front of me, is drinking the martini but I may find it expedient to *communicate* my *semantic* reference to Frank using this *false* description – perhaps *you* so believe and I know this phrase will make you attend to Frank. Stronger yet, perhaps both of us are 'in the know' (Frank is not drinking the martini) but both of us want the FBI – which is secretly recording the party – to believe that we both so believe, though in fact we don't. Much of this was made clear by Donnellan in his 1966 piece while discussing the case of "the King" who is not one but is referred by two of his aides using a description they *know* does not apply to the man. I cannot quite say – as Russell said of Quine – that Donnellan can now (or later) take care of himself. But I can say: he *did* take care of himself in his 1966 piece and I explain elsewhere just how well he took care of himself.

51. My concern here is with the phrase "semantic reference" because the way Kripke puts his own view (forgetting now the Donnellan part) proposes in effect not the relation of *reference* and not *semantics*, at least not as I understand those. What Kripke proposes is a (rigid) *denotation* theory. To unwind things, let us look at the case of Smith/Jones and the raking of the leaves.

I believe we encounter in this case the ambireferentiality of the name "Smith" (just as we do with "Aristotle" or earlier on, we imagined "Mary Kate (Olsen)" or "Madagascar" to be afflicted with). There are two historical chains affecting the user – as there might be with "Aristotle" or "bank". One chain is *local* – it runs from the surface leaf raker body to my eye and loads the proximate man Jones unto my use of "Smith",

"he" (even if it is a *she*) or "the leaf raker" (even if that individual is actually not raking the leaves but just feigning the motions).

On this local chain, my use of "Smith" is referential to Jones and this much is part of the *semantics* of "Smith" on such a use. The statement – on this use "Smith" refers to Jones – is part and parcel of how the English language works – what Kaplan called wisely "licensed uses of the language".

The second chain operative here is non-local. I have received for distal reference a certain man Smith brought to me by the common-currency-language loaded name "John Smith" (of course, "Smith", like "Aristotle" is itself ambireferential on its non-local uses, many global chains go back to many Smiths). But supposing the name is a bit more exotic, say it is "Barack Hussein Obama", this global chain ferries to me the name as loaded with our president, going back to a dubbing upon his birth (or conception) in Hawaii (!), 50 years ago or so. One use is induced by perception of the man by the tree, one by 'perception' of distal Barack Obama (or John Smith) via a global chain running back way beyond the current visual set up but ferrying to our distal perception an object uploaded far away from the current context.

This much is what the semantics of the natural-historical language should tell us, describing the *potential* uses. Which use is actually *operative*, which chain is dominant, is a further question. I do not believe it is the job of semantics to decide this further matter. Obviously, if we describe the case more, we can get information pertinent to the question of *which* chain has affected my brain/mind. But this much is to me a matter of cognitive psychology (of this agent and his specific psychological history). It is not the matter of the science of semantics. In like way, when I say "I am at the bank", semantics should describe the *potential* chains and uses. *Which* one is dominant in my case today at 2 pm is not the business of semantics or armchair philosophy. It will most likely turn out to depend on specificities of my cognitive psychology, much of it empirically determined by a host of events we have not specified or cannot glean from the 'case'.

52. What we can say is this: both potential uses are referential, neither is denotational (neither turn on satisfaction of beliefs or satisfaction of rules). Proper nouns are always – in any use – referential, whether it is to a local individual or to a distal individual. No use of a name *denotes* via

a rule. Names refer by an outside in(-coming) natural-historical process, a word ferrying-in an object – local or remote – that I refer-back to.

Consequences and extensions

In the remaining discussion, I would like to lay out consequences of the foregoing as well as extensions of the framework to compound expressions.

Consequences

Modal rigidity

53. One feature of proper names and indexical expressions as well as a subclass of descriptions that preoccupied modern semantics is their affording 'rigid designation'. It is often forgotten that "rigid designation" is a *technical* term casting for a relation in terms of the underlying relation 'designation at world w' and demanding the same designation at all worlds w. Intuitive tests for it in modal English – without the invocation of the abstract indices (worlds) and their set theoretic 'domains' – may be given and indeed have been given by Kripke.

54. I will not take us into another journey through the mysteries of rigidity, its many definitions, strong and weak, *de jure* and *de facto*, and umpteen other variations. In my estimation, the Carnap-Kripke casting of the whole semantics in terms of the modal notion of 'designation at w' has been a distraction. It is not that proper names or free variables on an assignment or many engineered definite descriptions – be it on metaphysical or semantical grounds (such as our "the actual F") – are *not* rigid. They are rigid all right. It is rather that this should not have been the key issue at the heart of semantics. A question about an *artifact* of a model theory of a logic – possible worlds – has become a decoy issue. When NN opens and discusses (pp. 28-30) the concerns of Mill, Frege and Russell in the semantics of names, modality is not key. To understand naming (vs. describing), we need not engage in modal metaphysics, let

alone in a reduction of it to what holds across constructed spaces of points, spaces coming from the algebras connected with formal operators.

In my view, we get to the key issues of the proper treatment of ordinary nouns in ordinary English in Lectures II and III of NN. I do not know why Lecture I has attracted all the glamour. Kripke says the fundamental things about names – and all correct in my view – in Lectures II and III.

55. Let us allow rigidity be for a moment and only a moment our focus. I would like to note one important point, rigidity unifies under one basic relation – designation – the very different categories of proper names (indexical pronouns) and compound definite descriptions. Indeed *predicates* with uniqueness built in e.g. "largest number strictly under 10" or "largest divisor of 27" both rigidly designate an extension. But a predicate and a proper name do not function semantically in similar ways.

56. I still insist that proper names have no *denotation/designation* at all and the entity they keep being associated with in all worlds is not a denotation/designation, if ... definite descriptions have a designation (remember Russell). The proper name "Nixon" and the description "the Us president in 1968" do not *co-designate* (co-denote). What is true is that there is a single *individual*, Nixon, that (1) the name relates to it in a certain relation R-1 and (2) the description designates/denotes it. In like way, the name Nixon" and the description "the actual president of the US in 1968" do not *co-rigidly designate*. The description rigidly designates the man Nixon and the name "Nixon" R-1 the man Nixon. I call R-1 "refers". Furthermore, by referring to it, it is it – Nixon – that is the object of all *subsequent* modal predications.

57. Kripke says in footnote 15 of NN that instead of saying "There is a possible world w where Humphrey wins the elections in 1968", it'd be good to say "Humphrey might have won". I totally agree. When you say "Nixon might have lost" or "Humphrey might have won", you are wedded – at the intuitive level if not the technical level of 'designation at w' – to the idea that this actual *referent* is the only thing that matters. Of course it is the one that matters – in the subject-predicate modal predications, where the "might have" acts as an *ad-verb*, a predicate modifier, it is clear that Nixon refers to Nixon (nothing modal here) and of that man – the referent – we ascribe now the *modal feature* – might

have lost. The proper name is outside the syntactical effect of the modal ad-verbs; what could be the question?

58. It is true that there are sentences of English where the modal locution modifies the whole sentence: "Possibly, Nixon loses" (read non-epistemically) or "It might have been the case that: Nixon loses". I find these sentential modifications rather engineered-English and they are of course engineered/modeled after the syntax of sentential operator languages, e.g. those of quantified modal ("Box") logic, in which an operator "Box" modifies a sentence. How are we to understand such sentential modal modifiers? When we started with "Nixon might have lost", we noted that the referent of "Nixon", the man, is ascribed the modal feature. Now, we consider "Possibly (It might have been that): Nixon loses". No matter how you care to give the semantics (and not just the model theory of a corresponding Box operator), I'd observe that the word "Nixon", even though now in the syntactic 'range' of the sentential modifier continues to refer to "Nixon". As I hear it, the truth of "It might have been: ... loses" turns on whether the referent of ... might have lost.

59. So where is the problem? The passion to proceed as in *some* quantified modal languages, by a *uniform* treatment for cases in which we slot descriptions in the ... must be controlled at this point. We must resist the temptation to work as in some of these formal languages precisely because the description "the president of the US in 1968" does not refer but *denotes* Nixon. One may be blinded at this stage by the fact that if I slotted a rigid description "the actual US president in 1968", we would, as with proper names, get the same truth value as if we used a proper name. To the modification of the whole sentence by "possibly" it'd be 'irrelevant' whether we use the name or the rigid description at least at the level of truth value. So it might seem that the key Hamletian issue for the semanticist is: to be rigid or not?

60. On the current perspective the "rigid or not?" question is subsidiary and Russell already saw it. The definite description denotes, the name refers: that was the ur-question for Russell, not rigidity. Of course, Russell went on further to doubt whether the description ... denotes. But when one refers with a name "N", the object loaded is carried and referred to and the *referent* is what is at stake in genuine predications. When the compound expression *denotes* – if we allow it to denote at all

(recall Russell's eliminations) – the story is different. It may be true that in all worlds it is one and the same object that is satisfying the denotation condition and that object affects the truth value. But this issue about truth value is *subsequent* – and post-semantical – to the issue of what each type of expression does and how.

61. Russell thought descriptions should be eliminated from being full terms. Others may believe they do denote, sometimes rigidly if the predicate has the same unique satisfier in all worlds. Yet others – famously Donnellan – think that prior to any modal or other embeddings, the description may be *used* in two different ways: one is *attributive*, wherein the attribute is key (we can – on that use – ask whether one and the same thing satisfies the attribute in all possible worlds). On the other reading – the *referential* reading – the compound predicate "president of US in 1968" – is inconsequential, for no satisfaction of it is involved. Rather, in the manner of a demonstrative, as if I am standing in front of Nixon and of him predicate "He might have lost", with the externalization "the American president in 1968" merely communicating the individual I already have in mind (by perception) and refer back to while using "he". I may, if I had too many drinks and you mis-present to me Nixon under the title "the Canadian president in 1968" use that title to say of him – Nixon – "But surely you (he) might have lost election, you lucky guy". On this use, the attribute in the description-used merely *communicates* (points to) the guy I already am bound to by the natural-historical relation of perception.

62. Be all this as it may with compound descriptions, they obviously raise new issues beyond those of the referential nouns. When a single word name refers, nothing 'deep' needs to be invoked when it comes to modal predications, let alone a whole machinery of possible worlds metaphysics, a semantics built on the metaphysics and what not. Nixon is loaded into the name and "Nixon" re-fers to the man. The man is always the only thing that matters, wherever you place the "possibly".

63. When we move on to descriptions, a different set of issues arises – not whether they are rigidly designating, but how *modality* operates. Modality is now the key. E.g. in the Russell account, the descriptions do not 'stand' for objects period and this much together with a simple treatment of the modal modifier, governing the whole sentence or just the predicate, gives the correct results. If Donnellan is right and there is

a duality of use, we get a duality of modal explanation – in the attributive use, the key is whether the attribute indeed has one and the same extension in all worlds (if worlds is what we use), but in the referential use we are back with non-designation. The compound refers – is loaded with that key entity, the referent – and it is of *it* that we check the modals, ignoring the descriptive predicate altogether.

64. In a nutshell: Quine and others have given a host of 'technical' arguments condemning the expression "possibly" as being a logically odd locution (odd compared to a modifier like "not"). The arguments, for all their authoritative use of 'logic', were confusing a host of issues about those modal locutions. E.g. if one really believed that "possibly" means the same as "is consistent" and "necessarily" means the same as "is provable" and the latter apply to sentences, not to objects and their bearing properties, then sure enough we have a different subject matter and the semantics of names and descriptions in not the issue.

65. But if "possibly" and "necessarily" mean what most of us hear them to mean, then that is what they *mean*. In any event, it is not a question about how the word "Nixon" functions or how the description "the US president in 1968" functions, it is a question about how "might have" functions. The questions about the name and the description are pre-modal and we can distinguish – we just did – how the two categories function. Those words – proper nouns whose sole linguistic function is to *refer* – bring in their referent and nothing but it into the discourse. What the truth value of the whole will turn on depends on other locutions, the embedding locutions such as "Pierre believes that . . . " or "It might have been the case that . . . ".

Indexicality

66. Though often taken to be vivid examples of 'direct reference', indexicals are in effect troublemakers for the idea of directness of reference. Kaplan hoped to model English "that" or "he" by the compound term "Dthat(the F)" or "Dhe(the F)", with the "the F" determining the denotation of the full demonstrative as whatever satisfies "the F" in the actual world. If we so read things, we'd have here another example of a modally rigid designator/denoter whose rule of denotation seals it off from the effect of modal operators. Another reading of Kaplan – offered

by Kaplan himself – is not to over read the model theoretic rules of de-
notation of the representing artifact and concentrate on how the original
words "that" and "he" function in English. Referentially, I would say –
an individual stands in front of me in the party, Frank is he. He is tipsy.
I say: he or this is a tipsy guy. You might say: who? And instead of
pointing, I whisper: the man drinking a martini. On the current account,
the description in the parenthesis in the formal artifact "Dthat(the F)"
is rationalized as my *communicating* (pointing) to you who it is I was *al-
ready* referring to. My use of "he" / "this" was loaded by natural-historical
process with Frank and it is of him that I am thinking (it is him I *came*
to have in mind). And it is him – ferried to me by perception – to whom
I now back re-fer. If, as luck would have it, the description I use to com-
municate to you my already taken-place referring, happens to denote a
man – invisible to me hidden in the kitchen – who is the unique man in
the party drinking martini, this denotation is irrelevant to *my referential
use*.[4]

67. And so it goes for the word "I". "I" refers all right. But it does not
refer *because* of an alleged denotation rule "In all contexts c, 'I' refers
at c to the agent of c". We could of course in the abstract model theory
introduce a term – it does not have to be a single word pronoun – that is
by design harping on that abstract coordinate of the context in the model
theory, "agent of c". That term "$A(c)$" denotes whoever is the agent of
c, provided we make c itself – as Frege urged us in "Der Gedanke" –
part of the *means of expression*. In "$A(c)$", we would have a denoter, for
every instance of the variable c. Using familiar scope-immunizing tricks,
we could make sure that this denoting term still denotes the agent of c
throughout all worlds w.

68. The English word "I" is *not* such a denoter. It refers. It refers
because a natural-historical process starting in full me – the human
being and no disembodied agent – has impacted me-the-user to think of
it (to have it in mind). Again it is that thing, me-JA, that I am thinking
of and referring back to. Wishing to *communicate* to you who it is I am
referring to and predicating as a fool, I say to you "I am a fool". Inasmuch
as I communicate to you that it is the *speaker* (or voice producer etc.)
who is here spoken of, this description of me – the speaker of the context,

[4] See Kaplan 1978.

helps you focus on who is that item, in the way that I can raise my arm inside a dense crowd when the question is asked – who left the coffee machine running? My referring is not 'fixed' or 'determined' by my arm raising or by my being the satisfier of "speaker of c". I refer to myself prior to this by having myself loaded into "I" by the historical process of myself coming to be my focused object. The further *communication* of who has been referred to is accomplished by the arm raising indication it is the speaker on whom you are now to focus.

69. Deictic "he" and indexicals like "I" are thus like *free* variables in one way but not in another, free variables being the formal artifact often said to represent them best. In free variables, *what* is loaded into the symbol – in the standard model theories – is just the object; *what* the device is infused with is nothing but the object. But this does not make the variable a *referring* device. Free variables have no meaning only an object as *value*, true enough. But they do not re-fer for they are not *intra-historical* ferriers of objects from past to future. So I submit – variables do not refer. Variables have values and those values are purely objectual. In contrast, as natural-historical devices, indexicals and deictic "he" re-fer – they allow the user to go back to the object uploaded by an intra-historical episode wherein the object, proximate or distal, *becomes* the subject of predication.

Extensions

70. In the current section I would like to attend to cases beyond the single word proper nouns and indexical pronouns. The successive enlarge-ments involve: other – bound, anaphoric – uses of the pronoun "he", the use of common nouns as in "Tigers are mammals" or "Tigers have been spotted around the train station" and mass nouns as used in "Water is wet". So much concerns simple nouns. Of course, nouns occur in com-pounds, as in the denoting phrases "every tiger" or in the plural "all (most) tigers". One type of enlargement concerns single word nouns oc-curring on their own or modified by determiners. The sequel of this paper dedicated to common nouns is dedicated to these enlargements of the reference all the way up thesis.

71. A second type enlargement concerns all single words in English: the word may be an adjective like "happy", a verb like "runs", a preposition

such as "at" or an adverb like "reluctantly" or even a 'connective' such as "and" or "not". Inherent to the present approach to semantics – as contrasted with denotational model theory with denotation rules – is that all single words re-fer – the word has been loaded in history with its referent (whatever type of entity the referent is) and it is that referent – and only it – that is contributed by the word.

72. A third type of enlargement concerns the full *sentential* level. We said that in "John is happy" "John" refers to John and "happy" to being-happy. We have not said anything about the word "is". But let us ask: what does the full combined unit, the sentence, do? Does it denote? Does it refer?

73. The present framework denies that any such combination denotes. But it also denies that sometime in history the full unit "John is happy" was loaded with its referent – let us say the *proposition* that John is happy – and its been ferried to me, so that when I now use it, I refer back to that proposition.

74. Eventually, the account of the linguistic function of the sentence "John is happy" is more along this latter – referential – line. But the sentence does not simply re-fer as do the nouns or single words. The sentence *combines* these simple referents depending on the syntactic mode of combination. Suppose we preceded naively and let *referential sequences*, sequences of referents reflecting the order of syntactic combination be of use here, thus letting ⟨Ref("John"), Ref("happy")⟩ be 'called upon' by "John is happy". We would get in tow different sequences for "John loves Mary", "Mary loves John" and "John (Mary) is loved by Mary (John)".

75. Of course, even at this naive phase, we run into problems: "John is eager to please" might be combined in a different manner than "John is easy to please", if linguists are to be believed. In like manner, many philosophers have taken "John appears to be approaching" to really be combined in the manner of: "It appears that: John is approaching".

76. Questions of syntactic modes of combination are rather complex. We need to embark on the treacherous path of defending the *appearances*, viz. the surface grammar does indicate well the syntactic mode of combination, including where we have phrasal ambiguity and *two or more* readings. But before we engage in such a demanding syntax of combinations, even in the simplest of all cases such as "John runs" or

"John is happy" the question of how a composed unit refers challenges reflection.

77. Last but not least are compounds such as "queen of England", "man who won the lottery", "happy and rich" "John and Mary", "walked and talked", let alone, "John and every student in my logic class". They sure all can be said to *eventually* denote, be true of an individual or a plurality thereof. But this cannot be the primary semantics. Are the phrases referring to some sort of *referential composition* made of e.g. the referents of "queen" "of" "England", perhaps thusly-referring to an attribute, which then, in turn, has a denotation, viz. all those satisfying at one time or other "be-queen of England"?

78. None of these enlargements can be provided, let alone justified, in a hurry, on one foot, at the end of a paper. Let me only say that on my reading of Russell, he believed (i) negatively – none of these harder cases, harder than say "John", do or could call on *denotation* for their primary semantics, denotation being a post-semantical truth-evaluation relation and (ii) positively – by combining simple referents – such as the referents of "John" and "happy" etc. – we can by syntax – in Russell's case not by English grammar but by *logical* syntax – account for the semantics of all these cases.

79. Speaking globally, all that is needed, from "John" and "happy", all the way up to "The first gentleman of Europe sought someone who wrote the Waverley novels" and "Every natural number has at most finitely many predecessors" or "Some critics admire only one another" are two things: the *referents* of the individual words and a theory of their syntactic mode of combination. In a *motif* form: semantics is reference theory + syntax. I too believe this and I hope to return in future work to some of these challenging 'higher up' cases to vindicate the conjecture.

References

Donnellan, K. (1966), "Reference and definite descriptions", *Philosophical Review*, 75, pp. 281–304, reprinted in his *Essays on Reference, Language, and Mind*, Oxford University Press, Oxford 2012, pp. 3–30.

——— (1974), "Speaking of nothing", *Philosophical Review*, 83, pp. 3–31, reprinted in his *Essays on Reference, Language, and Mind*, Oxford University Press, Oxford 2012, pp. 81–114.

Kaplan, D. (1978), "Dthat", in P. Cole (ed.), *Syntax and Semantics*, volume 9, Academic Press, New York, pp. 221–243.

Kripke, S. (1972), "Naming and necessity", in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Reidel, Dordrecht, pp. 253–355, 763–769, reprinted with a new introduction as *Naming and Necessity*, Blackwell, Oxford 1980.

——— (1977), "Speaker's reference and semantic reference", *Midwest Studies in Philosophy*, 2, pp. 255–276, reprinted in his *Philosophical Troubles: Collected Papers, Volume I*, Oxford University Press, Oxford 2011, pp. 99–124.

Napoli, E. (1995), "(Direct) reference", *Journal of Philosophical Logic*, 24, pp. 321–339.

Putnam, H. (1970), "Is semantics possible?", *Metaphilosophy*, 1, pp. 187–201.

A new account of the referential/attributive distinction and its semantic nature

Antonio Capuano

In "Reference and Definite Descriptions", Keith Donnellan pointed out that in ordinary English definite descriptions function in two distinct ways. They can be used *referentially* or *attributively*. Donnellan illustrated this duality of functions with some examples.

> [C]onsider the sentence, "Smith's murderer is insane". Suppose first that we come upon poor Smith foully murdered. From the brutal manner of the killing and the fact that Smith was the most lovable person in the world, we might exclaim, "Smith's murderer is insane". I will assume, to make it a simpler case, that in a quite ordinary sense we do not know who murdered Smith (though this is not in the end essential to the case). This, I shall say, is an attributive use of the definite description.
>
> The contrast with such a use of the sentence is one of those situations in which we expect and intend our audience to realize whom we have in mind when we speak of Smith's murderer and, most importantly, to know that it is this person about whom we are going to say something.
>
> For example, suppose that Jones has been charged with Smith's murder and has been placed on trial. Imagine that there is a discussion of Jones's odd behavior at his trial. We might sum up our impression of his behavior by saying, "Smith's murderer is insane". If someone asks to whom we are referring, by using this description, the answer here is "Jones". This, I shall say, is a referential use of the definite description. (1966: 8)

Generalizing from this case, Donnellan claimed that sentences of the form "The *F* is *G*" can be used in two different ways.

Nobody disputed that Donnellan's distinction exists in the language. In his extremely influential response to Donnellan's paper, "Speaker's Reference and Semantic Reference", Saul Kripke claimed he had "no

doubt that the distinction Donnellan brings out exists and is of funda-
mental importance" (1977: 102). Unfortunately, less agreement exists
about the nature of the distinction Donnellan made famous. Influenced
by Kripke, some viewed it as a pragmatic distinction. By contrast, others
viewed the distinction to be semantic.[1] As far I as I know, Ernesto Napoli
has not taken an explicit position on the issue and it is not completely
clear to me where he stands. If I have to guess, I suspect he would deny
that Donnellan's distinction is semantic, though I am not sure.[2]

My aim in this paper is to offer a new account of the semantic nature
of Donnellan's distinction. Napoli should be at least to some extent open
to my account, which relies on the difference between reference and deno-
tation. The distinction between reference and denotation can be traced
back at least to Bertrand Russell and Napoli has always emphasized
it as of central importance for the understanding of natural language.[3]
Before offering an account of the semantic nature of Donnellan's distinc-
tion, however, one has to motivate the need for such an account, i.e. that
indeed Donnellan's distinction is semantic and that Kripke's apparatus
of semantic reference and speaker's reference has not settled the issue
once and for all.

Since Kripke's seminal paper set the stage for any subsequent discus-
sion of Donnellan's distinction, I'll start by discussing it.

[1] Stephen Neale (2004: 69–70) mentions Simon Blackburn, William Blackburn,
Hector-Neri Castañeda, Donald Davidson, Martin Davies, Gareth Evans, Peter Geach,
Paul Grice, Stuart Hampshire, Saul Kripke, Stephen Neale, Mark Sainsbury, Nathan
Salmon, John Searle, Scott Soames and David Wiggins among those convinced that
Donnellan's is a pragmatic distinction. Joseph Almog, Jon Barwise, Anne Bezuiden-
hout, Robyn Carston, Michael Devitt, Keith Donnellan, Jennifer Hornsby, David
Kaplan, David Lewis, Chris Peacocke, John Perry, François Recanati, Marga Reimer,
Bede Rundle, Stephen Schiffer, Robert Stalnaker, Howard Wettstein believe that
Donnellan's is a semantic distinction.

[2] In "Riferimento diretto" he wrote: "To a referential use of a definite description
is associated a speaker's reference (the individual the speaker intends to refer). To
the attributive use is associated a semantic referent (where there is one)" (1992: 403,
translation mine). This suggests that like Kripke Napoli believes that the distinction
Donnellan introduced is a pragmatic distinction.

[3] See Napoli 1995 and Bianchi and Napoli 2004: 221–225.

1. The semantic nature of the referential/attributive distinction

In "Reference and Definite Descriptions" Donnellan tentatively conjectured that sentences containing definite descriptions in subject position are pragmatically ambiguous. Donnellan did not find it plausible to account for the difference between the two uses of definite descriptions as

> an ambiguity in the sentence. The grammatical structure of the sentence seems ... to be the same whether the description is used referentially or attributively: that is, it is not syntactically ambiguous. Nor does it seem at all attractive to suppose an ambiguity in the meaning of the words; it does not appear to be semantically ambiguous. (1966: 20–21)

Donnellan's cautious remark naturally invited Kripke's response. If definite descriptions are pragmatically ambiguous then Donnellan did not refute Russell's theory of definite descriptions. He did not, simply because Russell's aim was to offer a semantics for sentences containing definite descriptions. To point out that such sentences are pragmatically ambiguous is not to contradict Russell's theory. Therefore, Donnellan could not be right when he claimed that simple sentences containing a definite description in subject position are not semantically ambiguous but that he refuted Russell's theory.[4]

Like Kripke, I feel that "such phrases as 'semantic ambiguity', 'pragmatic ambiguity', 'syntactic ambiguity', and so on have partly ... the air of slogans until they are fleshed out" (Kripke 2013: 109). Unfortunately, most of the discussion that followed Donnellan's paper and Kripke's response has focused on whether sentences containing definite descriptions used referentially express or not a singular proposition. For instance, introducing the debate between Howard Wettstein and Nathan Salmon about the semantic significance of Donnellan's distinction Marga Reimer wrote:

---

[4] At the same time, later on Donnellan made clear that he meant the referential/attributive distinction as a distinction relevant to semantics. See especially Donnellan 1978. Unfortunately, this is not yet to say that the distinction itself is a semantic distinction.

> For the past 30 years, philosophers of language have been debating over whether or not Keith Donnellan's referential/attributive distinction has 'semantic significance'. To claim that the distinction has semantic significance is to claim that the proposition expressed by a referential utterance of a sentence of the form *The F is G* differs in kind from the proposition expressed by an attributive utterance of a sentence of the same form. Generally, those arguing on behalf of such a distinction have claimed that, while referential utterances of such sentences express *singular* propositions, attributive utterances of such sentences express *general* propositions. And, generally, those arguing against such a distinction have claimed that both sorts of utterances express *general* propositions. (1998: 130)

To claim that making the referential/attributive distinction semantic amounts to arguing that sentences containing a definite description can express either a singular or a general proposition is, I believe, a mistake. On the one hand, those who have argued, on behalf of Donnellan, that sentences containing a definite description used referentially express singular rather than general propositions have often made Donnellan's position hardly distinguishable from that of Peter Strawson in "On Referring" (1950). In doing so, in my view, they have missed what is radical and innovative in Donnellan's paper. On the other hand, arguing in terms of singular as opposed to general propositions forces one to cast the issue in a particular semantic framework, one which might ultimately not be the correct theoretical framework for studying natural languages.[5] In fact, in his paper, Kripke did not mention the notion of singular or general proposition. Instead he asked what truth conditions are assigned to sentences containing a definite description.[6] If one can make the case that sentences containing a definite description have different truth conditions depending on whether the definite description is used attributively or referentially then one has made the case that the referential/attributive distinction is semantic and that Russell's theory is at least incomplete, since it does not account for one use of definite descriptions.

---

[5] Inspired by Donnellan's observations about definite descriptions and empty names and David Kaplan's work on indexicals, Joseph Almog (2004) argues that the proper form of a semantics for a natural language cannot be given in terms of propositions. Kripke himself was no fond of propositions. See Kripke 1980: 20–21.

[6] Kripke (1977: 121 fn.29) voiced some uneasiness even about the concept of truth conditions.

To establish whether sentences containing definite descriptions have different truth conditions depending on whether a definite description is used attributively or referentially, Kripke suggested the following test:

> If someone alleges that a certain linguistic phenomenon in English is a counterexample to a given analysis, consider a hypothetical language which (as much as possible) is like English except that the analysis is *stipulated* to be correct. Imagine such a hypothetical language introduced into a community and spoken by it. *If the phenomenon in question would still arise in a community that spoke such a hypothetical language (which may not be English), then the fact that it arises in English cannot disprove the hypothesis that the analysis is correct for English.* (1977: 113)

Following Kripke, imagine a hypothetical community that speaks

> a language similar to English except that the truth conditions of sentences with definite descriptions are *stipulated* to coincide with Russell's: for example, "The present King of France is bald" is to be true iff exactly one person is king of France, and that person is bald. (113)

Let's call such a language "the weak Russell language".[7] Kripke's question is whether Donnellan's phenomenon would arise in a community that speaks such a hypothetical language. His answer is that it would. From this he concludes that the fact that Donnellan's phenomenon arises in English is no evidence against Russell's analysis of definite descriptions.

Why does Kripke think that Donnellan's phenomenon would arise in communities that speak the weak Russell language?

> Surely speakers of these languages are no more infallible than we. They too will find themselves at a party and mistakenly think someone is drinking champagne even though he is actually drinking sparkling water. If they are speakers of the weak ... Russell [language], they will say, "The man in the corner drinking champagne is happy tonight". They will say this precisely because *they think, though erroneously, that the Russellian truth conditions are satisfied.* (114)

It seems right to me that someone can be mistaken and be under the misimpression that the person in front of him is drinking champagne

---

[7] Kripke (1977) considered also what he called "the intermediate Russell language" and "the strong Russell language". Since, as far as I can see, nothing I say in the paper bears on the differences between the three languages, I ignore the intermediate Russell language and the strong Russell language.

while he is drinking water and use "the man drinking champagne" to talk about the person in front of him. This is not, however, to show that Donnellan's phenomenon arises in a community of speakers that speak such a weak Russell language. It is not, simply because the nature of Donnellan's phenomenon, as Donnellan understood it, had little to do with the beliefs of the speaker using a definite description referentially. What was crucial for Donnellan was that

> a speaker who uses a definite description referentially in an assertion ... uses the description to enable his audience to pick out whom or what he is talking about and states something about that person or thing. (1966: 7)

In doing so the speaker does not have to believe that the person or thing one talks about satisfies the description the speaker is using. This is illustrated by Donnellan's following example:

> Suppose the throne is occupied by a man I firmly believe to be not the king, but a usurper. Imagine also that his followers as firmly believe that he is the king. Suppose I wish to see this man. I might say to his minions, "Is the king in his countinghouse?" I succeed in referring to the man I wish to refer to without myself believing that he fits the description. (14)

Of course, Kripke was aware of Donnellan's example. However, he considered it "of a somewhat exceptional kind" (1977: 112 fn.21). I am not sure what Kripke thinks is so exceptional about the case. People refer all the time to Elvis Presley as "the King", presumably without believing that he is a king, let alone *the* King. Of course, one might object that in this case "the King" is used as a name rather than a definite description. But suppose one is an undercover agent in a dangerous fanatic terrorist cult whose members refer to their leader as "the Messiah". The agent needs to be accepted by the members of the group and gain their trust to collect information. The agent does not believe the leader of the group is the Messiah and still refers to him when using "the Messiah". Other similar cases can be imagined. Once a student asked a colleague of mine whether he could quote from the most historically accurate document of all times, i.e. the Bible. My colleague replied: "The most historically accurate document of all times, as you call it, has some contradictions in it. How could it be the most historically accurate document of all times then?". He failed to change the student's mind but this is not because the student could not understand what he meant in English. Does the

existence of such cases show that in a community of speakers speaking the weak Russell language Donnellan's phenomenon would not arise? Imagine the following dialogue:

(1)   Peter: The brilliant student we met yesterday at the coffee shop is not a student. She is a professor.
Charlie: Is that true?
Peter: Yes, she is so young that I never would have suspected it.

To me the dialogue appears to be proper English. It is unclear, however, how to make sense of it if it occurred between speakers speaking the weak Russell language. To begin with, Peter does not believe that the person he wishes to talk about is a student. It cannot be then that he uses "the brilliant student we met yesterday at the coffee shop" because he thinks, though erroneously, that the Russell truth conditions are satisfied. In fact, if he is a competent speaker, he cannot believe that the Russell truth conditions for the sentence can be satisfied and cannot believe that Charlie believes it either. Presumably, Kripke could refine his definition of speaker's reference to take care of this problem. But, how would he explain that Charlie replies to Peter's assertion with "Is that true?". In the weak Russell language Charlie could not have asked the question he asked because in the weak Russell language Peter could not possibly have said something true and could not possibly have thought to have uttered something true. Last, Peter could not have replied that yes it is true that the brilliant student they met yesterday at the coffee shop is not a student but a professor because he would have asserted that a contradiction is true. I would like to point out that if a definite description is used attributively, English and the weak Russell language are alike in making "The President of the United States, whoever she is, is not the President of the United States" nonsense. It would make even less sense for someone to reply "Is that true?".

Perhaps someone sympathetic with Kripke's line of argument could reply that if I am willing to use the dialogue to argue that English is not a weak Russell language, by the same token, I should also be willing to admit that English is not a Kripke language either, i.e. it does not satisfy the schema "The semantic referent of 'X' is X where 'X' can be replaced by a name".[8] Imagine that the CIA substituted the legitimate

---

[8] Andrea Bianchi suggested to me the reply together with the example.

President of the US with a lookalike person whose name is not "Barack Obama" but that people, being under the impression that the person that they see delivering speeches on TV is Barack Obama, call "Barack Obama". Peter finds out about the CIA's operation but, afraid of the consequences if he tells anybody, continues to call who appears to be the President of the US "Barack Obama". One time, he finds himself alone in a room with Charlie, who Peter trusts. Sure of not being heard by anybody else they have the following dialogue:

(2)   Charlie: I do not understand what's going on. Barack Obama is pushing for legislation that contradicts the platform that brought him to win the Presidential election.
       Peter: Barack Obama is not Barack Obama.
       Charlie: Is that true?
       Peter: Yes, the CIA kidnapped him to substitute him with a lookalike person. You cannot speak with anybody about this. If the CIA finds out we know what really happened, we are dead.

If the first dialogue shows that English is not a weak Russell language the second dialogue should show that English is not a Kripke language. But the second dialogue does not show that. There is no denying that, in some special circumstances, people can *use* the name "Barack Obama" to speak of someone who is not Barack Obama. This alone does not show that the name itself stands for the lookalike person. When someone sees Smith thinking he is Jones and ask "What is Jones doing?", in a sense she is referring to Smith even though, Kripke argued, the name "Jones" continues to be a name of Jones and to be *used* as a name of Jones in English.[9] When Charlie asks Peter whether it is true that Barack Obama is not Barack Obama obviously he is not asking whether a patent falsehood is true. Assuming that Peter is being truthful, he realizes that he is trying to pragmatically convey the information that the individual people are calling "Barack Obama" is not the real Barack Obama.

As is well known, one's *modus ponens* is another's *modus tollens*. If Kripke would argue on the basis of (2) that (1) does not show that English is not a weak Russell language, I believe that Donnellan would accept that English is not a Kripke language in the sense specified above.

---

[9] This is Kripke's famous example of Smith-Jones. See Kripke 1980: 25 fn.3.

That is, Donnellan would deny that the schema "The semantic referent of 'X' is X where 'X' can be replaced by a name" holds no matter what the speaker using "X" has in mind. Although their views have been often assimilated, in fact, Kripke and Donnellan differ on proper names.[10] I cannot defend it here but my inclination is to believe that Donnellan is right to deny the schema as Kripke understands it.[11] Therefore, to point out that if (1) shows that English is not a weak Russell language then (2) shows that English is not a Kripke language is not to create an embarrassment for a position *à la* Donnellan.

However, one should still account for the fact that some proper English dialogues can be easily explained under the hypothesis that English is a weak Russell language supplemented with the distinction between semantic reference and speaker's reference. However, the same dialogues are not as easily explained if one instead assumes that Donnellan's is a semantic distinction. Consider the following dialogue imagined by Kripke (1977: 121):

(3)   A. Her husband is kind to her.
        B. No, he isn't. The man you're referring to isn't her husband.

In the dialogue B uses "he" to refer to the unique individual that is married to her, what Kripke called "the semantic reference". According to him, under the hypothesis that the referential/attributive distinction is part of the semantics of English, the dialogue would be "perhaps impossible to explain" (122). By hypothesis, A was using

> "her husband" in the referential sense. Both the speaker's referent and the semantic referent would be the kind lover; only if B had misunderstood A's use as attributive could he have used "he" to refer to the husband, but such a misunderstanding is excluded by the second part of B's utterance. (122)

B then cannot use "he" to hark back to the unique individual that is married to her. By hypothesis the semantic referent of "her husband" is her lover.

Kripke thought that (3) is perhaps impossible to explain under the hypothesis that English is ambiguous between a referential and an attributive reading of "her husband". However, if English was the weak

---

[10] For a detailed reconstruction of their differences see Bianchi and Bonanini 2014.

[11] On this point see Almog et al. 2015: 374–376.

Russell language and the distinction between semantic and speaker's reference is admitted, as it should be since we have independent reasons to have it as part of our theory of language, one could easily handle the dialogue by saying, as Kripke did, "that pronominalization can pick up *either* a previous semantic reference or a previous speaker's reference" (121). Kripke could not think of any plausible move to handle the case under the hypothesis that "her husband" is ambiguous between two different readings. However, one move is simply to deny that "he", as used by B, is anaphoric. How plausible is this move? On some views of pronouns

> [t]here is ... no syntactic distinction that marks a pronoun as anaphoric, and the process of determining whether or not a pronoun is anaphoric is precisely the process of identifying the referent. Therefore a semantic rule that applies to anaphoric pronouns could not figure essentially in their interpretation. (Almog et al. 2015: 363)

If A had said "Her husband is kind to her" and B had simply replied "No, he isn't", the most natural reading would have been that A and B disagree about the kindness of the person in question. At the same time, B is aware that another reading is possible; it could be that "he" as used by A and B refer to two different individuals and after all A and B do not necessarily disagree about the kindness of the person A is referring to. B must be aware of it because this is the reading he intends. By adding "The man you are referring to is not her husband", B clarifies that the correct reading is the second one and his use of "he" is not anaphoric. It is true, as Kripke said, that B could not have misunderstood A's use of "her husband" to be attributive. But B is aware that A could have taken him to refer to the same individual A refers to. B exploits the ambiguity between a referential and an attributive reading of "her husband" and uses "he" to talk about the individual that would have been denoted if A had used "her husband" attributively. Consider the following dialogue:

(4)   A: The bank is my favorite place.
       B: It is not mine. But the place you're referring to is not a financial institution.

Nobody doubts that "bank" is ambiguous. B does not misunderstand A's use of it, but exploits the ambiguity of "bank". B clarifies that "it" is not anaphoric by adding that he is not referring to the same place A

is referring to. The same happens in (3). Since nobody doubts that in (4) "bank" is ambiguous and (3) and (4) look alike, we have some room to argue in favor of the hypothesis that, like "bank", "her husband" is also ambiguous.

In light of such considerations I think we have some reason to resist adopting a pragmatic explanation of the referential/attributive distinction and to take seriously the possibility that its nature is semantic.[12] Once we properly understand Donnellan's phenomenon for what it is, it is unclear that it arises in the weak Russell language and that therefore English can be analyzed as one analyzes the weak Russell language.

I am quite confident that with some philosophical ingenuity one can find a way to reformulate the notion of speaker's reference in a way to accommodate the above dialogue and hold the view that Donnellan's phenomenon is pragmatic rather than semantic after all. The question is not whether this can be done but whether this should be done. Kripke himself, who more than anyone else paved the way to treat Donnellan's distinction as pragmatic, warned us about the danger of using pragmatics as a "wastebasket":

> [U]nder the influence of Grice we have sometimes gone too far. Let's make the semantics as simple as possible, the rest is all pragmatics and we don't have to think about it. But, sometimes we have to think about it, and then maybe the phenomenon should be treated as part of the semantics of the language. Pragmatics is not a wastebasket. (2011: 342–343)

If pragmatics cannot be used as a wastebasket we perhaps should start looking at a semantics for English that can account for Donnellan's phenomenon no matter how simple and elegant an account that relegates it to pragmatics looks like.[13]

---

[12] Perhaps to those in the paper I should add the point, recently emphasized by Michael Devitt (2004) and Felipe Amaral, that "definite descriptions are *regularly* used referentially" (2008: 288).

[13] It is unclear to me to what extent Kripke disagrees: "Any conclusions about Russell's views *per se*, or Donnellan's, must be tentative. If I were to be asked for a tentative stab about Russell, I would say that although his theory does a far better job of handling ordinary discourse than many have thought, and although many popular arguments against it are inconclusive, probably *it ultimately fails* ... It may even be the case that *a true picture will resemble various aspects of Donnellan's in important respects* ... I will state my preference for one substantive conclusion

2. The problem

Any treatment that wishes to maintain that Donnellan's distinction is semantic needs to postulate a semantic ambiguity in definite descriptions. If definite descriptions are ambiguous, then a sentence can have different readings and different truth conditions depending on the semantic function of the definite description. But what do the different readings depend on? Influenced by Kripke, most of the discussion has assumed that there are two ways for a sentence to have different readings. Either a sentence is structurally ambiguous – like "Every man loves a woman" or "The number of planets is necessarily odd" – or some of its constituents are lexically ambiguous.

Some believe that the referential/attributive distinction arises from an ambiguity of scope, namely a structural ambiguity. Kripke mentioned Lauri Kartunnen as advocating "the view that the referential-attributive distinction arises from a scope ambiguity" (Kripke 1977: 123 fn. 34). However, as Kripke pointed out,

> [s]ince the refential-attributive "ambiguity" arises even in simple sentences such as "Smith's murderer is insane", where there appears to be no room for scope ambiguity, such a view seems forced to rely on acceptance of Ross's suggestion that all English assertive utterances begin with an initial "I say that", which is suppressed in "surface structure" but present in "deep structure". (123: fn. 34)

In such cases there are two possible readings of the sentence. In one, "Smith's murderer is insane" is read as: I say that Smith's murderer is insane. This reading would correspond to the attributive use. The other reading, instead, would correspond to the referential use. "Smith's murderer is insane" is read as: Smith's murderer is such that I say that he is insane.

However, in both readings – the one in which the definite description takes small scope and the one in which it takes wide scope – the definite description is used. One can still raise the question of whether it is used attributively or referentially. If so, the referential/attributive distinction

---

(*although I do not feel completely confident of it either*): that unitary theories, like Russell's, are preferable to theories that postulate an ambiguity" (1977: 100, italics mine).

cannot correspond to a scope distinction.[14] But suppose that no hidden operator is present, as Donnellan is inclined to think. In this case no scope ambiguity can explain the two readings of a sentence containing a definite description. However, if the referential/attributive distinction matters for semantics, then there must be two different possible readings.

The other option at hand is that some of the words in such a sentence – in particular, the definite article "the" – are lexically ambiguous.[15] Two meanings are associated with it by linguistic convention. When a definite description is used referentially it has one of these meanings. When it is used attributively it has the other. If that were the case, the definite article would be like the English word "bank", which is ambiguous between two different readings, meaning either the financial institution or the edge of a river. To assess whether "Jones went to the bank" is true or not, one must know whether "bank" is used to indicate the financial institution or the edge of the river. It might be true that Jones went to the bank to withdraw money and false that Jones went to the bank of the river. Analogously, to assess whether "The man drinking a martini is tipsy" is true or false, one must first assess what is the sense associated with "the".

As Kripke emphasized, it is very implausible that the definite article is like the English word "bank". We all know that "bank" is ambiguous. Because of that, we expect the ambiguity of "bank" to be disambiguated by separated and unrelated words in some other language. Do we expect the same for the definite article? Probably not. But if we do not expect that, it is probably because "our linguistic intuitions are really intuitions of a unitary concept, rather than of a word that expresses two distinct and unrelated senses" (Kripke 1977: 119). Of course,

> we can also ask empirically whether languages are in fact found that contain distinct words expressing the allegedly distinct senses. If no such language is found, once again this is evidence that a unitary account of the word or phrase in question should be sought. (119)

To give just one example, Italian disambiguates the word "bank" but does not disambiguate the definite article "the". When the financial in-

---

[14] See also Kripke 1977: 103–104 for criticisms.

[15] Neale takes the view that definite descriptions are semantically ambiguous between Russellian and referential interpretations to mean that "the definite article is lexically ambiguous" (1990: 65).

stitution is meant, "bank" is translated as "banca". When the edge of
a river is meant, "bank" is translated as "sponda". However "the man
drinking a martini" is translated into Italian as "l'uomo che beve un
martini" whether or not the definite description is used attributively or
referentially.[16]

It seems that whoever wishes to hold that the referential/attributive
distinction is semantic is in a kind of a quandary. To claim that the ref-
erential/attributive distinction is semantic, one has to claim that a sen-
tence containing a definite description used referentially has a different
reading from a sentence containing a definite description used attribu-
tively. However, one cannot say what the source of the two readings is.
The options at hand both seem wrong:

a.   the sentence is structurally ambiguous;
b.   the definite article is lexically ambiguous.

But no other option seems to be left.

My intuitions about the word "the" do not differ from those of Kripke
or, I suspect, those of Donnellan. I would not expect that in some other
language the definite article would be translated with two separate and
unrelated words, one corresponding to the referential "the" and the other
corresponding to the attributive "the". Nonetheless, I believe that def-
inite descriptions can give rise to two different readings of one and the
same sentence. It is just that the two readings depend on something
other than a structural or a lexical ambiguity. It is quite surprising that
this source of the two readings has gone unnoticed. In fact, the semantic
ambiguity of definite descriptions has its roots in a distinction which is
at the very heart of semantics.

---

[16] Amaral observes that "if the alleged ambiguity of 'the' were homonymy, it would
be remarkable that 'the' were translated into likewise ambiguous articles in many
other languages – just like it would be remarkable, for example, if 'bass' turned out
to be translated into single graphs meaning ⟨perch⟩ and ⟨deep sounding male voice⟩ in
many other languages. Yet, there is no good reason for the Ambiguity view to allege
that 'the' is homonymous ... Since no one has provided an argument showing that
'the' is not polysemous, no one has shown ... that it is remarkable or coincidental that
'the' should be translated into likewise ambiguous articles in many other languages"
(2008: 291–292). I find Amaral's observation interesting. In this paper, however, I
am trying to show that even if one grants that the definite article is not lexically
ambiguous one can still account for the semantic nature of the referential/attributive
distinction.

3. Reference as opposed to denotation

The distinction I intend to draw upon to account for the semantic ambiguity of sentences containing definite descriptions was already familiar to Russell. He clearly distinguished between reference and denotation and he clearly thought of it as a semantic distinction. According to Russell, in fact, one cannot do without the distinction. On the one hand, logically speaking, reference is more fundamental than denotation in that there could not be any denotation if there was no reference. On the other hand, if there is reference there must also be denotation, given speakers' ability of building complex expressions out of simple ones. Why is reference indispensable and denotation inevitable for Russell?

> This characteristic, that you understand a proposition through the understanding of its component words, is absent from the component words when those words express something simple. Take the word 'red', for example, and suppose – as one always has to do – that 'red' stands for a particular shade of colour. You will pardon the assumption, but one never can get on otherwise. You cannot understand the meaning of the word 'red' except through seeing red things. There is no other way in which it can be done. It is no use to learn languages, or to look up dictionaries. None of these things will help you to understand the meaning of the word 'red' ... [I]t is clear that if you define 'red' as 'The colour with the greatest wave-length', you are not giving the actual meaning of the word at all; you are simply giving a true description, which is quite a different thing, and the propositions which result are different propositions from those in which the word 'red' occurs. In that sense the word 'red' cannot be defined, though in the sense in which a correct description constitutes a definition it can be defined. In the sense of analysis you cannot 'define' red. That is how it is that dictionaries are able to get on, because a dictionary professes to define all words in the language by means of words in the language, and therefore it is clear that a dictionary must be guilty of a vicious circle somewhere, but it manages it by means of correct descriptions. (1918–19: 194–195)

I take Russell to be saying:

1. Complex linguistic expressions are made of simple expressions;
2. Complex linguistic expressions have a meaning only if the simple expressions that make them up have a meaning;
3. Simple expressions must have a meaning that is not derived from the meaning of other expressions but from standing for certain worldly entities. Otherwise one would run sooner or later in a circle;

4.   Once there are meaningful simple linguistic expressions, speakers
      can combine them to obtain complex linguistic expressions whose
      meaning depends on the meaning of the simple linguistic expres-
      sions.

For Russell that there must be two distinct semantic relations is really
a logical point. This alone does not imply that the language contains
linguistic expressions that refer to particulars, what Russell calls "logi-
cally proper names". For instance, "in the logical language set forth in
Principia Mathematica there are not any names, because there we are
not interested in particular particulars but only in general particulars"
(201). Reference must exist under the assumption that a description of
the world cannot be given in purely general terms but must mention
some particulars. If this is the case then linguistic expressions that refer
to particulars must exist and one cannot do without reference.

Although there can be no doubt that Donnellan saw a difference be-
tween reference and denotation, the logical point Russell brings home is
not one of Donnellan's points. However, something else that Russell says
about reference is relevant to Donnellan's point of view.

According to Russell, if simple expressions cannot be defined but their
semantic function is simply to stand for worldly entities then, as long as
the language has expressions that stand for particulars, i.e. it has refer-
ring expressions, their semantic function cannot rest on the satisfaction
of some predicate. In short, reference is true predication free. There can-
not be any doubts in fact that the mark of a referring use of a definite
description is that a speaker can use it to refer to an object that is not
correctly described.[17] Consider Donnellan's case of the man drinking a
martini.

> Suppose one is at a party and, seeing an interesting-looking person holding a
> martini glass, one asks, "Who is the man drinking a martini?" If it should turn
> out that there is only water in the glass, one has nevertheless asked a question
> about a particular person, a question that it is possible for someone to answer.
> (Donnellan 1966: 9)

---

[17] Some disagree. For instance Howard Wettstein writes: "The thesis that the
referential-attributive distinction has semantic significance should be distinguished
from another controversial claim that Donnellan makes in the course of the paper.
These two claims have proved difficult to keep apart. And in informal discussions
and even in the literature, 'Donnellan's distinction' is sometimes understood as if
it were just this second idea ... I have doubts about the correctness of this second

Why, for Donnellan, has one asked a question about a particular person even if he is not drinking a martini? This is because by the time the speaker comes to use the definite description "the man drinking a martini" she was already bound to the object by having it in mind. To some, the notion of having something in mind looks mysterious and unintelligible. To be sure, in ordinary English, locutions such as "Who do you have in mind?" sound perfectly intelligible and little mysterious. Of course, one does not have to think that one has something in her mind in the same sense in which a desk is in the room. In ordinary English, we also say that one has the sun in her eyes and ordinary speakers seem not to have problems understanding what is meant. In my view, to have someone in mind is like having the sun in one's eyes. Of course, this in no way counts as an analysis of the notion of having an object in mind. I am afraid this is not the place to attempt one.[18] This is just to say that the notion of having an object in mind does not have to be hopelessly unintelligible as some have felt it is.

That reference rests on a cognitive relation is a view that Donnellan share with Russell, who believes that the semantic function of a proper name essentially depends on the cognitive state of mind of the user in a particular circumstance of use:

> Suppose some statement made about Bismarck. Assuming that there is such a thing as direct acquaintance with oneself, Bismarck himself might have used his name directly to designate the particular person with whom he was acquainted. In this case, if he made a judgement about himself, he himself might be a

---

thesis, and even about how central Donnellan took it to be. In 'Reference and Definite Descriptions', Donnellan does not mention the idea that intention trumps literal meaning until well into the paper. He first explains the two uses in a very plain way: nothing about reference in the absence of the description's literal application. Only later does he add that the distinction can be usefully brought out by examples in which the description fails to apply literally. This makes it tempting to suppose that it's not the core idea" (2012: 96–97). Like Genoveva Marti, I would argue that "if a description can be used as a genuine device of reference, it must be possible to use it to refer to an object whether the object satisfies the attribute mentioned in the description or not" (2008: 46). In this sense, it is a core idea of Donnellan that definite descriptions can be used to talk about individuals even when the individual does not satisfies the attributes mentioned in the description.

[18] More on the idea of having an object in mind can be found in Almog (2008, 2012 and 2014: 61–86). Kaplan (2012) is also on the notion of having an object in mind but has a different take on it.

constituent of the judgment. Here the proper name has the direct use which it always wishes to have, as simply standing for a certain object, and not for a description of the object. But if a person who knew Bismarck made a judgment about him the case is different. What this person was acquainted with were certain sense-data which he connected (rightly, we suppose) with Bismarck's body. (1912: 37)

In such a case, the person could not use the name "Bismarck" as a genuine name but only as abbreviating a definite description denoting Bismarck, for instance "the first Chancellor of the German Empire". As Kaplan noted:

no one has claimed that Russell's theory of proper names is not a semantic theory, despite the fact that there are both descriptive and referential uses of the same name and that the use is dependent upon the epistemic state and intentions of the user. (2012: 165 fn. 3)

I hope that in section 2 I made at least a *prima facie* case for considering Donnellan's distinction a semantic one. Now, Kripke himself has to agree with Donnellan that referring is not the same as denoting. He claimed that, for all that Donnellan showed, Russell might be right about definite descriptions. However, Russell's analysis of definite descriptions presupposes a distinction between two distinct semantic relations. One of them must be reference unless one believes that it is possible to have a language of pure predicates not involving proper names. Kripke knows of

no convincing reasons to suppose that such a level can be reached in any plausible way, or that the properties can continue to be uniquely identifying if one attempts to eliminate all names and related devices. (1979: 150)

Thus, Kripke has to recognize that there are two different semantic relations, reference and denotation.[19] However, if there are two semantic

---

[19] I think that Kripke would recognize that reference is distinct from denotation. At the same time, in *Naming and Necessity* he had a tendency to underplay the distinction between reference and denotation and to emphasize that between rigid and non-rigid designators. "Some philosophers have thought that descriptions, in English, are ambiguous, that sometimes they non-rigidly designate, in each world, the object (if any) satisfying the description, while sometimes they *rigidly* designate the object actually satisfying the description" (1980: 59 fn. 22). This has been, I believe, unfortunate, although it is quite natural if one considers rigidity the main feature of proper names as opposed to definite descriptions.

relations then there might be, at least in principle, linguistic expressions that can stand in both relations with some entity. These hypothetical expressions are such that in some contexts, they would refer to an entity, and in others they would denote an entity. Finally, the fact that, according to Donnellan, reference rests on a cognitive relation should not disqualify his view as a semantic view unless this disqualifies Russell's view on proper names as well.

Still, I have not offered different sets of truth conditions for sentences in which definite descriptions are used referentially and sentences in which definite descriptions are used attributively. This is what I aim to do in the next section.

4. Two sets of truth conditions

Donnellan pointed out that if a speaker uses "the man drinking a martini" referentially, the speaker may succeed in referring to the man he has in mind and saying something true about him even if the man the speaker is referring to is drinking water. When one uses the definite description referentially, the problem is not that one cannot speak the truth. Rather "the difficulty is in the notion of a 'statement'" (Donnellan 1966: 24). What is the difficulty? Almog explains it quite clearly:

> There seems to be no systematic rule-governed way of associating with an arbitrary referential use of 'The man drinking the martini is G' the statement (proposition, content, etc.) it expresses. On different uses, different objects would be referred to (at my next party, I may well pick out Jones, who will be in front of me, with a martini glass, etc.). Thus the situation here seems even more worrisome than with indexicals. Consider the sentence 'I am sad'. We may wonder what is *the* proposition it expresses. The answer is: there is no such unique proposition. Different uses will determine different propositions. Nonetheless, there is a rule-governed determination of the pertinent proposition. We may write with David Kaplan:
>
> > For any context c, 'I am sad' is true at c iff the agent of c is sad at c.
>
> Donnellan observes that no such rule, conventionally associated with the lexical meaning of the quoted singular term, seems forthcoming for referential uses of descriptions. (2004: 396)

The problem of giving two sets of truth conditions for sentences containing definite descriptions goes beyond the fact that there do not seem to be two different meanings associated with the definite article. It seems that when it is used referentially the meaning of the definite description does not determine the truth conditions of the sentence. This seems to follow from the fact that the definite description is used to refer. If it refers, the object it stands for is not determined by the fact that the object satisfies the property associated with the definite description. Whom the use of the definite description refers to rests on who the speaker has in mind. This is why what matters to the truth conditions of a sentence in which the definite description is used referentially is who the speaker has in mind and whether, as a matter of fact, the object referred to by the speaker has or lacks the property the predicate stands for. Thus, the truth conditions of a sentence containing a definite description used referentially read as follows:

> On a referential use, "The man drinking a martini is happy" is true if and only if the individual the speaker has in mind is happy.

If the nature of reference enters into the truth conditions of a sentence in which a definite description is used referentially, it is only natural to expect that the nature of denotation enters into the truth conditions of a sentence in which the definite description is used attributively. Unlike reference, denotation essentially depends on true predication. This essential feature shows up in the truth conditions of a sentence in which a definite description is used attributively:

> In an attributive use, "The man drinking a martini is happy" is true if and only if the unique individual to whom the description applies is happy.

Clearly, the truth conditions of a sentence containing a definite description used referentially are different from the truth conditions of a sentence containing a definite description used attributively. This is obtained without postulating a structural ambiguity in the sentence or a lexical ambiguity in the definite article "the". Kripke wrote that

if the truth conditions supplied by Russell would apply in some cases, and some other set of truth conditions in another case, then that is as strong a case for semantic ambiguity as I can imagine to hold. (2013: 111–112)

If I succeeded in arguing for the semantic nature of Donnellan's distinction and in offering different truth conditions for sentences containing definite descriptions – admittedly two big ifs – then I made as strong a case for semantic ambiguity as one can image to hold. Let me hope that mine is at least a step in the right direction.[20]

References

Almog, J. (2004), "The proper form of semantics", in M. Reimer and A. Bezuidenhout (eds.), *Descriptions and Beyond*, Oxford University Press, Oxford, pp. 391–419.

——— (2008), *Cogito*, Oxford University Press, Oxford.

——— (2012), "Referential uses and the foundations of direct reference", in J. Almog and P. Leonardi (eds.), *Having in Mind*, Oxford University Press, Oxford, pp. 176–184.

——— (2014), *Referential Mechanics*, Oxford University Press, Oxford.

Almog, J., Nichols, P. and Pepp, J. (2015), "A unified treatment of (pro-)nominals in ordinary English", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 350–383.

Amaral, F. (2008), "Definite descriptions are ambiguous", *Analysis*, 68, pp. 288–297.

Bianchi, A. and Bonanini, A. (2014), "Is there room for reference borrowing in Donnellan's historical explanation theory?", *Linguistics and Philosophy*, 37, pp. 175–203.

Bianchi, A. and Napoli, E. (2004), "Poche parole", in M. Carrara and P. Giaretta (eds.), *Filosofia e Logica*, Soveria Mannelli, Rubettino, pp. 175–225.

Devitt, M. (2004), "The case for referential descriptions", in M. Reimer and A. Bezuidenhout (eds.), *Descriptions and Beyond*, Oxford University Press, Oxford, pp. 280–305.

Donnellan, K. (1966), "Reference and definite descriptions", *Philosophical Review*, 75, pp. 281–304, reprinted in his *Essays on Reference, Language, and Mind*. Oxford University Press, Oxford 2012, pp. 3–30.

——— (1978), "Speaker reference, descriptions and anaphora", in P. Cole (ed.), *Syntax and Semantics, Vol. 9, Pragmatics*, Academic Press, New York, pp. 47–68, reprinted in his *Essays on Reference, Language, and Mind*. Oxford University Press, Oxford 2012, pp. 115–146.

Kaplan, D. (2012), "An idea of Donnellan", in J. Almog and P. Leonardi (eds.), *Having in Mind*, Oxford University Press, Oxford, pp. 122–175.

Kripke, S. (1977), "Speaker's reference and semantic reference", *Midwest Studies in Philosophy*, 2, pp. 255–276, reprinted in his *Philosophical Troubles: Collected Papers, Volume I*, Oxford University Press, Oxford 2011, pp. 99–124.

——— (1979), "A puzzle about belief", in A. Margalit (ed.), *Meaning and Use*, D. Reidel, Dordrecht, pp. 239–283, reprinted in his *Philosophical Troubles. Collected Papers, Volume I*, Oxford University Press, Oxford 2011, pp. 125–161.

——— (1980), *Naming and Necessity*, Harvard University Press, Cambridge, MA.

——— (2011), "Unrestricted exportation and some morals for the philosophy of language", in *Philosophical Troubles: Collected Papers, Volume 1*, Oxford University Press, Oxford, pp. 322–350.

——— (2013), *Reference and Existence*, Oxford University Press, Oxford.

Marti, G. (2008), "Direct reference and definite descriptions", *Dialectica*, 62, pp. 43–57.

Napoli, E. (1992), "Riferimento diretto", in M. Santambrogio (ed.), *Introduzione alla filosofia analitica del linguaggio*, Laterza, Milano, pp. 385–429.

——— (1995), "(Direct) reference", *Journal of Philosophical Logic*, 24, pp. 321–339.

Neale, S. (1990), *Descriptions*, MIT Press, Cambridge, MA.

——— (2004), "This, that, and the other", in M. Reimer and A. Bezuidenhout (eds.), *Descriptions and Beyond*, Oxford University Press, Oxford, pp. 68–188.

Reimer, M. (1998), "The Wettstein/Salmon debate: Critique and resolution", *Pacific Philosophical Quarterly*, 79, pp. 130–151.

Russell, B. (1912), *The Problems of Philosophy*, Williams & Norgate, London, republished as *The Problems of Philosophy*, Dover Publications, Mineola NY 1999.

——— (1918–19), "The philosophy of logical atomism", *Monist*, 28–29, reprinted in his *Logic and Knowledge: Essays 1901–1950*, Routledge, London and New York 1956, pp. 177–281.

Strawson, P. (1950), "On referring", *Mind*, 59, pp. 320–344.

Wettstein, H. (2012), "Having in mind", in J. Almog and P. Leonardi (eds.), *Having in Mind*, Oxford University Press, Oxford, pp. 93–106.

On descriptive reference fixing*

Vittorio Morato

1. Introduction

Among the many morals one could draw from direct reference (or at least one of the morals I have drawn) is that there is a gulf between general and singular thoughts (or between what we might call "satisfaction-based thoughts" and "reference-based thoughts"). In addition, proper names are special because, *qua* proper names, they can only be used (in their typical uses) to express singular thoughts. The idea is that there is something in the nature of naming that grounds this connection between names and singular thoughts.[1]

To make my psycho-semantic 'morality' a bit more flexible, I relax it with the (Donnellanian) view according to which the association of other types of expressions (e.g., definite descriptions) with general or singular

---

[1] The moral I am drawing from direct reference should not be confused with a *consequence* of it. It would be a consequence only if some auxiliary theses are assumed. In this paper, I take direct reference to be the view that there are directly referential expressions (and proper names are among them) and that a directly referential expression contributes an object to the proposition expressed by the sentence containing it. The propositions expressed by sentences containing directly referential terms are singular propositions. This is, roughly, the characterisation of direct reference given by David Kaplan (1989: 568–569). To this Kaplanian view, one should add some theses connecting singular propositions to singular thoughts and general thoughts to non-singular propositions, and this should be enough to entail the moral I am discussing above.

thoughts might depend on previous intentions of the speaker (what the speaker 'has in mind') and not simply, as it seems to be for proper names, on the nature of the linguistic expression used or, more importantly, on the nature of the linguistic act associated with the expression.

A consequence of this picture is that attributive uses of definite descriptions and typical uses of names are associated with completely different kinds of thoughts and this is, basically, the reason why a psycho-semantical analysis of proper names cannot be given in descriptive terms.

Within this picture, the possibility of fixing the reference of a proper name by description (usually while introducing the name into the language) should be seen, to use again the moral metaphor, as something 'immoral'. Descriptively introduced names should be banned or, at least, taken as a 'deviant' phenomenon. The problem is, roughly, this: how could I have singular thoughts by means of expressions that allow me to have only general thoughts?

Unfortunately, my position is quite uncomfortable to defend, even within the direct reference tradition, mainly because some of the Founding Fathers of this tradition, notably Saul Kripke and David Kaplan,[2] seem all to be in favour of such an 'immoral' practice and they all believe that descriptive reference fixing should be allowed. In Kripke's case, for example, the existence of contingent *a priori* truths, a building block of *Naming and Necessity*, entirely depends on the possibility of introducing names by description. Furthermore, some of the Founding Fathers give no sign of taking such names ("Neptune", "meter", "Jack the Ripper") as somewhat deviant (they are taken to be genuinely directly referential) or of believing (even though they are not always very explicit on these matters) that the thoughts we have, using such names, are different in kind from the thoughts we have using properly (i.e., non-descriptively) introduced ones.[3]

It should come as no surprise that my psycho-semantic morality agrees instead, at least on its negative sides, with that of Ernesto Napoli. This is easy to explain: Ernesto has been the one who actually triggered my psycho-semantic morality in the first place (although he should not be held responsible for my subsequent deeds) as well as of much of my philosophical sensibility ... and, I have to say, of much of my philosophical

---

[2] Cf. Kripke 1980, Kaplan 1973.

[3] Donnellan is, again, an exception here. Cf. Donnellan 1977.

prejudices (some of which I am still having trouble freeing myself of). In "(Direct) Reference" he writes:

> As a consequence of the directness of the referring relation, we have that *no dubbing by description is possible.*
> [. . .] To suppose that there were a dubbing by description, is to suppose that given a description one can introduce an expression as a name of the putative unique satisfier of the description. Hence, there would be a case of conditional naming, i.e., a naming which takes place conditionally upon the satisfaction of a condition. But the satisfaction of a condition depends on the meaning of the condition and how the world is. Satisfaction is not a direct relation, in that it presupposes meaning rather than establishing meaning. Hence there is no naming by description in so far this would amount to grounding meaning on the satisfaction of meaning. (1995: 325-326)

In this paper, I want to show that the practice of introducing names by descriptive reference fixing is actually quite mysterious and at risk of generating strange results. In particular, I will show that tentative tamings of such a practice are not working, either because they would be helpless in problematic cases (or at least in the problematic case I am going to present in the next section) or because they would be so strong as to exclude (what are standardly taken as) non-problematic cases of descriptive name introductions. The kind of position I would like to put pressure on is one combining the view that names are directly referential and the view that descriptively introduced names are genuine names, and thus a position according to which descriptively introduced names are devices of direct reference. This does not mean, however, that my contribution should be taken in any way as an indirect defense of descriptive approaches to descriptively introduced names (both when they get combined with a Millian approach to non-descriptively introduced ones or, more coherently, are part of a full descriptivist approach to names).[4]

---

[4] Gareth Evans (1979) suggested that descriptively introduced names are equivalent to rigidified definite descriptions (where the description is the one used to introduce the name) while, at the same time, defending a causal theory for ordinary proper names (Evans 1973).

2. The case of "Pablo Hernandez"

The case of the descriptive introduction of the proper name "Neptune" is usually taken to be one of the standard cases. It can be described (like many of the others) in two stages. *First stage*: Urbain Le Verrier, prior to the telescopic observation of Neptune, on the basis of other empirical observations, performs a verbal ceremony – presumably private, but it could also be imagined as public – along the following (idealized) lines: "I will call 'Neptune' the planet (if any) causing the perturbations in the orbit of Uranus". The "if any" signals that the definite description "the planet causing the perturbations of Uranus" is used attributively. *Second stage*: . . . *yada, yada, yada*[5]. . . "Neptune" is used by Le Verrier, and by whoever happens to inherit the name through the right causal chain, to express singular thoughts about Neptune.

As you can guess, the mysterious part is what lies behind the "*yada, yada, yada*" in the second stage, and this is exactly what a direct reference theorist should find mysterious or irritating, namely the generation of singular thoughts by means of something general. As we will see, however, even the introduction stage (the first stage) is not without mysteries itself.

As I said, the case of "Neptune" is usually taken to be one of the 'standard cases' and, in many respects, it is even one of the most straightforward. Neptune was later empirically observed, it actually satisfied the description used by Le Verrier to introduce its name, so it was there to be successfully named when Le Verrier introduced (or tried to introduce) a name for it.[6]

But in the "Neptune" case, everything goes definitely too smoothly; everything is already in the right place in the imagined scenario, and so it is difficult to make explicit what are the preconditions for descriptive reference fixing (if there are any). Far more instructive, as is often the case with imagined scenarios and thought experiments, would be dealing with scenarios where something goes wrong. So let us now consider an

---

[5] https://en.wiktionary.org/wiki/yada_yada_yada; the *locus classicus* is this: https://www.youtube.com/watch?v=O6kRqnfsBEc.

[6] As we will see later on, some issues about descriptively introduced names depend on what conception one has of the introduction of a name into a language.

imagined scenario where the introduction of a name by description goes obviously wrong.

Imagine the following dialogue between a history professor and a student during an exam on Italian history (assume that the student does not know the correct answer to the question the professor is asking):[7]

> *Professor:* Who was that general, born in Nice in 1807, commanding the so-called 'Italian legion' during the Uruguayan civil war?
> *Student:* Well, letting 'Pablo Hernandez' refer to the general, born in Nice in 1807, commanding the so-called 'Italian legion' during the Uruguayan civil war, my answer is that Pablo Hernandez was the general, born in Nice in 1807, commanding the so-called 'Italian legion' during the Uruguayan civil war.
> *Professor:* This is not the right answer, it is not true that Pablo Hernandez was the general, born in Nice in 1807, commanding the so-called 'Italian legion' during the Uruguayan civil war.
> *Student:* Sorry professor, but it is true, I am thinking just of the right person and such a person is the very same person you are thinking of and, more importantly, I am saying something true of him!

How to react to this dialogue? On the one hand, the natural reaction would be that of claiming that the professor is right to say that "Pablo Hernandez" is not the right answer to the question she asked. To answer correctly to a "who-is" question (one that is asking someone's name) is, after all, to know the 'right' name of the person, where the 'right' name is that already introduced in the language and socially recognised as such (in the case of an exam, it is the name used in the relevant textbooks). On the other hand, those who admit descriptive reference fixing (on the model of the "Neptune" case, with all its *yada, yada, yada*s) should admit that the student is somewhat right when she claims that the person whom she is thinking about is the same person to which the professor is thinking about. To answer correctly to a "who-is" question (made by someone who already knows the answer) is, after all, to have the 'right' person in mind and, in case descriptive reference fixing is allowed, the student has surely a thought (actually more than that: a true belief) about the 'right' person when she claims that Pablo Hernandez is the general who did such-and-such.

---

[7] The case is an adaptation and an expansion of a case briefly (and almost incidentally) discussed by Scott Soames (2003: 411).

Friends of descriptive reference fixing (FDRFs, from now on)[8] should thus describe the situation as one where the student gives the wrong answer, but has the right person (and the right thought) in mind. When the student says that Pablo Hernandez did such-and-such, due to the nature of descriptive reference fixing, it is of Giuseppe Garibaldi that she is talking about, the same person the professor is asking about.

But given that the student and the professor are talking of the same person, how could there be a disagreement among them? If there is no disagreement, how could the professor claim that the student gave a wrong answer?

The "Pablo Hernandez" case is clearly an absurd case, but FDRFs could only explain part of its absurdity. They can explain why the student gave the wrong answer, but they have troubles, I think, in denying that the student, by introducing the name "Pablo Hernandez", can have a thought about the right person. But there is no way in which a student who does not know who was the general born in Nice in 1807 commanding the so-called "Italian legion" during the Uruguayan civil war could have a true belief about this person simply by means of the semantic mechanism of descriptive reference fixing.[9]

---

[8] As I said in the introduction, by "friends of descriptive reference fixing" I have in mind, as far as this paper is concerned, those who believe that names introduced by descriptive reference fixing are directly referential.

[9] Assume that, by chance, the student actually knew an ostensively introduced name of Garibaldi, a name completely unknown to the community of historians and to the general public, but still a genuine name for Garibaldi. Assume that this name is "Lello Raimondi". In such a case, one where the student would have answered "Lello Raimondi was the the general, born in Nice in 1807, commanding the so-called 'Italian legion' during the Uruguayan civil war", the reaction of the professor and the dialogue between the professor and the student would have developed, at least roughly, along the same lines: "Lello Raimondi" would not be accepted by the professor as an answer to her question and the student could protest that he has instead said something true of Garibaldi. This might be taken to show that a case such as that of "Pablo Hernandez" is not typical of descriptively introduced names, that it is not *because* of the descriptive nature of "Pablo Hernandez" that the dialogue between the student and the professor is problematic. Surely, the case of "Lello Raimondi" is as problematic as the case of "Pablo Hernandez". The difference, however, is that the mechanism of descriptive names introduction would render such cases *systematic*: it would always be possible to answer to a "who-is" question by introducing a new descriptively introduced name. The case of "Lello Raimondi" is instead a mere *contingency*, based on the contingent fact that a student already knew a not well-known

3. Possible reactions

The challenge for FDRFs is thus that of giving us a convincing explanation of why the "Pablo Hernandez" case is absurd. What they should be able to do is to single out this 'bad' case from commonly believed 'good' cases, such as "Neptune", "meter", or "Jack the Ripper" (as we will see, there is no unanimous consensus on whether all these descriptive names should be counted as good ones). It seems to me that, in order to do that, the two most natural explanations would be the following:

(A)   the case is absurd, because the student is not allowed to introduce the name "Pablo Hernandez".
(B)   the case is absurd, because the student, while allowed to introduce the name "Pablo Hernandez", cannot entertain, solely on the basis of such an introduction, singular thoughts and have true beliefs about the person who actually satisfies the description used by the professor.

Depending on what explanation, (A) or (B), is preferred, the dialogue between the student and the professor on page 93 might have taken a different twist. In case the first explanation, (A), is preferred, the professor could have stopped the student already at the very first stages of their discussion and said something like: "No, I am sorry, you are not allowed to introduce a new name for the person I am asking about, so whatever you are going to say next, your answer is wrong!". In case the second explanation, (B), is preferred instead, the professor could have reacted in the following way: "Listen, you are surely free to name the person I am asking about as you want, but by means of such a semantic trick, you will not be able to have thoughts about this person and, more importantly, you will not be able to acquire justified true beliefs about such person; the exam serves just this purpose (namely, to check whether you have justified true beliefs, or at least justified beliefs, about the person I am asking), so your answer is wrong!".

Whatever be the favoured explanation that FDRFs will choose in the end, it is worth noticing that the 'official' formulation of their view –

name of Garibaldi. So there is, after all, a level in which the case of "Pablo Hernandez" is problematic just because it involves descriptively introduced names.

as far as it emerges from the standard narrative accompanying it, for example, in the "Neptune" case – is deemed to be enriched or substantially changed.[10] In the case (A) is favoured, then, contrary to what it is usually told, there are some preliminary conditions that allow or do not allow a speaker to descriptively introduce a proper name. The problem, in such a case, is to determine what these conditions are and whether they are plausible. In the case (B) is favoured, the required change is more dramatic in that it forces the abandonment of the view that, by means of descriptively introduced names, singular thoughts are expressible and thus of the view that they are devices of direct reference.[11] We will see later on, in section 5 of the paper, what this dramatic change would amount to.

In what follows, I will try to show that neither (A) nor (B) have the chance of being good explanations of the absurdity of the "Pablo Hernandez" case.

4. Conditions on descriptive reference fixing

The aim of this section is to discuss the first kind of explanation, (A). The challenge for FDRFs is that of determining what conditions there are and how plausible they are to allow, or not allow, a speaker to introduce a descriptive name.[12]

I take the conditions that will be discussed as conditions regulating a *practice* and thus as having a *normative* character; they will be taken as specifying *felicity conditions* for the practice of descriptive reference fixing. Of course, I do not pretend that the list of conditions I am going to discuss be exhaustive, but I think it covers fairly well the spectrum of

---

[10] I think this to be further evidence that the use of thought experiments or idealised scenarios in philosophy is definitely more useful in its disconfirming (or at least 'problematising') role rather than in its confirming one.

[11] I am assuming here that there is an essential connection between a term being directly referential and the thoughts expressed by a sentence containing it being singular. This, I surmise, is a consequence (with the addition of some auxiliary theses) of a Kaplanian conception of directly referential terms (see footnote 1).

[12] A note on terminology: by "descriptive name" I mean a descriptively introduced name, not a name whose semantics is descriptivist.

possible cases. This will allow me to make some (prudent) generalisations at the end of the section.

So, let us imagine that there is a speaker s willing to introduce a descriptive name n by means of an attributive use of some descriptive material D.

A condition one might start with is the following semantic condition:

Semantic condition: s is authorised to introduce n into a (natural) language in case "the D" is uniquely satisfied.

The plausibility of such a condition depends on the notion one has of "introduction of a name into a (natural) language". The one presupposed in the semantic condition above is, admittedly, quite robust. There might be other, lighter conceptions that do not connect the felicitous introduction of a name by description to a 'semantic fact' like the satisfaction of a definite description (for some, descriptive names can be introduced even by indefinite or improper definite descriptions, for some others, no previous 'semantic fact' might be required to back up the introduction of a name).[13]

Note, however, that this condition is satisfied by almost all standard cases of descriptively introduced names ("meter", "Neptune") and should be taken as being satisfied even in the case of "Pablo Hernandez".[14]

Some doubts may be cast on whether the introduction of "Jack the Ripper" satisfies the condition and thus whether "Jack the Ripper" really is a name:[15] we do not know, and probably we are not anymore in the

---

[13] Note that for Kaplan, something analogous to the semantic condition seems to be the only condition needed for a descriptive name to be felicitously introduced. He writes: "The requirement for a successful dubbing [...] is simply that the dubbee be, in some way, uniquely specified" (1973: 506).

[14] "Vulcan" is one of the standard cases that does not satisfy the condition.

[15] See Salmon 1987. Even Ernesto seems to have doubts about "Jack the Ripper". He claims that if a description may fail to be proper (and such is the case for the description associated with "Jack the Ripper", but even for that associated with "Newman1") what has been performed is an abbreviation of a definite description, not a naming (Napoli 1995: 326). It is not clear, however, whether the claim that "Jack the Ripper" is an abbreviation of a definite description implies that "Jack the Ripper" has been felicitously introduced in the language. Furthermore, it is not clear whether Ernesto believes that there are definite descriptions that *could not fail*

condition to know, whether a single person was the author of the relevant murders, so we do not know whether the name satisfies the semantic condition.

What is important for us, however, is that this semantic condition, plausible or not, is not enough to distinguish good cases of descriptive names like "meter" or "Neptune" from bad cases like "Pablo Hernandez", so, in any case, it would not be a good condition by means of which FDRFs might explain the absurdity of the "Pablo Hernandez" case.

We might thus try with some *epistemic* conditions. Maybe an act of descriptive reference fixing is felicitous in case the speaker doing the introduction occupies certain privileged epistemic positions. The hope for the FDRFs is that bad cases of descriptively introduced names such as "Pablo Hernandez" could be singled out from good cases by pointing out some epistemic deficiency on the part of the speaker who does the introduction.

A first candidate might be the epistemic counterpart of the semantic condition just discussed.

Uniqueness condition: a speaker s is authorised to introduce a descriptive name n on the basis of some descriptive material D, if s has sufficient evidence that D is uniquely satisfied.[16]

According to the uniqueness condition, no one should be allowed to introduce a proper name by means of some descriptive material, unless one has sufficient evidence that "the D" is proper.

Now, what counts as "sufficient evidence" might be difficult to tell and it could vary from context to context. My inclination is not to take such evidence to be so strong as being *knowledge*. If one could not introduce a descriptive name unless one *knew* that "the D" is a proper definite description, many standard cases of descriptive names would not have been allowed in the first place. Consider the case of "Jack the Ripper". It is implausible to suppose that whoever introduced the name did have *knowledge* that the author of the murders was a unique individual. After

---

to be proper and what happens to names eventually introduced by means of these descriptions.

[16] From now on, by "introduction of a name" I will always mean "introduction of a name into a natural language".

all, the existence of such a unique individual was a *hypothesis* and hypotheses are not known. The case of "Neptune" is similar. Le Verrier's postulation of a planet, at the moment of introducing "Neptune" was, again, just a *hypothesis* and, again, hypotheses are not known.[17]

So the relevant evidence needed to satisfy the uniqueness condition is not knowledge. But what is it and, more importantly, how much of it do we need? It is not important that we settle the issue here. We might be generous enough and concede to FDRFs that those who introduced "Neptune" or "Jack the Ripper" did so on the basis of a sufficient amount of evidence (relative to their context of introduction). We might also concede that, in case the policemen had no sufficient (investigative) evidence that the author of the murders was a unique individual or in case Le Verrier did not have any (scientific) evidence that there exists a unique perturbator of the orbit of Uranus, their introductions would not have been taken seriously and thus they would not have been felicitous.[18] The same could be conceded for the case of "meter", especially in the case the construed scenario is one where the subject actually perceives stick s.

The problem is that even the introduction of "Pablo Hernandez" seems to satisfy this condition. Relative to the context of the introduction (with its associated standards of evidence), the student seems to have a lot of evidence that there is a unique individual satisfying the description used by the professor in the formulation of her question (assuming, as we did, that the descriptive material was uniquely satisfied). After all, the question has been made by a *professor*, one who (in most cases) knows a lot about her subject matter. The probability that the question is a trick (i.e., that there is not a single general born in Nice commanding the Italian legion in Uruguay) is actually quite low, so it is

---

[17] Of course, there might be contexts where it is just knowledge what is required to introduce a descriptive name: there might be cases of descriptive reference fixing in logic or mathematics where one needs first to *prove* that a certain descriptive condition is uniquely satisfied and only then one is allowed to introduce a name for the denotation of the description.

[18] Note, however, that the felicity of the introduction is a different thing from the name's being 'taken seriously' or the name's having 'success' in the linguistic community. Those who claim that the introduction of a descriptive name should not be taken seriously unless some conditions (epistemic or not) are satisfied are not yet proving that the introduction of such a name would not be felicitous.

rational for the student to assume that there is a unique satisfier of the description. If it is rational for her to believe it, she must have enough evidence to believe it, so the introduction of "Pablo Hernandez" should count as perfectly acceptable under the uniqueness condition. But then FDRFs would not be able to explain the absurdity of its introduction by means of such a condition. Something else (or something more) is needed.

Another kind of condition could be the following:

No-renaming condition: a speaker s is authorised to introduce a descriptive name n on the basis of some descriptive material D, if s has some evidence that the individual denoted by "the D" has not already been named.

This condition seems to be particularly suited for the "Pablo Hernandez" case. Indeed, the student has more than "some" evidence that the person the professor is asking for has already been named. The student is therefore violating the condition and her act of descriptive reference fixing should thus be taken as non-felicitous.

The problem for the no-renaming condition, however, is that it does not clearly seem to hold in other typical examples of descriptive names, such as "Jack the Ripper", "Julius", or "meter" and it is not clear whether it holds in the case of "Neptune".

Consider the case of "Neptune". It is surely true that Le Verrier's hypothesis (the existence of a planet) is presented as the result of a *discovery* (the calculated perturbations in the orbit of Uranus). So, it seems plausible to assume that Le Verrier was the first to know about the perturbations and the first to hypothesise the existence of a planet as the cause of such perturbations. Does it follow from all this that he had acquired also some evidence to the effect that Neptune had not already been named or that he would have been required to be in possession of such evidence?

We could imagine the case of an envious colleague of Le Verrier that, prior to the introduction of "Neptune", but after the hypothesis of the existence of a planet causing the perturbations of the orbit of Uranus has been made (assuming, for the sake of the argument, the two events happened in slightly different moments) anticipated Le Verrier and, unbeknownst to him, named the hypothesised planet "Jakku". Even though,

afterwards, Le Verrier came to know about the misconduct of his envious colleague, should we consider Le Verrier's descriptive naming ceremony infelicitous? Should we say that Le Verrier was not in an epistemic position to introduce "Neptune"? After all, in such a case, he might have reacted with a "Who cares?". In the "Neptune" case then, the no-renaming evidence seems to be at least less relevant for the felicity of the act than in the case of "Pablo Hernandez".

The irrelevance of the no-renaming evidence is more clearly revealed in the case of "Jack the Ripper". Would the descriptive naming introduction of, say, Scotland Yard's detectives have been infelicitous, unless they also had acquired some evidence to the effect that nobody in London already named the author of the murders? Why should they bother with that? In fact, quite the contrary was the case. Jack the Ripper, if the one and only author of the killings existed at all, quite certainly *had a name already* when the detectives descriptively introduced a name for him and, more importantly, the policemen had evidence of that. In fact, they gave him a descriptive name, waiting to discover the real name of the person. Their introduction was made knowing that the person already had a name.

This fact generalises. In many typical cases, descriptive names get introduced, in what seems to be a felicitous way, to name individuals who already have a name. Consider the case of 'on the spot' introductions. For example, I might felicitously introduce the name "BB-8" for whoever was the first person in Italy to have bought a ticket for *Star Wars: The Force Awakens* movie. In such uses, the no-renaming condition is quite clearly violated, because I have surely some evidence that the person who bought that ticket already has a legal name.[19]

What about "meter"? The scenario is not usually presented as one when some evidence is acquired to the effect that a name for the length of stick s has already been introduced. In any case, such the satisfaction of such a requirement would be quite implausible to ask. How could I acquire some evidence that an abstract entity has not already been named?[20]

---

[19] I could drop "BB-8" once I happen to be acquainted with this person, but this does not mean that my introduction was not felicitous.

[20] I am here assuming that "meter" is to be treated as a proper name. Kripke himself seems to assume this in *Naming and Necessity* and so do many authors working

It might be that a stronger no-renaming condition would do a better job. Let us try with this one:

Strong no-renaming condition: a speaker s is not authorised to introduce a descriptive name n on the basis of some descriptive material D in case s already knows a name of the individual denoted by "the D".

In effect, we might concede that it would have been pointless to descriptively introduce "Jack the Ripper" if one already knew that, e.g. John Smith, was the author of the murders (but, again, its being pointless does not mean that it would be infelicitous). Instead, we have seen above that it would not be pointless for Le Verrier to introduce "Neptune" even if he knew about the preemptive naming of the envious colleague. The same seems to apply to the case of "meter".

Note, however, that the strong no-renaming condition would not be of any help in the "Pablo Hernandez" case. The scenario should be construed as one where the student did not know the real name of the person the professor was asking for so she is not violating the strong no-renaming condition.

Furthermore, the condition is quite clearly violated in many cases of (what seem to be) felicitous introductions of descriptively introduced names. Again, cases of 'on the spot' introductions are helpful. For example, I can introduce the descriptive name "First" for the first among my friends (whose name I know for sure) who will be served at the restaurant, and I could keep calling her or him "First" for the rest of the evening. It would be obviously a joke, but surely not a case of an infelicitous introduction.

Finally, there is, I think, also a general reason why something like the no-renaming condition does not work, in its weak or strong version. Indeed, no such condition seems to be at work in the cases of ordinary, non-descriptively (i.e., ostensively) introduced names. Apart from very codified contexts of introductions (for example, scientific or legal), ordinary names can be non-descriptively introduced without anyone caring

---

on descriptively introduced names. Actually, I have some doubts: "meter" should probably better be treated as a *functor*, an operator that attaches to a singular (numerical) term to form a new singular term. On this, see Salmon 1987: fn. 5.

about possible cases of re-naming.[21] But then, in case the introductions of descriptively introduced names were regulated by something like the no-renaming condition(s), the conditions of their introduction would be stricter than those required for ostensively introduced ones. Descriptive names would then be some sort of 'super-names'. But FDRFs (at least those I am discussing in this paper) usually fight against discrimination, they believe that descriptively introduced names are genuine names and should be treated as such (with all their *yada, yada, yadas*); they usually do not fight for the recognition to these names of a special, privileged status. Ordinary names are not subjected to the no-renaming conditions, so we should not expect that descriptively introduced names be subjected to similar conditions.

Another condition might be the following:

Appropriateness condition: a speaker s is authorised to introduce a descriptive name n on the basis of some descriptive material D in case the context of the introduction is appropriate.

This condition presupposes that the introduction of descriptive names is context-dependent and that, among the contexts of introduction, there are some contexts that are appropriate (allowing the introduction of descriptive names) and some contexts that are not appropriate (not allowing the introduction of descriptive names).

The "Pablo Hernandez" case could, in effect, be described as a case where the appropriateness condition for the introduction of descriptive names has been violated. When the professor reacts by saying, "You are not allowed to introduce a new name for the person I am asking about", what she plausibly meant to convey was the idea that a history exam is not a context where the introduction of descriptive names (and, eventually, of ordinary names) is allowed. Appeal to the appropriateness condition could then be used by FDRFs to explain why the "Pablo Hernandez" case is illegitimate.

What about "Neptune", "meter" and "Jack the Ripper"? What features of the contexts of their introduction make such contexts appropriate? It seems that the context of a scientific hypothesis, and probably of hypotheses in general, should eventually be taken as appropriate for

---

[21] Again, it might be that the name that comes last does not get any 'success', but this does not mean that its introduction is not felicitous.

the introduction of descriptive names (after all, the whole practice is somewhat reminiscent of what happens in logic or mathematics when one introduces arbitrary names). So "Neptune" and "Jack the Ripper" seem to be fairly covered by the appropriateness condition. Even the context of the introduction of "meter" seems to be appropriate as far as we imagine the stipulator as being committed to the introduction of the new metric system. Under the appropriateness condition, it seems thus possible to single out the introduction of "Pablo Hernandez" from that of other standard descriptive names and explain why this introduction, unlike the others, is infelicitous.

The problem is that the appropriateness condition not plausible. In particular, I am sceptical over the existence of substantial, contextual appropriateness conditions for the introduction of descriptive names (and of names in general). It might be quite easy, in the case of "Pablo Hernandez", to spot the context of the introduction as inappropriate (and, in the case of "Neptune" to spot the context as appropriate). But it is difficult to generalise from that. What makes a context appropriate or inappropriate? What do the contexts that allow the introduction of such names, or those that do not allow their introduction, have in common? Is there a feature of a context of being *(descriptive) name-introductory*?

As in the case of the no-renaming condition, the plausibility of the appropriateness condition could be assessed by comparison with the case of ordinary, non-descriptive names. Indeed, the situation looks structurally similar. We seem to have these two possible options: (i) descriptive names are like ordinary, non-descriptively introduced names; (ii) descriptive names are 'super-names' (i.e., genuine names to which some special conditions apply).

If (i) is the case, then the appropriateness condition is not plausible for descriptive names, because it is not plausible for ordinary, non-descriptively introduced ones. The appropriateness condition would make the introduction of a name dependent on the satisfaction of some contextual condition, where this seems not to be the case. It might be helpful, at this point, to quote Ernesto again. In "Names as Predicates?", he claims that what makes an expression "a name of an individual" is just "to be used as a name of an individual". And he continues:

> There is no more substance to being a name than being put to a certain use, than being made to play a certain role, namely, that of a quite arbitrary (and hence a-contextual) tool for reference to an individual. (Napoli 2015: 214)

Names are *a-contextual* tools for reference to individuals, so nothing like the satisfaction of a contextual condition (such as the appropriateness condition) applies to their introduction. The decision to use an expression as a name is simply unconstrained. Thus, if the condition does not apply to non-descriptively introduced names, then, under (i), it should not be applied to descriptively introduced ones.

If (ii) is instead the case, descriptively introduced names are 'super-names' that differ from ordinary, non-descriptively introduced ones for the special conditions imposed on their introduction. In our case, they are 'super-names' that can be introduced only in contexts tagged as 'appropriate'. The problem in such a case is the following. On the one hand, as I have claimed above, FDRFs are not trying to defend the view that descriptively introduced names, if genuine names at all, are 'super-names'. On the other hand, even admitting that descriptive names are special, they seem to be capable of being introduced 'on the spot', exactly like their non-descriptive counterparts (the case of "First" or "BB-8" mentioned above are cases in point). 'On the spot' introductions seem to be incompatible with substantial contextual requirements or appropriateness conditions, because appropriateness conditions require a classification of all introductory contexts and such classification seems implausible. There are more ways to introduce names than there are classified contexts of introduction or conditions of appropriateness. Surely, names get introduced within contexts, but not within contexts whose features determine the felicity of the introduction.

My conclusion is that FDRFs could not appeal to the appropriateness condition to explain why the "Pablo Hernandez" is an absurd case.

The last condition we are going to discuss has been proposed by Scott Soames, originally in Soames 1998 and, more recently, at the end of his discussion of the contingent *a priori* done in ch. 16 of Soames 2003. According to Soames, the following condition holds for the introduction of descriptive names:

Soames's condition: a speaker s is authorised to introduce a descriptive name n on the basis of some descriptive material D only if s independently believes of the object denoted by "the D" that the description applies to it.

Something along similar lines (actually, something weaker) seems to be defended by Ernesto in "(Direct) Reference", where he writes:

> A description can occur in a dubbing only as a way of focusing on an independently given object, i.e., as attention caller to an independently individuated object. (Napoli 1995: 325)

Soames's condition is specifically designed to be applied to a certain reconstruction he does of the meter stick example in *Naming and Necessity*. According to Soames, the scenario could be reconstructed in two ways: (i) one where we are *seeing* stick s, we form an idea of its length and *then* introduce a name ("meter") for this length with the description "the length of stick s at t", or (ii) one where we introduce the term "meter" for whatever satisfies (an attributive use of) the description "the length of stick s at t". Scenario (ii) is compatible with us as never being perceptually in contact with stick s, while scenario (i) grants us with a perceptually justified true belief about stick s (i.e., the belief that one meter is its length). Scenario (ii) is one where the introduction of descriptive names is unrestricted, and where cases *à la* "Pablo Hernandez" are possible, while scenario (i) is one where descriptive names should be introduced only under what I have called "Soames's condition".

According to this condition, we need first to acquire the justified belief of the object denoted by "the D" that it is the D (in the case of the meter stick example, the justification is perceptual, but, presumably, we could have other forms of justification) and only afterwards we are allowed to introduce a name for the object denoted by "the D" (and we express singular propositions by using such a name).[22]

I have many doubts about Soames's condition and some doubts about his reconstruction of the meter stick example. However, as far as this paper is concerned, I think it is enough to make just the following two observations.

First. Soames's condition does not seem to fit nicely with the "Neptune" case. Did Le Verrier, before introducing "Neptune", have a justified true belief of Neptune that it was the perturbator of the orbit of Uranus? Surely, Le Verrier had some justified true belief prior to the introduction

---

[22] According to Soames, the fact that the introduction of "meter" relies on a previous perceptually justified belief counts against the fact that the proposition expressed by "the length of stick s at t is one meter" is known *a priori* by the stipulator.

of "Neptune", but they were not the beliefs required by Soames's condition. What Le Verrier justifiably believed was, presumably, that there were some perturbations of the orbit of Uranus. He then *hypothesised* that these perturbations were compatible with the existence of a planet and then introduced "Neptune" as the name of this *hypothesised planet*. As any good scientific hypothesis, Le Verrier's hypothesis was based on some justified true belief, but one might doubt that Le Verrier justifiably believed, of the hypothesised planet, that it was the perturbator of the orbit of Uranus. The existence of such a planet was a hypothesis and hypotheses need not be justifiably believed. Even if a hypothesis h is introduced on the basis of a justified belief b, or even on the basis of a piece of knowledge k, h does not 'inherit' the epistemic status of b or k, even though the epistemic status of h surely improves for being grounded on b or k. So it is doubtful that the "Neptune" case counts as an infelicitous introduction of a descriptive name under Soames's condition. The same seems to apply to the case of "Jack the Ripper".

Second. What does it mean "independently believing of the object denoted by 'the D' that the description applies to it"? In case *acquaintance* is the only way of believing *of* an object that a condition applies to it, Soames's condition applies only to those situations where a speaker s is, for example, perceiving, or has perceived in the past, an object o. Now, this situation is compatible with the following two scenarios: (i) s believes of o that o is the denotation of "the D", (ii) o is the D and s believes of o that o is the D.

The first scenario, (i), is compatible with a situation where o is not the D and o is believing falsely of o that o is the D. In such a case, the introduction of a name n for o would not satisfy the semantic condition and, more importantly, it would be a case where n is descriptively introduced by means of a *referential* use of "the D".[23] But such a case would not count as a one where descriptive reference fixing is the same phenomenon discussed in this paper. What we are discussing here is whether directly referential names can be introduced into the language by *attributive* uses of definite descriptions. The practice of introducing names by referential uses of definite descriptions, and the view that, using such names, we

---

[23] Note that even Ernesto's view, expressed in the quotation above, could be interpreted in a similar vein, namely as the view that a description can occur in a dubbing only in the case it is used referentially (as an "attention caller").

can express singular thoughts is not at all problematic and it is not in contrast with direct reference. In a case like that, we have never abandoned, so to speak, the 'realm of the singular', we have never generated singular thoughts by means of something general, and it was just this unduly generation of singular thinking that was taken to be problematic.

The second scenario, (ii), is one where s is acquainted with o, o is the D and o believes that o is the D. Under (ii), Soames's condition is the result of combining the semantic condition, the uniqueness condition *plus* an acquaintance condition (according to which only if s is acquainted with o, s is in the condition to believe *of* o that P, where P is about o). But even in such a case, I think it would be plausible to suppose that the description is used referentially. Even scenario (ii) is thus not a scenario where the relevant phenomenon is described, therefore neither a scenario where the relevant phenomenon is constrained.

Endorsing Soames's condition is thus just equivalent to claim that the introduction of names by attributive uses of definite descriptions is not possible. "Neptune", "Pablo Hernandez", "Jack the Ripper" (but even "Julius", "Vulcan", "Newman1" and all cases of on the spot introductions) would all come out as bad cases. But then, under Soames's condition, attributive descriptive reference fixing comes out as a 'deviant' phenomenon, a conclusion that I surely welcome.

Note that even in the case of the introduction of "meter", the applicability of Soames's condition seems to be problematic: assuming scenario (ii), one should claim that, to introduce the name "meter", one should be acquainted with the denotation of "the length of stick s at t". But is it plausible to suppose that we are acquainted with abstract entities such as lengths?[24]

Therefore, I do not think that Soames's condition represents a useful condition on the introduction of descriptively introduced names. The condition is so strong as to exclude almost all cases. Under this condition, the absurdity of the "Pablo Hernandez" case can surely be explained, but at the cost of considering absurd all other cases of descriptively introduced names.[25]

---

[24] Note that even Ernesto's condition seems to give similar problems: in what sense the length whose "meter" is a name of is "independently individuated" before a name for it has been introduced?

[25] For another way of criticising Soames's condition, see Jeshion 2006: 129–135.

The situation can be summed up with the following table telling us whether a certain descriptive name satisfies ("yes"), does not satisfy ("no") or probably satisfies ("maybe") a certain condition.

	"Neptune"	"Meter"	"Pablo Hernandez"	"Jack the Ripper"
Semantic	yes	yes	yes	no
Uniqueness	yes	yes	yes	yes
No-renaming	maybe	no	no	no
Strong no-renaming	maybe	maybe	yes	yes
Appropriateness	yes	yes	no	yes
Soames's	no	maybe	no	no

What this table reveals is that FDRFs are stuck in the following situation: either the candidate conditions are satisfied by "Pablo Hernandez", which thus comes out as a perfectly legitimate descriptive name, or, if they somehow happen to block the introduction of "Pablo Hernandez", they would also block the introduction of other descriptive names that are usually taken to be perfectly felicitous.

5. Descriptive names without singularity

This section is about the second kind of explanation that FDRFs could give to explain the absurdity of the "Pablo Hernandez" case.

According to (B), the professor could react to the student's semantic trick by claiming that, while she is free and perfectly entitled to introduce whatever name she wants with the descriptive material at her disposal, by means of such names she will not be able to think and, more importantly, have justified beliefs, about the person the professor is asking for, namely Garibaldi. The situation is thus one in which "Pablo Hernandez" is treated like a genuine descriptive name (the introduction is felicitous), but sentences containing "Pablo Hernandez" do not express singular propositions about Garibaldi. This situation could be generalised to the following general view of descriptively introduced names:

Mixed view of descriptively introduced names: a speaker s is always entitled to introduce a name n by means of an attributive use of

the description "the D", but sentences containing n do not express singular propositions about the denotation of "the D".

I think that the mixed view is untenable. I will put it in terms of a dilemma: either "Pablo Hernandez" is a genuine name and then it is, *qua* name, a vehicle of singular thought (sentences containing it express singular propositions) or "Pablo Hernandez" is not a vehicle of singular thought (sentences containing it do not express singular propositions) and thus it is not a genuine name. The best "Pablo Hernandez" could be is, eventually, a non-genuine descriptive name – i.e., a name that is non-genuine and descriptive. Given that by "genuine" I basically mean "directly referential" and by "directly referential" I basically mean "vehicle of singular thought" (see footnote 1), this position could only correspond to a descriptivist view according to which descriptive names like "Pablo Hernandez" are abbreviations of the definite description by means of which they are introduced.[26]

The mixed view is untenable, because it conflicts with what I take to be a fundamental psycho-semantical principle:

P1 If a competent, rational and sincere speaker s assertively utters, at time t, a sentence S, then s believes the proposition semantically expressed by S at t.

The role of **P1** is that of connecting the assertion of a speaker to the proposition semantically expressed by what she asserts. The principle presupposes the 'Moorean principle' according to which, if s assertively utters S at t, then s believes that S at t (on pain of contradiction), the existence of propositions and the usual relations between them, sentences and mental states such as *believing*.[27]

A consequence of **P1** is that, in case a speaker assertively utters something like "n is the D", then s believes the proposition semantically expressed by "n is the D". In case n is a genuine name, "n is the D" expresses a singular proposition and so, according to **P1**, s believes the singular proposition expressed by "n is the D".

---

[26] Where the original description is rigidified, the position corresponds to that famously defended in Evans 1979.

[27] Principle **P1** could be re-phrased without any (explicit) reference to propositions as the principle for which if s assertively utters S, then s believes what is semantically expressed by S.

The mixed view is a violation of **P1**, because it allows the descriptive introduction of "n" as a genuine name without the sentence "n is the D" expressing a singular proposition. A speaker s may thus assertively utter "n is the D" without believing the singular proposition expressed by "n is the D". But "n is the D" semantically expresses a singular proposition (if "n" is a genuine name) and so, a consequence of the mixed view is that a speaker can, by means of a descriptive name, assertively utter a sentence while not believing the semantical content expressed by the sentence. This goes against **P1**.

We might expect that there might be occasional, highly contextual, controlled violations of **P1**, but what is really problematic in our case is that the conflict with the mixed view would be *systematic*: if one endorses it, every assertive utterance of a sentence containing a descriptive name would be a case where one asserts something without believing what is semantically expressed. This systematic violation of **P1** sounds quite unacceptable.

The incompatibility between the mixed view and principle **P1** does not depend on my 'Kaplanian' conception of directly referential expressions as essentially connected with singular thought. One might, for example, characterise directly referential expressions as *rigid designators* and add to this a psycho-semantic thesis according to which ostensively introduced rigid designators are vehicles of singular thought, while descriptively introduced ones (where the description is used attributively) are not vehicles of singular thought. Descriptively introduced rigid designators may eventually become vehicles of singular thoughts as soon as one is able to ostensively refer to the denotation of the description. Under this view, "Neptune" would be a genuine name (i.e., a directly referential expression, i.e., a rigid designator), but not a name whose typical uses express singular thoughts (before its telescopic observation). The same would go for "Pablo Hernandez". This instance of the mixed view corresponds, roughly, to the one presented in Donnellan 1977.

I find this position problematic for a number of reasons. First, there are all the problems deriving from the characterisation of directly referential expression in terms of the modal notion of rigid designation.[28] Furthermore, and more importantly, even Donnellan recognises that such

---

[28] See Almog 1986, but see also Morato 2001 which is the result of many hours of discussion with Ernesto and owes a lot to him.

a view, and, in general, the possibility of descriptively introducing rigid designators, is "odd" (Donnellan 1977: 24), and the reason is just that it conflicts with a principle closely related to **P1**. The procedure of descriptively introducing rigid designators, Donnellan claims, puts one in the position of introducing into the language sentences expressing truths or falsehoods "without thereby putting the possessor of the apparatus in a position to know what they express".

The introduction of descriptively introduced names thus conflicts with the following principle:

P2 If a competent, rational and sincere speaker s introduces S into the language (such that the reference of any names and demonstratives in S has been fixed), then s knows what S expresses.[29]

In those cases where the introduction of S by s corresponds to, or entails, the assertive utterance of S by s and assuming that "s knows what S expresses" entails (modulo the apparatus of singular propositions) that "s believes the proposition semantically expressed by s" (and that the assertive utterance of S and the corresponding belief are temporally coordinated) principle **P2** entails principle **P1**. So even this mixed, 'non-Kaplanian' conception of descriptively introduced directly referential expressions clearly conflicts with **P1**.[30]

---

[29] Here are some of the relevant passages (both at p. 24 of Donnellan 1977):

> [...] the facts, as I have described them, are that the procedure sets up the apparatus necessary for a set of sentences to express truths and falsehoods without thereby putting the possessor of the apparatus in a position to know what they express.

> If I am right one result of introducing names via definite descriptions in the way suggested is that there are sentences introduced into one's language such that about any particular one "P," one does not know that "P" expresses a truth if and only if P.

[30] At the very end of his article, to better explain why the existence of descriptively introduced names conflicts with a principle like **P2**, Donnellan proposes, following a suggestion of Rogers Albritton, to treat descriptively introduced names as *reserved names* and the sentences containing them as neither true nor false (Donnellan 1977: 24–25). Under this view, a speaker s does not know the truth of S (assuming that S is a sentence introduced by s and containing a descriptively introduced name), because there is no truth to be known by S.

So those defending a mixed view of descriptively introduced names (and choosing (B) as the explanation of the absurdity of the "Pablo Hernandez" case) are in conflict with a principle like **P1**, both in the case they believe that directly referential expressions are essentially connected with singular thoughts, and in the case they believe that directly referential expressions are better characterised as rigid designators. In the former case, the reason is that what is semantically expressed by a sentence containing a directly referential expression would be essentially connected with singularity; in the latter case, the reason is that a gap would be opened between the introduction of sentences into one's language and knowledge of what is expressed by such introductions.

6. Conclusions

My conclusion is that FDRFs cannot explain the absurdity of the "Pablo Hernandez" case. To this conclusion they could react in various ways.

One way is by doing nothing. They might simply accept that sometimes, hopefully in some limited cases, the practice of descriptive name introduction generates strange results. What this paper has shown is that there does seem to be a systematic patch to avoid such strange cases. FDRFs could probably live with this conclusion in case they believe that the introduction of genuine names by (attributive uses of) definite descriptions is something that should be preserved, even on the face of recalcitrant results. After all, they might say, the theoretical benefits of preserving the mechanisms of descriptive reference fixing could, in the end, be greater than the costs of accepting strange cases such as that of "Pablo Hernandez".

Another reaction might be somewhat more dramatic. They could choose to have a two-tiered theory of proper names. According to such a view, only ostensively introduced names are directly referential, but descriptively introduced ones should be treated differently, both semantically and cognitively. In particular, they could claim that, semantically, "n" (where "n" is descriptively introduced by "the D") is synonymous with "the D" and that, cognitively, "n is the G" and "the D is the G" express the same thought. Under this two-tier semantic theory, when

the student answers "Pablo Hernandez was such-and-such" what she is saying (and thinking) really is "such-and-such was such-and-such" and this would be nothing the professor should be worried about. She could easily dismiss an answer like that.

I would not recommend such a view, but I cannot fully explain why in this paper.[31] Its more problematic presupposition, at least for me, is that the kind of thoughts we are able to entertain with names would come out as dependent on the conditions of their introduction. But the conditions in which names are introduced are, more often than not, forgotten, and become simply irrelevant once we acquire and start using them. Thus, if one wants descriptively introduced names, either one should treat them as genuine names (i.e., directly referential) or treat them *and* ordinary names as abbreviations of definite descriptions. But in this latter case, one would end up with a full-fledged descriptivist approach (with all its numerous problems), in the former, cases like that of "Pablo Hernandez" seem to be unavoidable.

References

Almog, J. (1986), "Naming without necessity", *Journal of Philosophy*, 83, pp. 210–242.

Donnellan, K. (1977), "The contingent *a priori* and rigid designators", *Midwest Studies in Philosophy*, 2, pp. 12–27.

Evans, G. (1973), "The causal theory of names", *Aristotelian Society Supplementary Volumes*, pp. 187–208, reprinted in his *Collected Papers*, Clarendon Press, Oxford 1985, pp. 1–24.

——— (1979), "Reference and contingency", *Monist*, 62, pp. 161–189, reprinted in his *Collected Papers*, Clarendon Press, Oxford 1985, pp. 178–213.

Jeshion, R. (2004), "Descriptive descriptive names", in M. Reimer and A. Bezuidenhout (eds.), *Descriptions and Beyond*, Oxford University Press, Oxford, pp. 591–611.

——— (2006), "Soames on descriptive reference-fixing", *Philosophical Perspectives*, 16, pp. 120–140.

Kaplan, D. (1973), "Bob and Carol and Ted and Alice", in J. Hintikka, J. Moravcsik and P. Suppes (eds.), *Approaches to Natural Language*, Kluwer, Dordrecht, pp. 490–518.

---

[31] Cf. Jeshion 2004.

——— (1989), "Afterthoughts", in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford University Press, Oxford, pp. 565–614.

Kripke, S. (1980), *Naming and Necessity*, Blackwell, Oxford.

Morato, V. (2001), "I problemi della designazione rigida", in M. Carrara, G. De Anna and S. Magrin (eds.), *Linguaggio, Mente e Mondo. Saggi di filosofia del linguaggio, filosofia della mente e metafisica*, Il Poligrafo, Padova, pp. 121–160.

Napoli, E. (1995), "(Direct) reference", *Journal of Philosophical Logic*, 24, pp. 321–339.

——— (2015), "Names as predicates?", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 211–224.

Salmon, N. (1987), "How to measure the standard meter", *Proceedings of the Aristotelian Society*, 88, pp. 193–217.

Soames, S. (1998), "The modal argument and rigidified descriptions", *Noûs*, 32, pp. 1–22.

——— (2003), *Philosophical Analysis in the Twentieth Century*, volume 2, Princeton University Press, Princeton, NJ.

Name or predicate?*

Paolo Leonardi

The predicate view of names (PV) has reached a high level of sophistication, closely competing with the referentialist account that distinguishes names from predicates – the name view of names (NV).[1] The two accounts split the heritage of the classical view that a name *denotes* what it names, i.e., that it picks out a particular by some of its features. According to NV, a name points out a *particular* by none of its features, but, according to PV, a name predicates of a particular just the specific *feature* of its name.[2] At the same time, PV and NV join in deeming a name not to categorize the namee.

---

[1] I shall indifferently speak of predicates or of nouns, and use PV and NV in the singular and in the plural. As Joseph Almog reminds me, conflating predicates and nouns is disputable. Yet, I shall not dispute it here.

[2] Denotation, according to Frege 1892, is achieved by means of the sense that a name expresses – the sense supplies the way of determining the namee. In 1905, Russell articulated his famous theory of descriptions and, reading ordinary names as disguised descriptions, developed a predicative view of names – "Dante" has to be paraphrased, say, by the description "the author of the *Comedia*", and any sentence in which the name of Alighieri occurs, such as "Dante was from Florence", has to be rewritten as "There is one and only one individual who wrote the *Comedia* and

Here, I shall avow my preference for NV over PV. PV is deemed more economical and simpler, as it offers a uniform treatment of names and

---

all who did were from Florence". A second step up in the development of PV comes with Quine who, in 1948, states that

> [i]f the notion of Pegasus had been so obscure or so basic a one that no pat translation into a descriptive phrase had offered itself along familiar lines, we could still have availed ourselves of the following artificial and trivial-seeming device: we could have appealed to the *ex hypothesi* unanalyzable, irreducible attribute of *being Pegasus*, adopting, for its expression, the verb "is-Pegasus", or "pegasizes". The noun "Pegasus" itself could then be treated as derivative, and identified after all with a description: "the thing that is-Pegasus", "the thing that pegasizes". (27)

The name is no longer substituted by some description fitting the namee. A predicate like "is-Socrates", Quine imagines, is introduced and applied to an individual, and this predicate later generates the name "Socrates" – a choice more radical than that of substituting "Socrates" with a predicate, such as "being named 'Socrates'", "being called 'Socrates'", or "being the bearer of the name 'Socrates'". More recently, linguists and philosophers have brought the predicative view of names much closer to natural language – see Sloat 1969, Burge 1973, who further adapts Quine's idea, accepting the name "Socrates" itself to be a predicate as is. (Cf. also Hornsby 1976, Crane 1992, Jubien 1993, Thomsen 1997, Geurts 1997, Katz 2001, Bach 2002, Elugardo 2002, Ghomeshi and Massam 2009, Ghomeshi et al. 2009, Sawyer 2010, Rieppel 2013, Rami 2014, Gray 2014, and especially Elbourne 2005, Matushansky, 2006, 2006a, 2008, and 2015, and Fara 2011, 2015, and 2015a. Fara maintains that in the so-called referring use, a name predicate is preceded by a null determiner, *the*.

Russell 1903's view of proper names, later restricted to logically proper names, originated NV. (See Donnellan 1966 and Kripke 1972). Russell suggests that denotation is not a semantic relation, but that reference is. A name directly picks out what it designates. (I shall use "designate" as hypernym for *denote* and *refer*. For the distinction between reference and denotation, see also Almog 2012 and 2014 and Capuano 2012.) The referential view, which can actually be traced back at least to Mill (1843) (and perhaps further back to Olivi and to Hobbes), was fixed between 1966 and 1970, by Donnellan and Kripke, and has been elaborated and refined by many (e.g., Kaplan (1989), Devitt (1981), Salmon (1981, 1986), Wettstein (1984), Perry (2001), Almog (2014)). The referential view has also been defended against the predicative view by Boer (1975), Cohen (1980), Englebretsen (1981), Rami (2013), Frederick (2014), Jeshion (2015, 2015a, 2015b), Napoli (2015), Sainsbury (2015), Predelli (2015), and indirectly by Kaplan (1990), who introduces a distinction between generic and specific names. Among linguists, several hold positions closer to NV, including Longobardi (1994, 2005), Anderson (2007), and Mathieu (2009). Hinzen, a philosopher who is very involved in contemporary linguistics, also takes a stand close to NV, as of 2007.

predicates. I shall argue two points: (i) If names are predicates, they are different from any other, because, if a predicate, a name does not *tell* a feature of what it names, but *is* the feature it tells. A predicate categorizes whereas a name labels; a name is arbitrary and a predicate is not.[3] The difference emerges considering name introductions and name shifts, leaving us with two kinds of nouns instead of nouns and names.[4] (ii) Is PV simpler and more economical? Formally, it is, waiving one category of expressions, though in combining a predicate and a determiner or a demonstrative, it turns names into complex expressions. It leaves out the role of names that originates NVs: names fix points of reference and bring back to their reference frames; predicates characterize these points and further ones in relation to them. A motivation for PVs is that names are shared – there are indefinitely many Annas, Josephs, Muhammads, and Nimrats. In closing, however, I shall hint at a way of looking at names, which reshapes the issue of the indefinitely many particulars sharing a name and shows them to be an almost perfect flexible coding system for individuating particulars. I shall also allude to viewing non-predicating predicates as names, an issue Frege (1892a) investigates.

1. Uniformity and simplicity are among the advantages generally claimed for PV over NV, as I remarked above. Uniformity is an essential element – without uniformities, there would be no knowledge. How much does the formal description make names like predicates? It shows them *formally* alike – it offers parallel representations of them. Does it show them *substantially* similar, as well?

Most names originate from predicates, but the predicates they originate from play no role once they become names. The predicate a name comes from might have been satisfied by an ancestor of the first bearer of the name. Until it was a predicate, it was no name. For instance, "Delia", which was originally Greek and means "from Delos", might very well have been at first used predicatively to describe one or more women

---

[3] Jeshion (2015b: 400–402) seems to think differently. But even PVs generally agree (see Burge 1973: 425, fn.1 and Matushansky 2015).

[4] There are differences at the syntactic level, too. All the linguists mentioned in fn.2 above acknowledge that names are a special kind of nouns – according to them, the issue is whether or not they are a subkind of nouns, a distinction that in a Montague grammar could be described as whether names are of type $\langle e \rangle$ or $\langle e, t \rangle$.

from that Greek isle. Other names might have been appropriate for a
newborn, but as the baby grew might easily have ceased to be appropriate,
as in the case of "Ernesto", from the German *eornost*, which means
"serious". If possible, this phenomenon is even more evident with family
names, like "Smith", a name that spread across Europe in its many languages.
The village smith might have learned his job and inherited his
name from his father for some generations, until the son took a different
job. There are family names that, as "Smith" or "Cooper", originate
from a profession, and others that come from other names – patronymic,
like "Davidson", or place of origin, like "Hamilton" (which formerly was
a town of Leicestershire). The noun from which a name comes, as Kripke
could have said, might originally have fixed the name's referent, but later
name and noun diverged. The noun categorizes, the name does not – it
names.

Differences between names and nouns, if both are predicates, pop up
in introductions and shifts. In between late Medieval time and the 16th
century, in English, for instance, the hypernym for dogs shifted from
"hound" to "dog". The shift involved all hounds/dogs. "Hound" was
'downgraded' to be the name of some specific kinds of dog, as it was
before in the case of "dogge" ("dogge" meant mastiff, basset(t),..., as
"hound" today groups greyhound, bloodhound...). This change was motivated,
because there came to be more dogs than hounds. Instead, a
name shift usually involves only one single individual so named. When
Norma Jeane changed her name, choosing to call herself "Marilyn", no
other Norma had her name changed.[5] Analogously, it goes with introductions.
When the English called "platypus" a platypus, any other platypus
came to be so called (in English), and when they called "bat" a
bat, any other bat came to be so called. But when the parents call their
baby "Anna", no other daughter of theirs and no other female baby
gets thereby to be called "Anna".[6] Names, besides, are more resistant to

---

[5] Rather different from a name change is a *nom de plume*, like Lewis Carroll, George
Elliot, Elena Ferrante, etc., or adjustments of names by adding titles – Sir, Lord,
Lady, etc. The first is no name change but the adding of a second name (and the
building of a distinct social identity); the second is really an apposition, as in "Arthur
Wellesley, Duke of Wellington", where the apposition is put in front out of deference.

[6] All this converges with the tension towards particular individuals, discussed by
Jeshion in her 2015b.

language change than predicates – so much so that nowadays, the pref-
erence is not even to translate names of persons, though a translation is
not a linguistic shift but a change of language.[7] The relation between a
name and a namee is arbitrary, but the relationship between a noun and
what it applies to is not. Anna, by being so called, turns into an Anna,[8]
whereas she does not become a runner by being so picked out – she is a
runner if she runs, and for that reason "runner" can be used to pick her
out. Particulars come to share a name by being so called; an item fits a
noun even if it is never identified as such.

Differences between names and nouns come out also in the truth con-
ditions of sentences involving two expressions. Burge writes:

> In holding that a name applies to an object just in case the object bears a
> certain pragmatic relation to that name, I am suggesting that the name itself
> enters into the conditions under which it is applicable. In this respect, proper
> names differ from many other predicates. Take, for example, the predicate "is
> a dog". An object could be a dog even if the word "dog" were never used as a
> symbol. But an object could not be a Jones unless someone used "Jones" as a
> name. (1973: 430)[9]

The truth conditions of a sentence involving proper names, according
to PV, require that naming be done in the appropriate way. "Baptism,
inheritance, nicknaming, brand-naming, labeling", Burge adds, "may all
be expected to enter into" an account of appropriate naming (435). How-
ever, a homophonic condition like the following, he claims, is sufficient for
semantics: "'O'Hara' is true of any object y just in case y is an O'Hara"
(435), and y is an O'Hara if this name has been given to y's family in the
appropriate way.[10] The name's truth, its being true of, depends on the
appropriateness of the name's endowment, because the name predicates
of the particular only its having that name, no more, and the appropri-
ateness of its introduction is what discriminates calling someone by one's

---

[7] Bach (2015) claims that there being no need to translate proper names suggests
that they "are not lexical items in particular languages" (776–777). I would rather
say that they are loans from foreign languages.

[8] Savage (2014) hints at this *performative* dimension of naming.

[9] Immediately after the quote, Burge adds: "This mild self-referential element in the
application conditions of proper names...".

[10] "A proper name is a predicate true of an object if and only if the object is given
that name in an appropriate way" (Burge 1973: 428).

name from doing it in any other way.[11] Are there, however, always appropriate ways to introduce a name? In almost any country, any human being has a legal name, which has been given to her or him according to a legal procedure; there is a second legal procedure, too, for changing that name. Yet, human beings often have other, non-legal names too, and many other non-human particulars have names. It is doubtful that there is a ratified procedure through which these other names are given to someone – the law, at most, legalizes some uses, as happens for many names of places. If we set aside what the law oversees, the appropriateness of a name becomes hard to define. We could maintain that a name is true of a thing just in case this is *called* by that name.[12] Is this a *condition*? Among names, there are nicknames, whose introduction and consumption is open, on the condition that it be understood who or what is so nicknamed.[13] A nickname may be made up on the spot. Ernesto's examples go far beyond Nunberg's cases,[14] and include "Leave me alone", "Silly", "Indeed", "I" – not anchored to a pragmatic function. Even "Brrr", which would show the emotional reaction of the name-giver, might be an appropriate (nick)name for a person whom the speaker finds icy.

> Ultimately, an expression acquires the property of being a name of an individual because of something that is done to it or with it: being used as (being made into) a quite arbitrary (and hence a-contextual) linguistic tool for reference to an individual (Napoli 2015: 213).[15]

[11] In the *Cratylus*, Socrates already shows concern for true and false names. Appropriateness introduces a normative aspect. Wittgenstein seems to have considered norms with some of his serious irony: "I can't be making a mistake; and if the worst comes to the worst I shall make my proposition into a *norm*" [Italics mine] (1969: 634).

[12] This is, again almost literally, Fara's *Being Called Condition*: "'*N*' (when a predicate) is true of a thing just in case it is called *N*" (2015: 64).

[13] The same word "nickname" originates from a mistake involving late Middle English: from *an eke-name* (*eke* meaning "addition"), misinterpreted, by wrong division, as *a neke name*.

[14] See, for instance, Numberg 1993: 27, fn. 28 and Numberg 2004. Nunberg's deference of meaning develops and extends an idea of Fauconnier (1985).

Ernesto distinguishes names only by syntactic means, without resorting to pragmatics.

[15] Notice how close this idea is to the origin of names mentioned at the opening of this section.

Ernesto, who argues that names can be neither properly characterized nor listed, concludes that "[a]ll that is necessary and sufficient for an expression to be a name of an individual is for it to be used as a name of an individual" (214), which is a possible reading of Fara's vague condition "being called by that proper name". Anything goes but what someone might challenge in a talk exchange. An extreme case of a name that works just because of how it is used, with no other dimension of appropriateness, is the name of a fictional character. In *Zeno's Conscience*, Zeno is just so called, with no introduction. Yet, calling n a particular is in general not enough to endow it with a name. Matushansky (2015: 337) writes that giving a name to a particular relates a phonological string to the particular – that is true, yet not sufficient to turn the string into a name.[16] First, as I have already claimed, the name can be contended; second, not all the phonological strings related to a particular are its name. Think of honorifics, for one instance. In Italy, any person who is or was President of the Republic is called "President" for life. Yet "President" is not the name of such people.

Be that as it may, if names are nouns, perhaps we have two kinds of nouns rather than names on one side and nouns on the other. Calling them collectively "nouns" highlights their similarities – a predicate may apply to indefinitely many particulars, a name can be shared by indefinitely many particulars. To suggest that there are two kinds of noun is to point at their dissimilarities. A predicate categorizes and its attribution is not arbitrary; a name does not categorize and its attribution is arbitrary. Learning a noun, I learn how to apply it in indefinitely many cases, but learning a name, I learn how a single particular is called. If a name is a predicate, and I am right, we have substituted two grammatical categories (name and predicate) for two different kinds of one category (predicate).[17]

2. In any PV, a striking aspect is the examples offered as central to understanding names – "There are relatively few Alfreds in Princeton", "An Alfred Russell joined the club today", "The Alfred who joined the club today was a baboon", "Some Alfreds are crazy; some are sane" (Burge

---

[16] See Gray 2014: 214.

[17] This was something Ernesto and I tangentially argued in Leonardi and Napoli 1995.

1973: 429). The idea that "Alfred" here might be a predicate derived from a proper name is almost never explored, though we have names derived from nouns, "Alfred" derives from a compound of the Old Saxon forms "alf" and "rath", meaning respectively "elf" and "counsel", and have nouns derived from verbs ("run" and "runner" from "run", for instance). Marginal cases can be revealing of how something works, as with optical illusions and visual perception. Are those examples revealing for understanding names? Jeshion (2015, 2015a, 2015b) has vigorously discussed their role, and I have nothing to add to her points. Claiming that it is ambiguous as to whether "Alfred" is a name or a noun (as in "few Alfreds", "an Alfred", "the Alfred", "some Alfreds") is close to the polysemic view sketched by Leckie (2013), a view that must be refined to be properly evaluated. Here, I shall rather develop some arguments critical of a *that*-PV and some other arguments critical of a *the*-PV, i.e., PVs claiming, respectively, a name to come with a possibly implicit demonstrative or with a possibly unpronounced determiner "the". My preferred *that*-PV is Burge's and my preferred *the*-PV is Fara's.

Burge (1973) argues for a predicative account of proper names, whose referential function is attributed to an implicit or, more rarely, explicit demonstrative.[18]

> [Proper names] play [...] the roles of a demonstrative and a predicate. (432)
>
> [O]ur predicate approach provides a single satisfaction rule [for the name], plus the set of primitive reference clauses applicable to all occurrences of demonstratives (implicit or explicit) in sentences. (438)

The use of a demonstrative is generally contrastive.[19] Leaving aside for the moment anaphoric uses of demonstratives, the contrast may be explicit – "I want that book, not this one" – or implicit – "I want that book". I would say the second, for instance, when there are many books around, in order to pick out the book I want to buy. Using an explicit demonstrative may signal distance – for example, because the person named is not liked by the speaker: "This Max wants to see me again".

---

[18] The demonstrative in Burge's formal theory is represented by a variable. A more sophisticated theory, which links directly proper names with variables, is Cumming's variabilism, which is not, however, a PV. (See his 2008). However, as Almog (this volume) writes, even assimilating proper names to variables is disputable.

[19] See Hinzen 2007: 209f.

If the person is an acquaintance and there is no contrast, no demonstrative goes with the name. In any case, can a substantial semantic role be attributed to an implicit demonstrative? Does an implicit demonstrative have an *occurrence*, or can it be *token* reflexive? The context determining the reference of a demonstrative is fixed by the circumstance of its utterance, that is, by its occurrence – but an implicit demonstrative does not properly occur. (*En passant*, Burge gives no account for the switch between an explicit and an implicit occurrence of the demonstrative.)[20] Moreover, can a demonstrative, even an explicit one, play that role? To the last question, my answer would be positive, but according to Burge, demonstratives do not by themselves grant a sentence a truth value:

> Apart from speaker-reference or special context, both
>
>> Jim is 6 feet tall.
>
> and
>
>> That book is green.
>
> are incompletely interpreted – they lack truth value (Burge 1973: 432).[21]

---

[20] Higginbotham comments:

> (10)   That Mary had fish for lunch
> (11)   Mary had fish for lunch
>
> On seeing a woman emerge from the seafood restaurant, and taking her for my friend Mary, I might volunteer either (10) or (11). Suppose that the woman is a Mary, but not my friend Mary. Then (10) is true if the woman had fish for lunch, but (11) is not verified thereby. (1988: 36)

Bach (2002) criticizes Burge's demonstrative account pointing out that "[w]hereas we might use 'that chair' to single out one chair from another, we would never use 'Jim Jones' to single out one Jim Jones from another" (91). Sawyer (2010: 214ff) counterargues Higginbotham's criticism, offering a pragmatic account based on two different saliences, a conversational and a contextual one. However, that implies a failure in Burge's project, because it shows that a fully satisfying, purely formal general account has not been given.

[21] Higginbotham comments:

> Acts of reference in the sense of Burge mark the fact that indexicals and demonstratives alike have no reference apart from contexts in which they are actually used. But no act apart from use is required to give an indexical a reference. (1988: 35)

Rather, it is formally reading pronouns as variables, and bringing in assignments, that demonstratives allow a sentence to have a truth value. If so, however, why not read names directly as variables?[22]

Burge's essay comes four years after Sloat's paper on "Proper Nouns in English", a 1969 paper the former quotes. Proposing what became relatively famous as Sloat chart,[23] which compares the syntactic behavior of nouns and names, Sloat offers a different solution, a *the*-PV, for proper

---

[22] This is exactly what Cumming does with his variabilism (see fn. 18). At least, *a posteriori*, it shows that there are more than the two only possibilities Burge envisages for a formal semantical theory with proper names – i.e., that they "play the role of constant, noncomplex singular terms [. . . or] that they play the role of predicates" (1973: 426). For variabilist theories of names, see also Yagisawa 1984 and Rami 2014.

A more sophisticated view involves variables (*à la* Burge, via Larson and Segal 1995, developed by Elbourne (2005)). A name is understood as having the following form: [[THE i] NP], in which the NP position is occupied by a name rather than a noun. As Elbourne writes,

> [o]ne should also note, of course, that in the current view names contain a free variable which can be assigned a referent directly. (177)

Variabilism, however, conflicts with the naïve understanding of names formulated by Napoli (2015: 213) quoted above in the text.

An aside on Burge's predicative view: his paper is ideological, connected with Davidson's philosophical program. One observes this in such statements as "theories of language should be no less general and precise (where feasible) than mathematical or physical theories".

[23] This is Sloat's chart, from Sloat 1969: 27:

A man stopped by.	A Smith stopped by.
*Some man stopped by.	*Some Smith stopped by.
Sóme man stopped by.	Sóme Smith stopped by.
Some men stopped by.	Some Smiths stopped by.
Sóme men stopped by.	Sóme Smiths stopped by.
Men must breathe.	Smiths must breathe.
The clever man stopped by.	The clever Smith stopped by.
The man who is clever stopped by.	The Smith who is clever stopped by.
A clever man stopped by.	A clever Smith stopped by.
The men stopped by.	The Smiths stopped by.
The man stopped by.	*The Smith stopped by.
*Man stopped by.	Smith stopped by.

"Some" and "sóme" are, respectively, the "some" used with mass terms and plurals and the one captured by the existential quantifier.

names accounted as *proper nouns*, writing the following rule in full (kept in the original capitalization here):

THE DEFINITE ARTICLE WILL APPEAR AS ZERO BEFORE SINGULAR PROPER NOUNS, EXCEPT WHEN IT IS HEAVILY STRESSED OR THEY ARE PRECEDED BY RESTRICTIVE ADJECTIVES OR ARE FOLLOWED BY RESTRICTIVE RELATIVE CLAUSES. (1969: 28)

This solution has been preferred recently by most linguists and philosophers supporting the predicative view of names.[24] Higginbotham (1988), Larson and Segal (1995), Thomsen (1997), Bach (2002), Matushansky (2006a, 2008, 2015) discuss the issue, with Matushansky examining a rich evidence from many languages and developing some syntactic arguments in support of *the*-PV. Fara advocates the most philosophically sophisticated version of *the*-PV.

Fara's analysis of names has a *being called condition*, which I quoted above in fn. 12 and repeat here: "'N' (when a predicate) is true of a thing just in case it is called N", adding that

in every case where a word occurs as a name, it is a predicate with (potentially) multiple application. The condition of its application is given by the being-called condition. (2015: 64)

From Matushansky, Fara derives a syntactical argument for reading names as predicates, arguing that in

   My parents called me Delia

"me Delia" is a small clause, and that in the small clause, "Delia" is the predicate and is not quoted, as it is in "bearer of '*Delia*'". A name gains the meaning it has by being used the way it is. The syntactic argument is relevant for investigating the predicative role of the name, which comes to mean what it does by being used as it is, with no need of any introduction with *to call* or anything similar. Fara, moreover, proposes that a name, "N", is semantically equivalent, though not identical, to "called N", equivalent in turn to "bearer of 'N'".[25] In this way, Fara avoids

[24] Larson and Segal offer a criticism of Burge 1973 and express a preference, among PVs, for *the*-PV, without endorsing it (1995: 353–355).

[25] Notice that Fara thus shifts the central role of the argument for names as predicates from a set of examples, as did Burge (1973), to a grammatical issue concerning names and independent from determiners, quantifiers, plurals, etc.

any formal circularity present in "N is the particular called 'N'" or in
"N is the particular that is the bearer of 'N'" – which is the complaint
Kripke (1972) has with Kneale 1962.[26] Hence, by the small clause anal-
ysis, Fara avoids two traps because in the small clause, the proper name
itself occurs and not its metalinguistic quote,[27] and in the condition for
the appropriateness of a name, she does not resort to metalanguage. A
caveat, however: if "Delia" is a predicate in "My parents called me Delia",
that does not establish by itself that "Delia" is a predicate in any other
occurrence.

Yet one may object to the small clause analysis on several grounds.
Fara makes a full parallel between

> My parents called me Delia

and

> My parents called me stubborn.

Regarding the first sentence, Biro remarks that

> saying "is called N" is a way of displaying – mentioning, if you like – "N", just
> as much as is enclosing "N" in inverted commas. Inverted commas are just
> one way of indicating that one is mentioning, rather than using, an expression.
> (2012: 290)

Dropping the inverted commas "may look awkward, conditioned as we
are by the standard philosophical usage" (292). The small clause does
not allow a predicative reading, he adds, if a name occurs in it.[28]

---

[26] Kripke, however, was not attacking PV, but rather the idea of fixing the reference
of a name by a descriptive condition such as "the man that I call 'Glunk'" (see Gray
2014: 211).

[27] However, see Biro 2012.

[28] If "Delia" is a predicate in "My parents called me Delia", it is supposed to agree in
number with the subject. Yet, it seems that there is not such an agreement in general.
Consider the following example, which describe an extreme case, *à la* Foreman (see
Bach 2002: 93):

> Susanna e Giorgio hanno chiamato tutte le loro figlie Anna.
> Susan and George called all their daughters Anna.

Predelli (2015) too contends some analyses of language offered by PVs, giving, for
instance, comparative counterexamples to the suffixing of names for number agree-
ment.

The difference can be stressed contrasting

   My parents named me Delia

with

   *My parents named me stubborn.[29]

An ambiguous passage emerges in Fara's discussion about the incompleteness of definite descriptions and proper names. She writes:

(87)   Olga enjoyed the party.
[. . .]

If you were to sincerely utter (87), you would be talking about some one particular party—call it BASH—and saying of *it* that Olga enjoyed it. (2015: 97)

Here, Fara's incidental remark "call it BASH", affixed to "talking about some one particular party", has a small clause format, but the supposedly name predicate is stressed as if it were quoted, inviting a NV reading – since in her small clause argument no stress and no quote were involved.[30]

   Finally, whereas the *that*-PV does not account for when the demonstrative is implicit, rather than explicit, the *the*-PV does not provide a *reason* for unpronouncing the determiner. Sloat's chart comparing proper names and nouns contains two problematic lines concerning the occurrence of the determiner:

   The man stopped by.   *The Smith stopped by.
   *Man stopped by.   Smith stopped by.[31]

---

[29] In Italian, the translations of the two English sentences diverge:

   I miei genitori mi hanno chiamato Delia
   *I miei genitori mi hanno chiamato testarda
   *I miei genitori mi hanno detto Delia
   I miei genitori mi hanno detto testarda.
   *I miei genitori mi hanno dato della Delia
   I miei genitori mi hanno dato della testarda.

[30] Fara quotes Donnellan 1966 at the end of the paragraph.

[31] In the chart, there are some doubtful parallels, one of which is noticed by Sloat himself. The preferred understanding of "Smiths" is "members of the Smith family", which seems to me the case not only in "The Smiths stopped by", but also in "Some

Perhaps a motive for dropping a determiner can be offered. In vocatives, any name or noun occurs without a determiner. Indeed, a vocative is always a referential use *à la* Donnellan. There is a person there and the addresser assumes that the vocative – be it "Mary", "Lady", "Professor", "President", "Dear", "Honey", etc. – pick that person out. The general occurrence of names without determiners depends on names' univocal identification of their bearers. So we can say not only "Alfred" but also "Young Alfred", without using a determiner, and we say "the young Alfred" or "Alfred, the young" where "the young" is claimed to be restrictive (is it really so?), because there is an issue in identifying of which Alfred we are speaking.

See the multi-part example below:

Alfred,
Alfreds,
Smith,
Smiths,
Professor,
Lady,
Man,
President,
Kind man,
Wise and kind lady,
*The President,
*The wise man,
*The wise and kind lady,
*This President,
*This wise man,
*This wise and kind lady,

– can you help me?[32]

Smiths stopped by". I do not know whether the case noticed by Sloat, and some others, can be read how Sloat does, i.e., as Smiths belonging to different families. Yet the problem could be easily solved substituting "Alfred" for "Smith", and then having "Alfreds" instead of "Smiths".

[32] In Ancient Greek, ὤ, which prefixes vocatives, is an allocative particle rather than an article. See Scott 1903, 1904 and 1905 and Lepre 2000, all quoted in Donati 2008.

Though descriptions too can be used referentially, only pronouns and names, in non-vocative contexts, go without determiner. I shall come back to this issue in the next section.[33] True vocatives are to be distinguished by fake ones, which do not summon but have rather a descriptive function. An instance of a fake vocative is "Au travail, les filles!" ("The girls, to work!"), which *describes* which persons in the audience have to go to work. (Cf. Espinal 2013: § 3 – the example is hers, see p. 119.)[34]

Last, Fara's account, as I mentioned, cannot be charged of being circular as Kripke (1972) thinks to be Kneale 1962's view – "'N' (when a predicate) is true of a thing just in case it is called N" is neither formally nor substantially circular. That was already true of Burge's account. Burge and Fara substitute the predicate "called N" with "N" itself. "N" is equivalent to "called N" or "named N". In none of these does "N" occur in quotes. It is by using "N" that "N" comes to name a particular. That a particular is named "N" is a metalinguistic reflection, as it is telling that naming links a particular and a phonological string. The circularity for Kripke anyway is not a problem, because it

---

[33] Almost all the modifiers of a name can be offered in a non-restrictive format, as appositions. The hardest case is that of "the famous Marc Jacobs", but I believe that to depend on "famous" – one is not famous because of his name, but because he is a scholar, a politician, a businessman, a movie director, etc. This and what I claim in the text put in doubt the idea that

> [w]hen names appear as bare singulars in argument position, they constitute the predicative component of a *denuded determiner phrase*, a determiner phrase with an unpronounced determiner. (Fara 2015: 60)

[34] Fara's argument, as I have already acknowledged, is generally brilliant and accurate. Yet, it is, I think, biased. To give just one specific instance: Fara discusses at length the issue of null versus non-null "the + modifier + name", which has two peaks, in her generalization and in her claim to produce some counterexamples. According to the first, "[w]hen a singular name is preceded by a modifier, the definite article must appear as overt 'the' whether the modifier is used restrictively" and if used non-restrictively "it can also occur with overt 'the'" (2015: 83–90). In addition, consider that, (a), at least in some languages an honorific, which acts restrictively, like "Miss" in "Miss Jones", does not require the determiner; and (b), this is taken to go with the PV outlined, but it was never denied that a determiner phrase can be a composite out of names and other terms – whether you understand the name as a name or as a noun.

shows that naming, as he maintains, is a primitive – but that naming is a primitive is exactly what PVs deny.[35]

3. Names, both if predicates and if referring expressions, seem to face the *Arabian Nights* robber problem. Mill cites the *Arabian Nights* story in which a robber marks with chalk a house so that he can identify it again, and adds:

> Morgiana chalked all the other houses in a similar manner, and defeated the scheme: how? simply by obliterating the difference of appearance between that house and the others. The chalk was there, but it no longer served the purpose of a distinctive mark. (Mill 1843: 41; quoted in Katz 2001: 148)

Naming a particular, we introduce a 'mark' with a similar purpose.[36] Marking out is occasionally jeopardized by other particulars marked the same way, and PVs take into account that the marking comes with no exclusive right. To overcome the problem, PVs explicitly or implicitly resort to indexes – explicitly, Burge and Elbourne and, implicitly, Fara, who views the context as adding the information to make the mark individuative.[37]

With signs, which are artificial features, there seems to be no principled way out of Morgiana's trick.[38] If my wife and I introduced a unique proper name for our daughter, any other parents could choose to give the same proper name to their daughter. Is it a problem to share one's

---

[35] Gray 2014 is entirely dedicated to the circularity issue, reallocating it to presuppositions about the extension of a (name) predicate – more exactly that x satisfies the name predicate N *iff* the group of people speakers interact with presuppose that x satisfies N. (Cf. especially p. 227.) On the circularity objection see also Loar 1976 and 1980, Recanati 1993 and Elbourne 2005.

[36] "Signifying" comes from Latin, *signum facere*, which translates almost literally to "mark out".

[37] In general, it is language that adds something to context, and not the other way around. (See Leonardi 2013.)

[38] Bach (2002: 93) quotes the case of George Foreman, the great boxer, who had five sons and called all his sons "George". Actually, he called them George Jr., George III, George IV, George V and George VI, and they, George Jr. aside, had all nicknames – Monk, Big Wheel, Red, and Little Joey. The great boxer wanted to call all his sons after himself, and failed twice – he had to add counters, and the family used nicknames, showing that when needed, we introduce a naming solutions capable of distinguishing.

name with others or very many others? In practice, it is not. There are indefinite many ways to supplement our proper name, like family names or the proper names of the parents (Vittoria's and Michael's daughter), a reference to time (to age, for instance, "the older", "the younger"), to places (from Venice, from Edinburgh), etc. These additions can each require a further supplement. Time exploits points of reference – *now*, Christ's year of birth, or the year when Muhammad migrated from Mecca to what later was called Medina, or the *Anno Mundi*, or the Year after Creation. Giving each particular a brand new name would be cognitively overwhelming, if not *tout court* impossible. PVs take the problem into account, as I acknowledged, but are their solutions simpler or more efficient than NVs'? I already called into question whether they are in fact simpler, suggesting that they substitute the name kind with a second predicate kind. In addition, instead of a phrase constituted by a single name, the phrase contains a determiner or a demonstrative plus a name predicate. Formally, this change unifies (the predicate element) and adds a tool (the determiner / the demonstrative) to find out which particular is named among the many particulars that share the name. In comparison with the communication chain of NVs, how much better does the PVs' determining tool manage to pick out a named particular? Even if the tools are different, their workings are basically the same as the communication chain of NVs. Tracing back to Plato via a demonstrative chain, or an assignment, we would use a chain connecting with people in direct contact with Plato.[39]

That may look better than claiming that names are ambiguous – an idea Kripke considers but does not endorse. Lexical ambiguity is an external acknowledgment of the problem, which has an external solution in the communicative chain. An ideal dictionary would disambiguate by listing for any name as many distinct items as many distinct particulars are so named, and adding subscripts to distinguish the homographic names. Yet PVs' improvement depends on having offered a schema, with no instruction about how to yield what is supposedly demonstrated, whereas a dictionary, or encyclopedia, of names shows individual solutions, listing an indefinite number of items, or an open list, as an encyclopedia

---

[39] The communication chain is schematically reproduced in a D-theory anaphoric chain, or simulations of such an anaphoric chain. The D-theory is, for instance, Elbourne's choice, in a format close to that of Heim 1990.

requires. There are other alternatives that are simpler than Burge's view
and formally clean, such as directly understanding names as variables.
Fara's proposal is more sophisticated than Burge's. It does not require
any implicit variable – the variable being the context of use, with no
trace in the linguistic representation. The solution here is external, as it
is in Kripke. If Kripke's proposal is no solution, Fara's is not either. Our
names do not individuate us, and our name-predicates do not either. The
appeal to contextual indices is proper in both accounts. Complemented
with predicates derived by names – Kripke is a Saul, because Saul is
his name – NVs look just as simple as PVs. A different look at names
could suggest that it is simpler. This view uses naming as a sophisticated
distributed coding device, capable of producing strings that individuate
each particular.

Ernesto and I once discussed, what is our name? Is his name *Ernesto*
or *Ernesto Napoli*? If we regard the whole system of naming (persons,
places, events, etc.), in many cases we can claim, for instance, that the
proper names in our family are the minimal differences added to a dis-
tributed code to distinguish each individual member within the fam-
ily. The combined names of our parents serve as the minimal difference
to pick out our family in contrast to that of our cousins, and so on.
Looking at our identity card we can reconstruct the following code for
each of us: ProperName(MotherFamilyName)FatherFamilyNameCityof-
BirthDateofbirthCountryofBirth(CityofResidenceCountryofResidence) –
where elements within brackets are optional.

"Ernesto" is the last minimal difference distinguishing Ernesto in the
intersection of his mother's and his father's families, i.e., the nuclear
family he comes from, and accordingly we call it his *proper* name.[40] Out-
side the family, using only the proper name of a person is a pragmatic
way to claim to be familiar with the namee. Each of those particulars
(persons, families, dates, cities, etc.), have a proper name, and by assem-
bling proper names of different things we yield a code unique to each
particular. This coding system is easy to run. Legal procedures come
out of natural ones – our legal name, the one on our ID, standardizes a
natural naming practice. When we deal at once with many Annas, we
distinguish them by resorting to parts of their distributed name – we

---

[40] A proper name of a person is then the minimal difference which in the last named
entity distinguishes the subentity from any other one.

speak of Anna Stewart and Anna Laurent, Anna S. and Anna L., Anna senior and Anna junior, or Anna from Belgium and Anna from Australia, etc. An analogous coding system exists for addresses. I live in Country-CityStreet/SquareNumber, the number being the last minimal difference distinguishing my house in the street or square I live in. Just as proper names do not assemble people in a natural or even artificial way, so is the case for address numbers, which as names are labels. The code system combines names of points of reference – places, dates, families – to come to individuate persons by proper names. We name persons because they too are points of reference. The code system of addresses combines places that are points of reference – country, town, street – and then individuate houses by numbers. Sharing a proper name is no different than sharing the street number in an address. It is like sharing the head digit or the tail one in a code number.

In addition, with addresses as with names, there are remedial practices if we miss some parts. If we forget the name of the street, we can say something like "the first street south, parallel to High Street". If we do not remember the address number, we can say "the second house on the left after the grocery, going east". If we forget the town, we say, "the closest city north of Edinburgh". Here, we use names that resort to denoting a street, a house, or a city. If our audience is unsure of our reference, we can speak of Anna Smith Cooper as "Anna, the daughter of Victoria" or "Anna, the daughter of Michael". Both systems – names and addresses – are extremely efficient and cognitively light to run.

As a last turn, I would hint, no more than hint, at that non-predicating predicates can be viewed themselves as names – a view Frege 1892a adopted investigating the concept *horse*. Imagine someone saying "Blue is the color of your sweater", "Blue is my preferred color", "That is the square", "The square is the form of that wood piece", "That is the tiger", "The tiger is the animal on the left", "Water is that liquid", "Friendship shows in their gestures", and so on. Consider the simplest case: The species tiger is not a tiger, so by indicating the species with "The tiger is the animal on the left", we are not predicating of it that it is a tiger; we are naming the species. If we distinguish between labeling a feature of things and categorizing a thing by a feature, we can yield a view of predicates as names proper to non-predicating predicates. "Blue" and "square" are names of features added to the features themselves to dis-

tinguish them (whatever features are); predication characterizes things by means of these features, a fact which cannot happen with names of objects, because objects do not characterize objects. In the examples above, if we preferred still to look at "blue", "square", "tiger" as predicates, we would have to say that these predicates are ambiguous, because "blue", for instance, has two different meanings in "That is blue" and in "Today the sky is blue". If I am right, Mill was wrong in claiming that the name "is said to signify the subjects *directly*, the attributes *indirectly*" (1843: 38). Notice that in some languages, including Italian, the determiner in this case is required to indicate that "blue", for instance, is not predicated.

Distributed names and predicates' names are connected. In an NV, by naming we fix points of reference that are our elements of concern, be they particulars or features, and thereby link words and things.[41] The predicate view by contrast seems to leave this issue up in the air. It is a formal, not a substantial, advance, and it cares for our representation more than for what we represent.[42]

References

Almog, J. (2012), "Referential uses and the foundations of direct reference", in J. Almog and P. Leonardi (eds.), *Having in Mind*, Oxford University Press, New York, pp. 176–184.

——— (2014), *Referential Mechanics: Direct Reference and the Foundations of Semantics*, Oxford University Press, New York.

Anderson, J. (2007), *The Grammar of Names*, Oxford University Press, New York.

[41] Both Mill and Austin stress this relevance; Mill does so specifically for names, which he deems to be "our clue to the things", bringing "at once before us all the distinctions which have been recognized, not by a single inquirer, but by all inquirers taken together" (1843: 29). Austin stresses that "our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations" (1956: 182).Yet, here I distinguish between highlighting particulars and highlighting features.

[42] I thank Sebastiano Moruzzi for commenting on a draft of my paper, Joseph Almog for some more cutting comments, and the three editors for the care they put in spotting where I had been sloppy.

Austin, J.L. (1956), "A plea for excuses", *Proceedings of the Aristotelian Society*, 57, pp. 1–30, reprinted in *Philosophical Papers*, Oxford University Press, Oxford 1979, pp. 175–204.

Bach, K. (2002), "Giorgione was so-called because of his name", *Philosophical Perspectives*, 16, pp. 73–103.

——— (2015), "Predicative view of names", *Philosophy Compass*, 10/11, pp. 772–784.

Biro, J. (2012), "Calling names", *Analysis*, 72, pp. 285–293.

Boer, S. (1975), "Proper names as predicates", *Philosophical Studies*, 27, pp. 389–400.

Burge, T. (1973), "Reference and proper names", *Journal of Philosophy*, 70, pp. 425–439.

Capuano, A. (2012), "The ground zero of semantics", in J. Almog and P. Leonardi (eds.), *Having in Mind*, Oxford University Press, New York, pp. 7–29.

Cohen, J. (1980), "The individuation of proper names", in Z. van Straaten (ed.), *Philosophical Subjects*, Clarendon Press, Oxford, pp. 140–163.

Crane, T. (1992), "Names, sense and Kripke's puzzle", *From the Logical Point of View*, 2, pp. 11–26.

Cumming, S. (2008), "Variabilism", *Philosophical Review*, 117, pp. 525–554.

Devitt, M. (1981), *Designation*, Columbia University Press, New York.

Donati, M. (2008), *La categoria del vocativo nelle lingue classiche: aspetti teorici, storici, diacronici*, PhD Thesis, Dottorato di Ricerca in Linguistica Sincronica, Diacronica e Applicata, Università di Roma Tre.

Donnellan, K. (1966), "Reference and definite descriptions", *Philosophical Review*, 75, pp. 281–304.

Elbourne, P. (2005), *Situations and Individuals*, MIT Press, Cambridge, MA.

Elugardo, R. (2002), "The predicate view of proper names", in G. Preyer and G. Peter (eds.), *Logical Form and Language*, Oxford University Press, New York, pp. 467–503.

Englebretsen, G. (1981), "Predicates, predicables and names", *Critica*, 13, pp. 105–108.

Espinal, M. (2013), "On the structure of vocatives", in B. Sonnenhauser and P. Noel Aziz Hanna (eds.), *Vocative! Addressing between Structure and Performance*, De Gruyter, Berlin/Boston, pp. 109–132.

Fara, D.G. (2011), "You can call me 'stupid'...just don't call me stupid", *Analysis*, 71, pp. 492–501.

——— (2015), "Names are predicates", *Philosophical Review*, 124, pp. 59–117.

——— (2015a), "'Literal' uses of proper names", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 251–279.

Fauconnier, G. (1985), *Mental Spaces*, Cambridge University Press, Cambridge.

Frederick, D. (2014), "Singular terms, predicates and the spurious 'is' of identity", *Dialectica*, 67, pp. 325–343.

Frege, G. (1892), "Über Sinn und Bedeutung", *Zeitschrift für Philosophie und philosophische Kritik*, 100, pp. 25–50, translated as "On sense and reference" in in *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black (eds.), Blackwell, Oxford, 1960, pp. 56–78.

———— (1892a), "Über Begriff und Gegenstand", *Vierteljahresschrift für wissenschaftliche Philosophie*, 16, pp. 192–205, translated as "Concept and object" in *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black (eds.), Blackwell, Oxford, 1960, pp. 42–55.

Geurts, B. (1997), "Good news about the description theory of names", *Journal of Semantics*, 14, pp. 319–348.

Ghomeshi, J., Ileana, P. and Wiltschko, M. (2009), "Introduction", in J. Ghomeshi, P. Ileana and M. Wiltschko (eds.), *Determiners: Universals and Variation*, John Benjamins, Amsterdam/Philadelphia, pp. 1–21.

Ghomeshi, J. and Massam, D. (2009), "The proper D connection", in J. Ghomeshi, P. Ileana and M. Wiltschko (eds.), *Determiners: Universals and Variation*, John Benjamins, Amsterdam/Philadelphia, pp. 67–95.

Gray, A. (2014), "Name-bearing, reference and circularity", *Philosophical Studies*, 171, pp. 207–231.

Heim, I. (1990), "E-type pronouns and donkey anaphora", *Linguistics and Philosophy*, 13, pp. 137–177.

Higginbotham, J. (1988), "Contexts, models and meanings: A note on the data of semantics", in R. Kempson (ed.), *Mental Representations: The Interface between Language and Reality*, Cambridge University Press, Cambridge, pp. 29–48.

Hinzen, W. (2007), *An Essay on Names and Truth*, Oxford University Press, Oxford.

Hornsby, J. (1976), "Proper names: a defense of Burge", *Philosophical Studies*, 30, pp. 227–234.

Jeshion, R. (2015), "Names not predicates", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 225–250.

———— (2015a), "A rejoinder to Fara's '"Literal" uses of proper names'", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 280–294.

———— (2015b), "Referentialism and predicativism about proper names", *Erkenntnis*, 80, pp. 363–404.

Jubien, M. (1993), *Ontology, Modality, and the Fallacy of Reference*, Cambridge University Press, Cambridge.

Kaplan, D. (1989), "Demonstratives", in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford University Press, New York, pp. 481–563, circulated as a typescript since 1971, and revised in 1977.

———— (1990), "Words", *Aristotelian Society Supplementary Volumes*, 64, pp. 93–119.

Katz, J.J. (2001), "The end of Millianism: multiple bearers, improper names and compositional meaning", *Journal of Philosophy*, 98, pp. 137–166.

Kneale, W. (1962), "Modality *de dicto* and *de re*", in E. Nagel, P. Suppes and A. Tarski (eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*, Stanford University Press, Stanford, pp. 622–633.

Kripke, S.A. (1972), "Naming and necessity", in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Reidel, Dordrecht, pp. 253–355, 763–769, reprinted in book format with a Preface, Oxford, Blackwell, 1980.

Larson, R. and Segal, G. (1995), *Knowledge of Meaning*, MIT Press, Cambridge, MA.

Leckie, G. (2013), "The double life of names", *Philosophical Studies*, 165, pp. 1139–1160.

Leonardi, P. (2013), "What language adds", in A. Capone, F. Lo Piparo and M. Caparezza (eds.), *Perspectives on Pragmatics and Philosophy*, Springer, Berlin, pp. 277–290.

Leonardi, P. and Napoli, E. (1995), "On naming", in P. Leonardi and M. Santambrogio (eds.), *On Quine*, Cambridge University Press, Cambridge, pp. 251–266.

Lepre, M.Z. (2000), "Le emozioni, la voce e i gesti. Le interiezioni in greco antico: un capitolo dimenticato", *Linguistica e Letteratura*, 25, pp. 9–45.

Loar, B. (1976), "The semantics of singular terms", *Philosophical Studies*, 30, pp. 353–377.

——— (1980), "Names and descriptions: A reply to Michael Devitt", *Philosophical Studies*, 38, pp. 85–89.

Longobardi, G. (1994), "Reference and proper names: A theory of N-movement in syntax and logical form", *Linguistic Inquiry*, pp. 609–665.

——— (2005), "Toward a unified grammar of reference", *Zeitschrift für Sprachwissenschaft*, 24, pp. 5–44.

Mathieu, E. (2009), "From local blocking to cyclic agree", in J. Ghomeshi, P. Ileana and M. Wiltschko (eds.), *Determiners: Universals and Variation*, John Benjamins, Amsterdam/Philadelphia, pp. 120–157.

Matushansky, O. (2005), "Call me Ishmael", in E. Maier, C. Bary and J. Huitink (eds.), *Proceedings of Sinn und Bedeutung 9*, NCS, Nijmegen, pp. 226–240.

——— (2006), "Head-movement in linguistic theory", *Linguistic Inquiry*, 37, pp. 69–109.

——— (2006a), "Why rose is the rose", *Empirical Issues in Formal Syntax and Semantics*, 6, pp. 285–308.

——— (2008), "On the linguistic complexity of proper names", *Linguistics and Philosophy*, 31, pp. 573–627.

——— (2015), "The other Francis Bacon: On non-bare proper names", *Erkenntnis*, 80, pp. 335–362.

Mill, J.S. (1843), *A System of Logic*, John W. Parker, London, reprinted by eBooks@Adelaide, Adelaide 2011.

Napoli, E. (2015), "Names as predicates?", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 211–224.

Numberg, G. (1993), "Indexicality and deixis", *Linguistics and Philosophy*, 16, pp. 1–43.

——— (2004), "The pragmatics of deferred interpretation", in L. Horn and G. Ward (eds.), *Handbook of Pragmatics*, Blackwell, Oxford, pp. 344–364.

Perry, J. (2001), *Reference and Reflexivity*, CSLI Publications, Stanford, CA.

Predelli, S. (2015), "Who's afraid of the predicate theory of names?", *Linguistics and Philosophy*, 38, pp. 363–376.

Quine, W.O. (1948), "On what there is", *Review of Metaphysics*, 2, pp. 21–38.

Rami, D. (2013), "On the unification argument for the predicate view of proper names", *Synthese*, 191, pp. 841–862.

——— (2014), "The use-conditional indexical conception of proper names", *Philosophical Studies*, 168, pp. 119–150.

Recanati, F. (1993), *Direct Reference: from Language to Thought*, Blackwell, Oxford.

Rieppel, M. (2013), "The double life of 'The Mayor of Oakland'", *Linguistics and Philosophy*, 36, pp. 417–446.

Russell, B. (1903), *The Principles of Mathematics*, Cambridge University Press, Cambridge.

——— (1905), "On denoting", *Mind*, 14, pp. 479–493.

Sainsbury, M. (2015), "The same name", *Erkenntnis*, 80, pp. 479–493.

Salmon, N.U. (1981), *Reference and Essence*, Princeton University Press, Princeton.

——— (1986), *Frege's Puzzle*, Ridgeview, Atascadero, CA.

Savage, H. (2014), "On being called names", MS, `https://www.academia.edu/3658560/On_Being_Called_Names`.

Sawyer, S. (2010), "The modified predicate theory of proper names", in S. Sawyer (ed.), *New Waves in Philosophy of Language*, Palgrave Macmillan, Houndmills/New York, pp. 206–225.

Scott, J.A. (1903), "The vocative in Homer and Hesiod", *American Journal of Philology*, 24, pp. 192–196.

——— (1904), "The vocative in Aeschylus and Sophocles", *American Journal of Philology*, 25, pp. 81–84.

——— (1905), "Additional notes on the vocative", *American Journal of Philology*, 26, pp. 32–45.

Sloat, C. (1969), "Proper nouns in English", *Language*, 45, pp. 26–30.

Thomsen, H.E. (1997), "On the proper treatment of proper names", *Nordic Journal of Linguistics*, 20, pp. 91–110.

Wettstein, H. (1984), "How to bridge the gap between meaning and reference", *Synthese*, 58, pp. 63–84.

Wittgenstein, L. (1969), *On Certainty*, Blackwell, Oxford.

Yagisawa, T. (1984), "Proper names as variables", *Erkenntnis*, 21, pp. 185–208.

A syncretistic theory of proper names

Alberto Voltolini

1. Introduction

In this paper, I want to show that, far from being incompatible, a Predicate Theory of proper names and the Direct Reference thesis can be combined in a syncretistic account. There are at least three plausible such accounts – one which compares proper names in their referential use to referentially used proper definite descriptions, another one that compares them in this use to demonstratives, and a third one which, although it is as indexicalist as the second one, conceives proper names in this use as a *sui generis* form of indexicals, *indexinames*. Finally, I will try to give both technical and substantive reasons as to why the third account is to be preferred to the other two.

2. The incompatibility between Predicate Theory and Millianism

A Predicate Theory of proper names claims that such expressions have the same kind of meaning as general terms, so that they may formally count as predicates, namely as expressions that predicate a predicable (a property) of something else (Burge 1973: 428–429). If this is the case, the theory may well account for the so-called *predicative* use of proper names,

the use that occurs when a proper name figures in non-argumental positions of a sentence, as in the following cases:

(1)   I'm happy to be an Alberto;
(2)   The Voltolini I'm curious about is not myself;
(3)   There is an Alberto Voltolini in Italy that I don't know.

For the proper names involved in such positions seem to predicate a property of something else, as Burge (1973: 430) originally envisaged.

As such, this theory is clearly incompatible with Millianism, *if* Millianism is the position that radicalizes in a certain way the so-called Direct Reference thesis.[1] As is well known, the latter is the thesis that the meaning of a proper name ends by collapsing onto its referent, i.e., what the name stands for. For Millians, proper names are counterparts in non-formal languages of individual constants. In so doing, from a semantical point of view they clearly privilege the *referential* use of a proper name in an argumental position, the use in this position where the proper name purports to designate something.[2] They say that in that use, not only i) the *main* semantically relevant contribution of a proper name is its truth-conditional contribution, i.e., its referent (if any),[3] as any direct referentialist maintains,[4] but also ii) that contribution is the

---

[1] I stick to the traditional formulation of Millianism (see below). In presenting an alternative formulation of Millianism as the view according to which a proper name "is endowed with a constant character – the function that determines a particular individual as its content at any [context] c", Predelli (2015: 369) may disagree on this assessment of Millianism as a radicalization of the Direct Reference thesis.

[2] In this paper, I refrain from considering another way in which a proper name may be used in argumental position, an *attributive* use (as when one utters "Alberto, make a step forward!" to mean that whoever is so named has to make such a step). If there were such a use, this would be further evidence for a Predicate Theory, especially if combined with the Direct Reference thesis as I do below. For in such a case, the truth-conditions of the sentential token involving such a use would be (*inter alia*) constituted by (roughly) the property that is predicated in the name's predicative use.

[3] This way of putting things notoriously raises the problem of how to account in direct referentialist terms for the truth-conditional contribution of empty names, those names who lack a referent. I won't deal with this problem here. For a proposal, see my Voltolini 2014.

[4] By putting things this way, it follows that a proper name is rigid, i.e., it designates the same individual across all possible worlds. For the name has the same referent, *qua*

only semantically relevant contribution – that referent is given via *no* meaning mediators.[5] For in the Millianist perspective, not only from a semantical point of view the predicative use of proper names is utterly split from their referential use, but there is no chance for a name in the latter use to be understood as counting as a predicate.

An evidence in favour of Millianism is that it is hard to see how a proper name in its referential use may be taken as counting as a predicate. The role the name has in referring to something cannot be a predicative role, not even if that something were the property a corresponding predicate predicates. In such a case, the name would name the property, it would not predicate it (cf. Napoli 2015: 217, 219).

The answer predicativists typically give to this problem is that in its referential use, a proper name involves a predicate of the kind "to be (an individual) called 'N'", where "N" is a schema for the name in question.[6] For reasons that will be clear in Section 5, I prefer this formulation to the merely seemingly analogous formulation of the predicate in question as a predicate of the kind "to be a bearer of 'N'". Yet this answer can really account for the referential use of a proper name only once a Predicate Theory is nested into the Nominal Theory of proper names. According to this theory, in its referential use a proper name "N" is a schema for involves the above predicate insofar as it is synonymous with the nominal definite description of the kind "the individual called 'N'".

Clearly enough, the Nominal Theory may well accommodate the predicative use of a proper name. In saying something of the kind that some-

its truth-condtional contribution, across all worlds. By following Marti (1995), I put thesis i) in terms of a name's having its referent as its truth-conditional contribution rather than in the both ontologically and metaphysically more demanding terms of the name's referent as (*inter alia*) constituting a singular proposition, the structured entity made of individuals and properties to be evaluated throughout possible worlds.

[5] See also Napoli 1995. For a similar way of formulating the Millian claim see Marti 1995, which distinguishes between a truth-conditional and a cognitive sense of "direct reference" and claims that Millianism defends both. I have not appealed to that formulation, for one may say that, once that cognitive sense is turned into a normative sense, Millianism has to stick only to the first, truth-conditional, sense of "direct reference" just as any direct referentialist. On this see Voltolini 2004. The formulation of Millianism Predelli gives in (2015) (see fn.1) may be interpreted as going along this direction.

[6] This idea is sketched in Burge 1973; it returns (in different forms) e.g. in Bach 1981, Fara 2011.

one is a N, what we really mean is that that very someone is an individual called "N", so that in such a use the property of the kind *being called "N"* is what the name the schema "N" is for predicates of the relevant individual. Thus, the above sentences turn out to say respectively the same as

(1a) I'm happy to be an individual called "Alberto";
(2a) The individual called "Voltolini" I'm curious about is not myself;
(3a) There is an individual called "Alberto Voltolini" in Italy that I don't know.

Yet such a theory accounts for the referential use of a proper name in a way that is still incompatible with both Millianism and the Direct Reference thesis, *if* the nominal descriptions in question are taken to semantically work *à la* Russell.[7] For then not only in that use the only semantical contribution of a proper name is not its referent, but also such a referent is not the name's main semantical contribution. For the semantical contribution of the nominal description synonymous with the name is rather given in terms of its contextual definition yielded by the Russellian paraphrase of the sentence involving it, whose truth-conditions notoriously appeal to properties and (the meaning of) connectives, not to the designation of the name.

This theory raises the problems Kripke (1980) notoriously put forward: proper names seem to be synonymous with definite descriptions neither in general, nor when such descriptions are nominal descriptions of the above kind. Yet independently of Kripkean worries,[8] one may put into question the idea that lurks behind endorsing such a theory, i.e., the idea that a proper name involves a predicate of the kind "to be (an

---

[7] In what follows, I equate the Nominal Theory with a nominal descriptivist theory of proper names, along the lines originally proposed by Kneale (1962). By itself, this is not exactly the case – the Nominal Theory limits itself to saying that a determiner equivalent to the definite article is a phonologically covert element of a sentence containing a referentially used proper name (cf. Leckie 2013: 1141–1142). Yet unless it opts for a nonstandard, e.g. a Montagovian, reading of a definite description, Russellian descriptivism naturally suggests itself as the correct way of cashing out the Nominal Theory. Unless one wants to read the Nominal Theory *à la* Predelli (2015); yet in such a case the determiner figures as a covert element that qualifies the constant character of a proper name, not its truth-conditional contribution. See below.

[8] I tried to answer such worries in my Voltolini 1995.

individual) called 'N'". To put things in this way is clearly tantamount
to giving up the original idea a Predicate Theory wanted to defend that
a proper name formally counts as a predicate. For in the above predi-
cate the name occurs as a quoted constituent. Cf. Napoli 2015: 222. So,
what is the legitimacy of such a move? If the reason is just to provide
a unified account both of the referential and of the predicative use of
proper names, a Millian may well endorse Jeshion's (2015) perplexities:
better to deny that there is a uniformity between the two uses and to
provide an alternative explanation of the latter use, so as to ultimately
reject that account.

In point of fact, Millians have tried to provide such an alternative ex-
planation, for instance by appealing to pragmatics, as Napoli (2015: 222–
223) does. There are many non-literal uses of proper names, both in the
referential and in the predicative use, which neither Millians nor pred-
icativists are able to account for. Consider

(4)   Aristotle is on the top shelf,
(5)   There are three Aristotles on the top shelf,
(6)   He is no Einstein,

taken to mean, respectively, that a certain book authored by Aristotle
is on the top shelf, that three such books are on the top shelf, and that
a certain person is not as clever as Einstein. Once one allows for such
non-literal uses,[9] why not simply saying that all predicative uses are
non-literal?[10]

Yet a predicativist may reply by noting, first, that in the referential
use, non-literal uses are such insofar as they somehow depend on other
uses, which thereby are the literal ones. As Fauconnier (1985) originally
stressed, there is a pragmatic function from authors to their works, or
in other and more general terms, a pragmatic local process, notably a
metonymical one (as Recanati (1993, 2004) would say), that enables a
referential transfer from the author of the *Metaphysics* to one of his books
to occur as far as "Aristotle" is concerned in (4). Yet if this is the case,
second, why can't there be an analogous pragmatic process that allows
a name in a predicative use to shift its meaningful contribution from

---

[9] Burge (1973) claims that, unlike (1)–(3), (4)–(6) involve non-literal predicative uses
of proper names.

[10] For similar examples and worries, cf. Jeshion 2015: 238ff.

the property it originally predicates to another property?[11] As Recanati (2004) has shown, this certainly happens when one utters

(7)   That statue is a lion

meaning that the statue over there is not a lion, but a lion-representation. In such an utterance, the meaning of the predicate "to be a lion" is modulated by a local pragmatic process (metaphorical in this case) that leads the literal meaning of that predicate to shift to its non-literal one.[12] So, why cannot there be an analogous shift to predicating the property of *being an Aristotelian work* (in (5)) and the property of *having an Einstein-like character* (in (6)) in virtue of some local pragmatic process or other (a metonymical one in the case of (5), a metaphorical one in the case of (6))? But then a shift having the above properties as their respective targets may well take the property of *being called "Aristotle"* and *being called "Einstein"* as (*inter alia*)[13] their respective sources, as a Predicate Theory ultimately claims. If this is the case, then not only (5)-(6) display non-literal predicative uses of proper names, but such uses are such for they *depend* on another use, precisely the predicative use of such names (1)-(3) display. The latter use is therefore the literal, hence the privileged, predicative use[14] that a Predicate Theory can account for.[15]

---

[11] By so considering (4)–(6), for the purposes of this paper I remain neutral as to whether such cases involve a truth-conditionally relevant pragmatics, as Recanati (2004) among others (including myself) believes, or a proper post-semantical pragmatics, as followers of Grice (1975) believe.

[12] For many other such examples, cf. Fara 2015.

[13] I say "inter alia", for the pragmatic function must here take as argument both the relevant nominal property and the fact that a certain individual is the ordinary (or at least the most relevant) referent of the proper name in question in order for it to take as value the relevant non-literal property. Jeshion (2015: 245–246) seems to consider a variant of this predicativist reply, yet she does not seem to see the problems that it raises for her own criticism of predicativism.

[14] *Pace* Jeshion (2015: 241–245). In (2015a: 288), Jeshion adds that when a sentence having a proper name in non-argumental position is used to say of something that it is called by that name, as in (1)–(3), this use is as derived as the use that occurs in (5)–(6). Yet this is not the case. For the relevant local pragmatic process operates in (5)–(6) by taking a property of the kind *being called "N"* as its source, whereas as to (1)–(3) there is no further property with respect to which that very property would be its target.

[15] In (2015), Fara defends a similar account. Yet she does not seem to see, first, that *all* the examples of non-literal predicative uses she provides – i.e., both examples of

At this point, in order to attack a Predicate Theory by still appealing to a pragmatic, or at least to a pragmatically-based, account of the predicative use of proper names one may appeal to the following move, as Leckie (2013) does. The very privileged predicative use of a proper name that (1)–(3) exhibit, one may say, is precisely a target having the name's referential use as its source, either in virtue of some pragmatic process or better because of the lexicalization of that process: the process is turned into a lexical rule that connects the predicative use of a proper name with its referential use. Yet this move doesn't seem to work. First, there seems to be no evidence of such a pragmatic process; as I have just said, the privileged predicative use of a proper name is the *literal* one. To be sure, this literality is compatible with the idea that the pragmatic process is lexicalized, as Leckie ultimately holds. Yet second, as we have seen before the pragmatic processes such as the ones Leckie herself appeals to in order to make her proposal plausible are *local* processes; namely, they

---

type (5) and examples of type (6) – may be ranked as property shitfs induced by local pragmatic processes, and second, that by appeal to such processes one can rank as non-literal predicative uses of proper names even those uses that she instead ranks as literal predicative uses involving different predicates having to do with family membership (e.g., "Waldo Cox is a Romanov" when used to mean that Cox the gardener belongs to the well-known imperial family) (for a similar criticism, see Jeshion 2015a). Fara thinks so for she believes (along with Jeshion (2015a)), that family names are proper *nouns* (2015: 268). But alleged proper nouns are just proper names of *types*, as any name of a genus or a species. Cox is a Romanov just as he is a Homo Sapiens, the species referred to when e.g. one says "Homo sapiens is the only surviving species of hominids". (Incidentally, the latter sentence shows that a type can be referred to by a proper name even without prefixing a definite article to the name. Thus, it also shows that the supposed syntactic evidence that should show when referring to a type, a purported proper name is just a proper noun (cf. e.g. Jeshion 2015a: 290), is rather controversial. As Northern Italians well know, even the opposite syntactic evidence to the effect that proper names in a referential use are not prefixed by the definite article is rather controversial.) Thus, a local pragmatic process may well involve a shift from the property of *being called "N"* (plus the fact that a certain type is the ordinary, or at least the most relevant, type referred to by that name) to the property of *being a member of the type called "N"*. Indeed, if you address me by saying that I am a Mussolini, you may mean not only that I am so-called, but also either a) that I have a Mussolinian physiognomy viz. that I am an instance of a certain *physiognomical type*, or b) that I have a Mussolinian character viz. that I am an instance of a certain *psychological type*, or c) that I belong to the Mussolini family, a certain *anthropological type*. If a)-b) depend on the relevant local pragmatic process, it is hard to see why c) does not so depend.

are processes that affect just *one* subsentential element – whether it is a singular term or a predicate – in order to provide another interpretation different from its standard one. Yet the process that should obtain here is a process from a *certain* subsentential element (the name in argumental position) to *another* subsentential element (the name in predicative position). Thus, why should one postulate the existence of such a *sui generis* pragmatic process when a simpler interpretation of the data in question is available, as the one that a Predicate Theory provides?

If the above is the case, then the fact that a Predicate Theory manages to explain how the predicative and the referential uses stem out of a common semantical root may lead one to think that dispensing with such a theory is like throwing the baby with the bathwater. We have hitherto seen that, once it is strengthened in the Nominal Theory, a Predicate Theory is incompatible not only with Millianism, but also with the weaker Direct Reference thesis. Now, it is worth exploring whether there is no weaker version of a Predicate Theory that makes it compatible, if not with Millianism, at least with the weaker Direct Reference thesis. In other terms, I will look for a *syncretistic* account that combines the best of a Predicate Theory and the best of Direct Reference while dropping their radicalizations making them theoretical antagonists.

3. The compatibility between Predicate Theory and Direct Reference

As I said before, Millianism radicalizes the Direct Reference thesis, the thesis that the meaning of a proper name ends by collapsing onto its referent, in a particular way; namely, by saying not only that i) the *main* semantically relevant contribution of a proper name is its truth-conditional contribution, i.e., its referent (if any), but also that ii) such a contribution is the *only* semantically relevant contribution.

Yet this way of articulating the above thesis is not mandatory at all for being a direct referentialist. For there are other expressions, or at least certain uses of them, which are taken to be directly referential and yet fail to satisfy thesis ii) of Millianism. Indeed, they possess a further dimension of meaning that enables them to have certain referents as their

truth-conditional contributions. As a result, for such expressions their meaning ends by collapsing on their referents, yet it is not absorbed by them.

These expressions are indexicals and (possibly) proper definite descriptions when referentially used. Indexical expressions – pure indexicals like "I", "here", and "now", demonstratives like "this" and "that", etc. – are directly referential expressions insofar as their truth-conditional contribution is provided by their contextual referents. Yet they do not function as Millian devices of reference, for they also possess another dimension of meaning, their so-called *nonconstant character*, which is the function that enables one such expression to get a referent given a certain context of truth-conditional interpretation – typically, a different referent for any such different context of interpretation.[16] The referential use of a proper definite description may be seen as the use in which that description is again a device of direct reference for its referent provides its truth-conditional contribution. Yet in such a use the description again is no Millian device of reference, for it gets such a referent insofar as it uniquely satisfies the *constant character* that description may be seen to possess, namely the constant function that, given *any* context of interpretation, yields to such a description the very same referent. According to this way of putting things, the definite description figuring in the famous Donnellan (1966) example "the man drinking a Martini" has a referential use that yields no individual as its truth-conditional contribution, for neither the individual the speaker has in mind when so using the description, nor anybody else in the relevant situation for that matter,[17] drinks a Martini. Hence, the description is not proper, for nobody satisfies its constant character. Yet the description "the Italian Prime Minister in 2015" has a referential use whose truth-conditional contribu-

---

[16] I follow Predelli (2005) in distinguishing context of interpretation from context of utterance. For, as he has convincingly shown, there are contexts of utterance that are irrelevant for yielding a sentence its contextual truth-conditions.

[17] I put the Donnellan's example in this way in order to rule out the possibility that the definite description in question is still proper, for in the situation at issue there is someone else who uniquely drinks a Martini, so that this someone may work as the referent of the referentially used description.

tion is Matteo Renzi, for in its being used to refer to Matteo, he uniquely satisfies its constant character, so that the description is proper.[18]

Once things are put this way, there is a way to make a Predicate Theory compatible with the Direct Reference thesis. For one may go on saying that, as far as proper names are concerned, in their referential use such expressions have their referents, if any, as their truth-conditional contributions. Yet while so used, proper names also express a character which (contextually) determines such contributions. Moreover, such a character amounts to a property that either coincides with or entails[19] the property that is predicated of them when they are predicatively used.

If this is the case, one manages not only to give a semantically uniform account of the referential and the predicative use of a proper name, but also to explain why the property that is expressed in the referential use as the name's character coincides with or entails a property of the kind *being called "N"*. For the latter is the property that is predicated in the predicative use of a proper name. To see why this is the case, consider to begin with indexical expressions. Also indexical expressions have predicative uses.[20] Consider sentences like

(8)   That cat is a he;
(9)   The today I'll someday want back remains vivid in my mind;
(10)   Once one has journeyed half of her life's way, there are for her more
       yesterdays than tomorrows.

It would be odd to say that the predicative use of the indexical "he" in (8), that of "today" in (9), and those of the indexicals "yesterday" and "tomorrow" in (10), have nothing to do with the referential use of such indexicals. For what is predicated in the former use are respectively the properties of (roughly) *being a male individual in the surroundings of the contextual speaker and interlocutors, being the day of the context, being the day that precedes the day of the context, being the day that follows the day of the context.* Now, such properties are precisely the properties that

---

[18] For this way of drawing a semantically relevant distinction between the referential and the attributive use of a definite description, cf. e.g. Recanati 1993.

[19] By considering later the problem of homonymity, we will see why this specification is important.

[20] In my Voltolini 1995, I already drew this comparison between indexicals and proper names in favour of the indexicalist theory of proper names I will appeal to later.

constitute the respective characters of such indexicals, as expressed by their respective referential use. Indeed, given a context of interpretation, the referent of "he" will be a male in the surroundings of the speaker and her interlocutors in that context; *mutatis mutandis*, one can say the same as to "today", "yesterday", and "tomorrow".

When we come to proper names, the situation is pretty similar. The property of *being called "Alberto"*, which is predicated by "an Alberto" in (1) once it is shorthand in such a predicative use for "an individual called 'Alberto'", either coincides with or entails the property that constitutes the character of "Alberto" as expressed in its referential use. Indeed, given a context of interpretation, what the name refers to is the individual that is called "Alberto" in that context; myself, in such a case. Likewise for "Voltolini" and "Alberto Voltolini" in (2) and (3) respectively.

Thus, once a Predicate Theory is suitably weakened so as to disentangle it from the Nominal Theory, it is quite compatible with the Direct Reference thesis once this latter thesis is also suitably modulated, namely, it is taken to coincide just with thesis i) and not also with thesis ii), as Millians instead pretend. Insofar as this compatibility not only allows for a basically uniform account of the predicative and of the referential use of proper names but also explains why a property of the kind *being called "N"* is the property both uses mobilize (as respectively predicated and expressed), defending such compatibility clearly constitutes a theoretical advantage. Thus, a syncretism about proper names that articulates such compatibility naturally suggests itself. Yet *which* form of syncretism exactly?

4. The best syncretistic account of proper names

As far as I can see, there are at least three, if not four, syncretistic ways of combining a Predicate Theory and the Direct Reference thesis. According to the *first* way, a proper name is actually like a referentially used proper definite description. For the truth-conditional contribution of a proper name in its referential use is just its referent, yet such a referent is determined by its *constant* character. Given any context of

interpretation, the referent of the proper name always remains one and the same individual, the one that uniquely satisfies the nominal property constituting such a character.[21] According to the *second* way, demonstratives are the paradigm of proper names. For again, the truth-conditional contribution of a proper name in its referential use is just its referent, yet such a referent co-varies with the context of interpretation, since the *nonconstant* character of a proper name determines a different referent for any different context of interpretation. Depending on whether the referential use of a proper name is taken or not to be strictly tantamount to that of a complex demonstrative of the kind "that N", which is in its turn shorthand for "that individual called 'N'", we have two options of how to articulate this second syncretistic way.[22]

These two ways basically differ insofar as they differently account for the phenomenon of *homonymous* proper names, i.e., the fact that the same morpho-syntactic string of letters that constitutes a proper name may be used to refer to different individuals. For instance, the surname "Voltolini" may be used to refer to me but also to Dario, my writer friend, who is the Voltolini I'm curious of.

According to the first way, homonymous names are just semantically different expressions insofar as they possess *different constant* characters. In this perspective, a constant character expresses a *relativized* nominal property, namely a property of the kind *being called "N" according to a certain naming convention*, where this naming convention is basically individuated in terms of a certain dubbing context. Since in the case of homonymy we have different dubbing contexts, hence different naming conventions, we have different constant characters, each for any homonymous proper name. Such characters respectively determine, in any context of interpretation, different referents for any such name. So in the "Voltolini" case, we have (at least) two different "Voltolini" names, one whose character includes a *certain* naming convention (basically, a proce-

---

[21] We owe this way to Predelli (2015). According to him, this way makes not only a Predicate Theory, but also the Nominal Theory, compatible even with Millianism. Yet this depends on the fact that he both means the Nominal Theory not as a form of Russellian descriptivism and provides a non-standard characterization of Millianism (see fn. 1).

[22] One can trace back those two options to Castañeda (1990) and Burge (1973), respectively.

dure tracing back to the dubbing my parents did that determines myself as its referent in any context of interpretation) and another whose character includes *another* naming convention (basically, a procedure tracing back to the dubbing Dario's parents' did that determines Dario as its referent in any context of interpretation). Cf. Predelli 2015: 371–375.

According to the second way instead, properly speaking there are no homonymous proper names, there is simply just one name that, in virtue of its nonconstant character, in different contexts of interpretation has different referents. Difference in reference for a proper name is basically a matter of indexicality. Just as it may be the case that in different contexts of interpretation the demonstrative "that" has different referents in virtue of its nonconstant character, roughly *being in the surroundings of the contextual speaker and interlocutors*, so it may be the case that the surname "Voltolini" refers to myself in a context of interpretation, to Dario in another such context, in virtue of its nonconstant character, roughly *being called "Voltolini"*.

The *third* syncretistic way agrees with the second one in its being an indexical account, basically accounting for homonymy in the same way as the second does.

In this respect, both the second and the third way have a slight theoretical advantage on the first way. For while they can account for both the referential use and the predicative use of a proper name by saying that *the very same property*, i.e., a property of the kind *being called "N"* or even *being contextually called "N"*, is mobilized as predicated of something in the latter use and as expressed as a nonconstant character in the former use, the first way is forced to say that while this very property is mobilized in the latter use, the former use expresses different more fine-grained properties, namely different convention-relative properties as the respective constant characters of the proper names involved. As the following sentence clearly shows:

(11) There are a few Voltolinis in the Turin's telephone directory.

(11) clearly states that the number of the individuals called "Voltolini" mentioned in the Turin telephone directory is small. Yet the constant character that is associated to any "Voltolini" name is always a different one, each time more fine-grained than the property of *being called "Voltolini"*. It is the convention-relative property of *being called*

"Voltolini" according to a **certain** *convention* in the case of *my* surname, the different convention-relative property of *being called "Voltolini" according to* **another** *convention* in the case of Dario's surname, and so on. Cf. Predelli 2015: 374–375.

Yet the third way disagrees with the second one in that for it, proper names are a *sui generis* form of indexicals, *indexinames* to give them a label. Indexinames are like pure indexicals on the one hand, since their reference is *automatically* determined by a context of interpretation given their nonconstant character. Yet they are also like demonstratives on the other hand, since in certain contexts of interpretation they are *referentless.* Let me clarify.[23]

As to the first point, the context of interpretation of an indexiname is a *narrow* context, i.e., a set-theoretical construction made of certain parameters, hence it is not a context in a *wide* sense, the concrete overall situation of discourse.[24] Moreover, it is an *enlarged* narrow context, for instead of limiting itself to containing the parameters that traditionally constitute such a context, i.e., agent, space, time and world of interpretation, it contains a *further* parameter, the parameter of naming conventions or better of dubbings. In a nutshell, what according to the first syncretistic way is put *outside* the context of interpretation, as constituting a prior and independent context – a dubbing context – according to the third way is put *within* the context of interpretation, as constituting another of its parameters. Furthermore, since the nonconstant character of an indexiname is again constituted by a property of the kind *being contextually called "N"*, the referent of an indexiname is automatically, i.e., non intentionally, fixed given a certain context of interpretation. For such a character directly points to the dubbing parameter of that context so as to provide the referent of the indexiname in that context, one's intentions on this concern notwithstanding. As it happens with pure indexicals: e.g. the character of "I", roughly *being the agent in context*, directly points to the agent parameter in narrow context so as to automatically provide the referent of "I" in that context.[25]

[23] I have appealed to this form of syncretism in my Voltolini 1995, 2014. It may be also found in Recanati 1993, Pelczar and Rainsbury 1998.

[24] For the distinction between these notions of context cf. e.g. Perry 1997.

[25] For such an interpretation of the automaticity of narrow context, cf. Recanati 2004a. To be sure, language users may well ignore which property a character of an indexi-

As to the second point, the nonconstant character of an indexiname is also a *partial* function, in that given a certain context of interpretation, there may be no referent in that context, that is, when the dubbing fails in want of an individual to be dubbed. As it happens with demonstratives, which fail to refer to anything when there is no relevant individual in the context of interpretation (cf. Kaplan 1989, 1989a).

The third syncretistic way is definitely better than the second way. To begin with, equating the character of a proper name to the character of a demonstrative, as the second way does, leaves the contextual reference of the name indeterminate even once a certain narrow context is fixed until further factors coming from the wide context are appealed to. The context of interpretation, the one providing the relevant contextual truth-conditions for the sentence involved, must be in that case the wide, not the narrow context, as so-called *contextualists* as to the semantics/pragmatics divide maintain (cf. e.g. Recanati 2004). If one utters

(12) That comes from North-Western Italy

the reference of "that" in the above token of (12) is indeterminate even if one fixes the agent, the space, the time, and the world of interpretation, in a nutshell even if one fixes a certain narrow context, until the utterer specifies e.g. that among all individuals that contextually satisfy the character of "that", roughly *being in the surroundings of the contextual speaker and interlocutors*, she had *one* of those individuals in mind (as Kaplan (1989, 1989a) originally envisaged), where the utterer's intentions are a wide-contextual factor. In the second syncretistic way, the situation does not basically change if one utters

(13)   As to his family's origins, Voltolini comes from North-Eastern Italy.

For according to such a way, the character of "Voltolini" notwithstanding (roughly, *being called "Voltolini"*), given a certain narrow context it is still indeterminate whether "Voltolini" in the above token of (13)

---

name amounts to, in particular they may well ignore that such a property appeals to certain psychological facts of attention (see later). Yet *pace* Leckie (2013: 1156) this is not a problem. For not even users of a pure indexical such as "I" must know which property the contextual referents of "I" share in order for them to be such referents; nevertheless, those referents are automatically fixed as well, given such a character.

refers to me, to my friend Dario, or to any other individual contextually satisfying that character, until some further wide-contextual factors are appealed to (for instance, the fact that the utterer of (13) has Dario rather than myself or any other Voltolini in mind).

Granted, on behalf of the second form of syncretism one might try to pair a demonstrative, hence a proper name as well, with a pure indexical, namely by supplementing the expression's character in such a way that it may point to a *further* element in *narrow* context, so that its contextual reference is automatically fixed. As to demonstratives, so-called *traditionalists* as to the semantics/pragmatics divide, namely people who believe that at most narrow context is the relevant context of truth-conditional interpretation, often appeal to such a move. For instance as to "that", the idea is that its character roughly amounts to *being the* **demonstrated** *individual in the surroundings of the contextual speaker and interlocutors*. As a result, by pointing to a further parameter in narrow context, a parameter of *demonstrations* (perceptual ostensions) or of *demonstrata* (ostended individuals), the character of "that" automatically provides a referent for that demonstrative in such an enlarged narrow context. For such strategies, cf. e.g. Caplan 2003.

Yet independently of whether this move works for demonstratives, as I strongly doubt (cf. Voltolini 2009), it is unlikely that it works for proper names. Suppose that one so supplements the character of a proper name, by making it roughly identical with a property of the kind *being the demonstrated individual contextually called "N"*. Thus, the name sounds equivalent to a complex demonstrative of the kind "that N", i.e., "that individual called 'N'". One then adds a parameter of demonstrations or of *demonstrata* in the enlarged narrow context of a proper name. Yet this move does not suffice to settle contextual reference for a proper name. For even if one fixes that further parameter, that reference is not settled yet. Suppose that two stars both named "Starry" send to the earth light rays that ultimately collapse into just one ray, thereby appearing as just one dot in the sky's vault. By pointing to that dot, one utters

(14) Starry is shining.

Yet by fixing her perceptual ostensions or such an ostended dot, it is not settled yet to which star the utterer of (14) is referring by "Starry". Clearly enough, no such problem arises as to the third way. For if a

proper name is an indexiname, then its character points to the parameter of dubbings in a narrow context enlarged by adding *that* parameter, so as to automatically get a referent in such a context. If "Starry" in (14)'s token mobilizes a *certain* dubbing, it refers to *one* of the two stars; if it mobilizes *another* dubbing, it refers to the *other* one.

To be sure, on behalf of the second syncretistic way a contextualist might reply that if the above problem merely shows that the context of truth-conditional interpretation for proper names treated as demonstratives must be the wide context, the worse for a traditionalist. This would not be an exciting reply, since contextualism as to meaning phenomena forcefully suggests itself only when no traditionalist account of the phenomenon at issue sounds plausible.[26] Yet the above counterexample shows that the problem is even deeper than that. A proper name cannot work exactly like a demonstrative, for its range of application widely exceeds what can be perceptually given. Unlike demonstratives, we use proper names to refer not only to what is out there, but also to what is not out there, for it is very far away or it is not there at all (it is a nonpresent entity, an abstract entity, a fictional entity, a merely possible entity . . .).

Moreover, the third syncretistic way is also better than the first way. As we have seen, in the third way the contextual referent (if any) of the function that constitutes the proper name's character is automatically given, for the character points to a parameter in the argument of that function, a certain enlarged narrow context, that nonintentionally provides the value of the function, the name's contextual referent (if any). Yet in the first way there is no element in the relevant function that is mobilized there, a function from dubbings to constant characters, which points to the function's argument, a given dubbing, so as to automatically provide the function's value, a certain constant character. For in such a case, what is at our disposal is only the *shape* of a certain expression. From that shape, one cannot get to a dubbing so as to automatically retrieve a constant character. Consider an example provided by Napoli (2015: 215):

(15)   Leave me alone is in a bad mood.

---

[26] As Recanati (2004: 116) convincingly says.

In the relevant token of (15), the expression "Leave me alone" is used as a nickname of a not very sociable fellow. Yet by merely considering the shape of that expression, we cannot get its syntactic form, which may be that of a sentence, as in its usual reading that would provide another yet ungrammatical token of (15), as well as that of a proper name, as in its unusual reading that interests us. If on the basis of that expression's shape we could have that syntactic form at our disposal, we might single out a syntactic element in that expression's shape that points to a dubbing, so as to automatically get the constant character of "Leave me alone" we are looking for. But since we do not so have such a form at our disposal, we cannot single out such an element. This may have a dramatic consequence, as in the famous Polyphemus story. If his fellow Cyclops could have singled out the relevant syntactic form on the basis of the uttered shape, they might have immediately helped Polyphemus when he uttered

(16)   Nobody has blinded me.

However, since they did not so have such a form at their disposal but they merely had to rely on an intentional factor that they (luckily) ignored, i.e., Polyphemus' intention to refer by "Nobody" to Ulysses, they limited themselves to mocking at their unfortunate fellow, by privileging the ordinary, quantified interpretation of (16).

On behalf of the first syncretistic way, one might reply with Predelli (2015: 374) that this is as should be. Semantic interpretation starts only when a constant character, hence a certain proper name, is at our disposal, so as to allow one to provide a certain truth-conditional interpretation for that name. Whatever happens before, dubbing procedures included, just belongs to the domain of pragmatics, as involving a *pre-semantic* role of context.

Yet it is not clear to me why this (let me call it) *second-order* contextualism, a contextualism as to (constant) character rather than as to (truth-conditional) content, should be the right game in town, if an overall traditionalist account of proper names such as the one the third syncretistic way provides is available. As a matter of fact, there are cases in which we may avoid such a second-order contextualism. For instance, we may relativize shapes to languages, so as to automatically get different lexical representations, including constant characters of the terms

involved, in virtue of the different syntactic structures such languages respectively mobilize. E.g., once we relativize a certain string of sounds to English, English makes us point to a certain syntactic structure that automatically gives us a certain lexical representation, including *certain* constant characters of the terms involved:

(17) Humpty Dumpty sat on a wall.

Yet once we relativize the same string of sounds to French, French makes us point to another syntactic structure that automatically gives us another lexical representation, including *other* constant characters of the terms involved:

(18) Un petit d'un petit s'étonné aux Halles.[27]

In such a case, we may give up a second-order contextualist account of lexical representations, for we have a second-order traditionalist account of such representations. Yet in point of fact, we can hardly so relativize shapes to syntax, as the following example Napoli (2015: 216) provides clearly shows. Only wide context can tell us which is the relevant reading (vocative hence nominal, imperative hence non-nominal) of "Leave me alone" in

(19) Leave me alone, do not try to fool me.

If we cannot analogously relativize shapes to syntax as far as (15), (16) and (19) are concerned and so ascend to a second-order traditionalist account of proper names, then it is better to stick to the third rather than to the first syncretistic way. For second-order contextualism may be in order when what has to be both wide-contextually and pre-semantically settled is just the expression to which a semantical interpretation must be subsequently applied, as traditionalists ordinarily maintain.[28] Yet it

---

[27] By pointing to this second-order traditionalist account of the case at issue, I do not want to say that it is its best account. I prefer a contextualist account that interprets the sound that utterances of (17)–(18) roughly share as an intentionally ambiguous punny utterance of one and the same sentence, on a par e.g. with "Condoms should be used in any conceivable occasion". For this account cf. Voltolini 2012.

[28] This is how Perry (1997) and Predelli himself (2005) deal with the phenomenon of lexical ambiguity; when a lexically ambiguous expression is at stake, actually wide context in a pre-semantic role makes one select *which* homonym among those that disambiguate the expression is at stake in order for semantic interpretation to start.

is less in order when what has to be both wide-contextually and pre-semantically settled is rather a *meaning* dimension, namely the alleged constant character of proper names, as supporters of the first syncretistic way hold. For as we have seen before, contextualism as to meaning phenomena forcefully suggests itself only when no traditionalist account of the phenomenon at issue sounds plausible. Yet the third syncretistic way precisely provides such an account.

5.   A speculative story

Up to now, first, I've limited myself to saying that a syncretistic theory of proper names that combines the virtues of the Predicate Theory and those of the Direct Reference Thesis is preferable to both the Nominal Theory and to Millianism, which for some reason or other unsatisfyingly radicalize the above doctrines. For it accounts for the very idea from which a Predicate Theory started and the Direct Reference thesis by itself does not deny, namely the idea that the predicative and the referential use of a proper name share a common semantical root. Second, I've tried to show that among the three (if not four) syncretistic approaches I have considered, the third one is the best. For it accounts in an utterly traditionalist way for the homonymity of proper names, by taking them as a *sui generis* form of indexicals, namely indexinames, whose character is a partial function from an enlarged yet narrow context to referents.[29] Yet is there a non-technical, but a philosophically substantive, reason, as to why we should stick to the syncretistic approach based on indexinames?

I guess that there is one. As I hinted at before, I take that the property a proper name's character mobilizes is a property of the kind *being contextually called "N"*, where "called 'N'" has to be meant as *called to the interlocutors' attention by means of 'N'*. To begin with, this move has the further technical advantage that it purportedly saves the analyticity, i.e., truth in all contexts of interpretation, of a sentence of the form

(20)   If it exists, N is called "N".[30]

---

[29] In my Voltolini 2014 I presented another argument in favour of the indexinames theory. Such an argument revolves on how to account for empty uses of proper names, especially the fiction-involving ones. Yet probably such an argument has counterparts in each of the other syncretistic approaches.

[30] As originally envisaged by Burge (1977: 344, fn.7).

Being analytic in the sense of being true in any context of interpretation is what is expected by a sentence that predicates of an individual the property that allegedly qualifies the relevant indexical's character (as for instance happens with "I am the context's agent", which is true in any context of interpretation for "I"). Now if (20) meant that if it exists, N bears "N", then there might surely be a context in which it is false, namely a context in which the name "N" is a scheme for is not the official name of a certain individual: recall the "Leave me alone" case given before (cf. Napoli 2015: 218). No such problem arises if "called 'N'" means *called to the interlocutors' attention by means of 'N'*. For instance, in the "Leave it alone" case, although the relevant individual is not the official bearer of that nickname, she is still the individual the speaker is calling her interlocutors' attention to by uttering "Leave me alone".[31]

Yet more importantly for my present purposes, the above move may suggest a theoretical justification as to why names are indexinames. Consider whistles, claps, or any other way to attract someone's attention. One may well take them as context-sensitive forms of expression, just as indexicals are. In one context, a certain whistle attracts my attention; in another context, a qualitatively identical whistle attracts the attention of someone else. Yet a whistle, or any functionally similar sound, is not a proper name yet. In order for such a sound to become a proper name, it has to be so to speak *stabilized*.[32] Unlike the previous case, when a sound

[31] For more about this, cf. Voltolini 2014: 303. Fara (2011) distinguishes between *being called N* and *being called "N"* roughly in the same way as I did between *being a bearer of "N"* and *being called to one's attention by means of "N"*. Yet she also says that her two notions are independent of each other. Although I agree that *being called to someone's attention by means of "N"* does not entail *being a bearer of "N"*, I think that the converse entailment holds. For, as I will try to show below, it is only in virtue of being called to someone's attention by means of a name that one also becomes a bearer of that name. Fara's (2011: 496–497) alleged counterexample – she is called Delia Ruby Graff Fara but she never heard anyone so calling her to someone's attention – is not a real counterexample, for it presupposes that the only way to attract someone's attention is the *auditory* one. Yet this of course is not the case. As she admits, at a certain point she added "Fara" to "Delia Ruby Graff". But this addition could take place only if she *manifested* it somehow – for instance, by writing it. For this is enough as a means to attract someone's attention to her.

[32] I here focus on sounds, on the reasonable genetic hypothesis that oral language precedes other forms of language. Yet, as I have shown in the previous footnote, nothing would basically change if I focused say on written language, which mobilizes visual rather than auditory perception.

is stabilized as a proper name, the subject of attention – the speaker's interlocutor – may well not coincide with the object of attention (if any)[33] – the speaker's addressee.[34] For not only the addressee may even be an inanimate thing or anyway an entity unable to display attention, but also the interlocutor manages to allow for an attentional focus to the addressee (if any) to persist not only when the addressee is no longer there, but also when the interlocutor herself is no longer there and has been replaced by another person, i.e., a later interlocutor. There are indeed two ways of attracting the earlier vs. the later interlocutor's attention to something: an *in praesentia* and an *in absentia* one. They enable the original sound to be used with respect to the addressee (if any) out of the original dubbing circumstance as a name of the addressee itself. For by means of such ways, not only the circumstance in which that sound is originally uttered may be reinforced in its characteristic of being a dubbing (of the addressee), but also that very dubbing may be appealed to as the relevant contextual parameter in order for the later interlocutor to assess *what* (if any) her attention has to be directed upon. Since the later interlocutor has not attended the original dubbing circumstance, she has indeed to settle to *what* her attention has to be directed upon when she hears a proper name. She can do that only by selecting a certain dubbing circumstance rather than other ones. (Such a selection may, but not must, be performed by appealing to certain referential intentions, as the traditional causal-intentional theory of naming stemmed out of Kripke (1980) claims.) Hence, the referential value of the name changes, depending on which dubbing is pointed out. As the indexiname approach predicts.[35,36]

---

[33] Remember that (uses of) names may well be referentially empty if the relevant dubbing fails when there is no individual one's attention is attracted to.

[34] The distinction between interlocutors, addressees, and speakers is functional. Sometimes, either the first two roles or the second two roles are instantiated by the same individual.

[35] In so giving an account of the mechanism of name reference, the third kind of syncretism copes with the methodological worries Jeshion (2015: 236–237 and fn. 15) raises against a Predicate Theory of names.

[36] I warmly thank Stefano Predelli for some comments to a previous version of this paper.

References

Bach, K. (1981), "What's in a name", *Australasian Journal of Philosophy*, 59, pp. 371–386.

Burge, T. (1973), "Reference and proper names", *Journal of Philosophy*, 70, pp. 425–439.

——— (1977), "Belief *de re*", *Journal of Philosophy*, 74, pp. 338–362.

Caplan, B. (2003), "Putting things in contexts", *Philosophical Review*, 112, pp. 191–214.

Castañeda, H.N. (1990), "The semantics and causal roles of proper names in our thinking of particulars: The restricted-variable/retrieval view of proper names", in K. Jacobi and H. Pape (eds.), *Thinking and the Structure of the World*, De Gruyter, Berlin, pp. 11–56.

Donnellan, K. (1966), "Reference and definite descriptions", *Philosophical Review*, 75, pp. 281–304.

Fara, D.G. (2011), "You can call me 'stupid',... just don't call me stupid", *Analysis*, 71, pp. 492–501.

——— (2015), "'Literal' uses of proper names", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 251–277.

Fauconnier, G. (1985), *Mental Spaces*, Cambridge University Press, Cambridge.

Grice, H.P. (1975), "Logic and conversation", in P. Cole and J. Morgan (eds.), *Syntax and Semantics. Speech Acts*, volume 3, Academic Press, New York, pp. 41–58, reprinted in his *Studies in the Way of Words*, Harvard University Press, Cambridge, MA 1987, pp. 22–40.

Jeshion, R. (2015), "Names not predicates", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 225–249.

——— (2015a), "A rejoinder to Fara's '"Literal" uses of proper names'", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 280–294.

Kaplan, D. (1989), "Demonstratives", in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford University Press, Oxford, pp. 481–563.

——— (1989a), "Afterthoughts", in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford University Press, Oxford, pp. 565–614.

Kneale, W. (1962), "Modality *de dicto* and *de re*", in E. Nagel, P. Suppes and A. Tarski (eds.), *Logic, Methodology and Philosophy of Science*, Stanford University Press, Stanford, pp. 622–633.

Kripke, S. (1980), *Naming and Necessity*, Harvard University Press, Cambridge, MA.

Leckie, G. (2013), "The double life of names", *Philosophical Studies*, 165, pp. 1139–1160.

Marti, G. (1995), "The essence of genuine reference", *Journal of Philosophical Logic*, 24, pp. 275–289.

Napoli, E. (1995), "(Direct) reference", *Journal of Philosophical Logic*, 24, pp. 321–339.

——— (2015), "Names as predicates?", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 211–224.

Pelczar, M. and Rainsbury, J. (1998), "The indexical character of names", *Synthese*, 114, pp. 293–317.

Perry, J. (1997), "Indexicals and demonstratives", in B. Hale and C. Wright (eds.), *A Companion to the Philosophy of Language*, Blackwell, pp. 586–612.

Predelli, S. (2005), *Contexts: Meaning, Truth, and the Use of Language*, Oxford University Press, Oxford.

——— (2015), "Who's afraid of a predicate theory of names?", *Linguistic and Philosophy*, 38, pp. 363–376.

Recanati, F. (1993), *Direct Reference*, Blackwell, Oxford.

——— (2004), *Literal Meaning*, Cambridge University Press, Cambridge.

——— (2004a), "Indexicality and context-shift", paper presented at the *Workshop on Indexicals, Speech Acts and Logophors*, Harvard University, 20.11.2004. `http://jeannicod.ccsd.cnrs.fr/docs/00/05/36/04/PDF/ijn_00000552_00.pdf`.

Voltolini, A. (1995), "Indexinames", in J. Hill and P. Kot'àtko (eds.), *Karlovy Vary Studies in Reference and Meaning*, Filosofia Publications, Prague, pp. 258–285.

——— (2004), "Can there be a uniform application of direct reference?", *Erkenntnis*, 61, pp. 75–98.

——— (2009), "How demonstrative pictorial reference grounds contextualism", *Pacific Philosophical Quarterly*, 90, pp. 402–418.

——— (2012), "Puns for contextualists", *Humana Mente*, 23, pp. 113–140.

——— (2014), "Fiction and indexinames", *Journal of Literary Theory*, 8, pp. 292–321.

A criticism of the syntactic solution to Frege's problem

Elisabetta Sacchi

1. Introduction

In this paper I want to deal with the issue of the cognitive value of referential devices in order to assess which account of this issue turns out to be more adequate to address what is called Frege's problem. My main critical focus will be on the non-semantic accounts of the cognitive value that most people adhering to the direct reference picture have put forward. Also Ernesto Napoli (1997) endorses this approach and in a passage that I deem very representative of this anti-Fregean strategy he claims:

> Frege has thought that to solve his puzzle, modes of presentation, i.e., senses, should be brought in. After him generations of philosophers have worked hard to make senses affable and effective. Senses have become all the more cumbersome and fat. To no appreciable effect. In the end the only successful move is to make the very expression itself part of the sense [. . .]. Semantic values have indeed modes of presentation. Disappointingly the mode of presentation of a semantic value is no other than the expression of which it is the semantic value. (196)

In the same vein John Perry (2015), in the attempt to defend referentialism against the charge that it does not provide an account of the cognitive contribution of expressions adequate to solve Frege's problem,[1]

---

[1] What Perry means by "referentialism" is the view that what an expression contributes to the proposition expressed by statements in which the expression is used is just its referent.

claims that the cognitive contribution of a name is how it sounds (if the name is spoken) and how it looks (if the name is written) and that, therefore, "the cognitive contribution of names neither requires nor motivates something in the semantics that associates names with ideas, or Fregean senses, or anything else except what they stand for" (200).

In critically addressing the syntactic account put forward by Frege's critics, I do not intend to defend either Frege's semantic account or any other semantic account, such as for example Fine's unFregean referentialist account.[2] I shall take the criticisms of the direct referentialists against the semantic approach for granted and consider whether the syntactic account they put forward is the best non-semantic account available. The gist of my criticism will be that syntactic entities are ultimately unable to properly address Frege's problem and, since to be able to address it is a *conditio sine qua non* in order for something to play the mode of presentation role, I shall conclude, against the advocates of the syntactic account, that modes of presentations cannot be expressions. My reason for claiming that the syntactic accounts do not answer Frege's problem is that they leave unaddressed an issue which I deem at the heart of that problem, namely the question of *what grounds the fact that a given subject takes/does not take what she is (mentally or linguistically) referring to on a given occasion to be the same as what she is referring to on another occasion*. If this question is not addressed, no non-circular and non-vacuous solution to Frege's problem can be provided. Or so I shall claim.

After having presented my criticism to the syntactic approach I shall put forward my positive proposal. As it will turn out, my suggested account will be neither a semantic nor a syntactic one. I shall qualify it as phenomenological because it characterizes *modes of presentation as ways of experiencing* the objects referred to. Even though my account is

---

[2] According to Fine (2007), Frege and Fregeans are right in claiming that cognitive differences must be reflected by semantic differences. What they are wrong, in his view, is in claiming that the semantic difference in question must be an intrinsic one. What Fine means by "intrinsicalism" is the idea according to which whether two utterances say the same thing is fully determined by features intrinsic to the individual utterances. Fine's alternative to intrinsicalism is relationism, the claim according to which "the fact that two utterances say the same thing is not entirely a matter of their intrinsic semantic features; it may also turn on semantic relationships among the utterances or their parts which are not reducible to those features" (2007: 3).

unFregean in so far as it does not treat modes of presentation as belonging to the semantic content, it aims at addressing what in my view is the main point behind Frege's problem, namely the issue of what grounds the aspectuality of the referential/intentional relation. This is the issue that in my view is left unaddressed in the non-semantic accounts that have been put forward within the direct reference framework. Actually, Ernesto seems to recognize that the syntactic account cannot be the definitive answer to Frege's problem and that, behind linguistic modes of presentation, one has to acknowledge other modes of presentation which are more basic. As he says, "the distinctness of proper names is the linguistic trace of an extra-linguistic distinction. Two coreferential proper names are distinct linguistic modes of presentation which reflect and are grounded in distinct extralinguistic modes of presentation" (1997: 198). What I shall consider is how the extra-linguistic modes of presentation that are responsible for and ground cognitive differences have to be conceived. Even though my answer to this question diverges from the one that Ernesto endorses in his paper,[3] I think that it is one that a referentialist could espouse, because as a matter of fact it does not ask one to commit to any cumbersome semantic framework.

2. What Frege's problem is really about

In this section I want firstly to summarize Frege's approach to the puzzles raised by co-referential expressions. That will help me to highlight what I take to be the main point behind Frege's problem. I think that a proper appreciation of the extent and scope of this problem is of the ultimate importance here and to my view this appreciation is almost missing in most works that have addressed this topic in the more or less recent literature. After having illustrated Frege's strategy I shall briefly present the main considerations that have led people to reject it. My main claim in this part will be the following: even though it can be granted that

---

[3] What Ernesto is thinking at are distinct uses or distinct baptisms, that is events of name introduction performed in distinct situations, i.e., as he says, "under presentations of the object which, no matter how similar, are not different *solo numero*" (1997: 199).

Frege's strategy is neither necessary nor sufficient to deal with the issue of the cognitive value, still Frege was right in maintaining that what is needed is something that is suited to play the 'mode of presentation role'. This is a point that even those who reject Frege's account of the cognitive value in terms of senses are willing to accept. The question we shall then consider in the next section, after having illustrated the main candidates put forward by Frege's opponents to deal with Frege's problem, is whether those candidates qualify as the kind of entities suited to play that role. In addressing this question I shall make use of what I take to be a necessary requirement that something must satisfy in order to qualify as a legitimate candidate for the mode of presentation role. Helping myself with this requirement I shall claim that the candidates suggested by the advocates of the syntactic account do not pass muster.

Let me now briefly consider what the issue of the cognitive value concerns and how Frege addresses it in *Über Sinn und Bedeutung* (1892). In this work Frege introduces the issue as one concerning how to account for the difference between statements of the form "$a = b$" as opposed to statements of the form "$a = a$". While the former can be informative (in the sense of being able to impart new knowledge) and a posteriori, the latter do not impart knowledge and are always a priori. In addressing this question Frege rejects the metalinguistic account he had provided in his *Begriffschrift* according to which the expressions flanking the identity sign stand for themselves, and presents his new account based on the distinction between sense and reference. According to his new account, the names flanking the identity sign stand for an object, which is the same if the identity statement is true. But, and here comes the novelty of his mature account, referring to something is only one of the semantic functions that linguistic expressions play, the other being that of expressing a sense that indicates (or as Frege sometimes says "contains") the mode of presentation of the referent. With this distinction in place Frege can answer the question raised in the opening paragraph of his article by claiming that the difference between the two identity statements is to be accounted for by appealing to the notion of sense. Even though the singular terms "a" and "b" which figure in the true identity statement refer to one and the same object (they have the same *Bedeutung*), they differ at the level of sense (they are associated with different modes of presentations of the referent) and it is precisely this difference

that accounts for the potential informativeness of those statements as compared to those which are mere instances of the logical principle of identity. With this move, occasionally motivated by the need to provide a more satisfactory account of identity statements, Frege put forward his two levels semantic approach which was to become the cornerstone of all the subsequent reflections in the philosophy of language.

The application range of Frege's proposal covers not only identity statements (those which occasioned the introduction of the distinction), but any other statement as well: what his proposal aims at explaining is not just how true identity statements of the form "$a = b$", as against those of the form "$a = a$", can be informative and not always a priori true, but more generally how two any statements which differ only for the fact of containing distinct coreferential expressions can differ in their cognitive value, that is (to put it in Frege's terms) in the kind of knowledge about the referent that people who master the expressions used in those sentences possess. Moreover, his proposal is meant to apply not only at the level of language but also at the level of thought or, as we would say, at the mental level. At the thought/mental level the problem is to account for how it is possible for a rational subject to take different (and even conflicting) doxastic/epistemic attitudes towards contents that ascribe the same properties to the same entity (for example believing that Mark Twain authored the *Adventures of Huckleberry Finn* while not believing or being agnostic about Samuel Clemens authoring that book).

I think these considerations concerning the extent of Frege's proposal are important in order to properly appreciate the nature of the issue that Frege was facing. What do phenomena such as informativeness of true identity statements, cognitive value and the possibility of rationally taking different (and even conflicting) attitudes towards contents that ascribe the same properties to the same entity have in common so as to be considered as aspects of a unique problem, what we now call Frege's problem? I think that it is fair to the spirit of Frege's proposal even though not to the letter of it – Frege would not have put things in this way – to say that what Frege's problem is about is the *aspectuality* of the intentional relation, that is the relation between some representational device, be it linguistic or mental or whatever, and the entity the given device is about. Both in language and in thought we aim at referring to items in

the world, be they individual objects, their properties or relations, and complexes thereof (states of affairs). We can refer to a given object in language by mobilizing an expression which stands for that object, but we can also direct our attention to that object in thinking by mobilizing some device which stands for the object. Both cases instantiate what we call intentionality or aboutness. Intentionality is characterized by two features, directionality and aspectuality. Any bearer of intentionality is directed towards something (an intentional object in the case of a mental bearer of intentionality, a referent in the case of a linguistic bearer) and what it is directed to is always presented in a given way, or as people say (Searle 1983, Crane 2001), under a given aspect. The fact that an object is always presented in a given way (that is: the fact that the intentional relation is aspectual) is what grounds the *perspectuality* of our relation to the world either in thought or in language. What we refer to is always presented/given from a particular perspective. The perspective which one enjoys in referring to an object is what makes some features of the object available to the subject. The object referred to is never completely 'on view': some of its features come into view while other stay hidden and become available as soon as a different perspective is taken. It is the aspectuality of the intentional relation that is responsible for the difference in the attitudes that a given subject may take towards the propositions expressed by sentences containing different co-referential terms. A subject may believe that Lucia, the woman he saw one night with his friends in an equivocal tavern dancing with a bird mask on her face, is a very disinhibited person while not believing that his daughter Rebecca (who is Lucia) is such, because the aspects under which he thinks of that person on the two occasions are not taken by him to belong to one and the same person (he does not 'see' Rebecca under the Luisa-aspect and he does not 'see' Luisa under the Rebecca-aspect).

In dealing with the topic of the aspectuality of the intentional relation, Frege impressed a radical twist in the way in which that topic, and more generally the topic of intentionality, was dealt with by his contemporaries, in particular by people working within the phenomenological movement. One important element of difference is that for Frege the primary bearers of intentionality (that is the entities which originally, not derivatively, possess that property) are not mental items, but rather linguistic ones. A second important difference, strictly connected with

the previous one, is that for Frege both features of intentionality, directionality and aspectuality, have to be accounted for in semantic terms. His proposal is to provide an explanation of both features in terms of his notion of sense: the sense of an expression is both what makes the expression referring to something and what accounts for the aspectuality of the referential relation.

To recap. I think it important to see Frege's problem as one hinging on the issue of aspectuality. According to Frege, what accounts for aspectuality is something which is suited to play a given role, namely, the mode of presentation role. He deemed senses to be able to play that role and on this ground he claimed that senses are modes of presentation. But are they? Moreover, and in the negative case, are there other entities that are suited to play the mode of presentation role?

As regards the first question, there is an almost general consensus in the analytical philosophical community that, *pace* Frege, senses are neither necessary nor sufficient for dealing with the issues that Frege was facing. They are not sufficient because, as Mates (1952) has shown, Frege's cases can arise even with synonymous expressions (expressions which, by having the same meaning, should also have the same sense, if it is true, as Frege claimed, that the sense of an expression is what competent speakers grasp when they understand the expression and so know its meaning).

Actually, there have been many attempts by the Fregeans to address these criticisms, either by divorcing meaning and sense, by claiming that in doubly (or higher level) indirect contexts (such as those mobilized by Mates' cases) the sense of an expression is not its ordinary sense but an indirect one which can be different even for expressions having the same direct sense, or by making senses more and more fine-grained and cumbersome in order to deal with cases such as Kripke's (1979) "London"/"Londres" one and "Paderewski" one. As things stand no such attempt has been immune to further and even more radical criticisms. As Kent Bach (1997) has shown, no matter how fine-grained senses are taken to be it is always possible to devise further versions of Frege's puzzles. So much for the sufficiency issue. As for the necessity issue, Frege's critics have claimed that senses are not necessary either, because what they have been introduced to explain can be accounted for without adverting to them and so without having to introduce in the semantics

of the expressions any other level besides the referential one. That is good news for people wanting to keep semantics to the minimum, like those who adhere to the direct reference picture and/or are willing to pursue some kind of naturalization project of linguistic or mental content. The general consensus among people working in this framework is that whereas senses are not needed, what is needed is some other kind of mode of presentation.

In the next section we shall take into account some varieties of the syntactic account which represents the most widespread alternative to Frege's approach. The question we shall consider is whether an expression (or more generally, a vehicle of content) is suited to play the mode of presentation role. In order to answer that question we need to put forward the requirements that something is expected to satisfy in order for it to be eligible as a candidate for that role. What are then those requirements? On the ground of what has been said above about the extent and scope of Frege's problem and on the ground of the functional nature of the technical notion of mode of presentation the three following requirements suggest themselves:

(i)   A mode of presentation has to account for the informativeness of true identity statements of the form "$a = b$";

(ii)   A mode of presentation has to account for differences in cognitive values between coreferential expressions;

(iii)   A mode of presentation has to account for how it is possible for a rational subject to hold different/conflicting attitudes towards one and the same object.

The question we shall consider in the foregoing is whether satisfying these three requirements is sufficient in order for something to prove acceptable. As far as I know, all the people who defend some version of the syntactic account deem the satisfaction of those three requirements not only as necessary but also as sufficient. What I want to show is that they are wrong in so claiming. Actually, on my view, one more requirement is needed, namely the following:

(iv) A mode of presentation has to account for how a given subject
      takes the object she is referring to on a given occasion.[4]

Since, as I shall try to show, I deem this requirement the most basic
one (in the sense that it is the prerequisite in order for the other three re-
quirements to be satisfied in a non-circular and non-vacuous way) I shall
characterize *a mode of presentation as the entity sameness/difference of
which accounts for the subject's taking/non-taking what she is referring
to on a given occasion (either in language or in thought) to be the same
as what she is referring to on another occasion.* My thesis is that repre-
sentational vehicles are unsuited to satisfy this requirement. Therefore,
by making use of the characterization provided of the notion of mode of
presentation, I shall claim that representational vehicles do not qualify
as the right kind of entity suited to play the mode of presentation role.

3. From semantics to syntax: modes of presentation as vehicles of representation

In this section I shall consider the main features of one of the most impor-
tant strategies put forward in the attempt to deal with Frege's problem
without introducing senses. What characterizes this kind of non-semantic
strategy is the idea that the modes of presentation that are needed to
account for (i) the informativeness of true identity statements of the
form "$a = b$", (ii) differences in cognitive values between coreferential
expressions and (iii) how it is possible for a rational subject to hold
different/conflicting attitudes towards one and the same object, are the
expressions themselves (or more generally the vehicles of content). This
proposal is meant to apply both at the linguistic and at the mental level.
As for the linguistic level the idea, as it comes out for example from
Napoli's and Perry's quotes we considered in the introduction, is that
what accounts for sameness and difference of cognitive contribution is
just sameness and difference in the words used. The idea that expressions
and not senses have the right kind of fine-grainiess to address Frege's

---

[4] The notion of *taking* (how a given subject takes the object she is referring to) will
be central in my positive account of the issue, as it will come out in the fourth section.

problem is what, according to many people, teach us Mates' and in a certain sense also Kripke's cases (in particular the "London"/"Londres" one). As I said, this move, which can be qualified as syntactic, is meant to apply also at the level of thought. At first stake this could look puzzling, in so far as it is not obvious in which sense expressions, or in any case symbolic material, could be involved in thinking (not just in ascribing to oneself or to other people contentful mental states, but in entertaining mental states themselves). If the extension of the syntactic strategy from the linguistic to the mental level is not automatic, it becomes quite natural once a certain picture of the nature of mental states is taken on board. This picture is the representational-computational theory of the mind (RCTM) which has played and still plays a crucial role within cognitive sciences. According to this theory, to be in a mental state is to stand in a peculiar relation with a mental representation. The relation in question, which is characterized in functional/causal terms (i.e. in terms of the several input/output and internal relations that the state entertains with the world and with other mental states), is meant to account for the way in which a given content is entertained (in the belief modality, in the desire modality and so on and so forth) and therefore to account for what is called the "intentional mode" of the act. The representation (or better its semantic properties) is instead taken to account for the content of the mental state; a mental state represents what it does (that p, that grass is green for example) in virtue of its relation with a mental representation that means that p. With this picture of the nature of mental states in place it is easy to see how the extension of the syntactic strategy from the linguistic to the mental level is possible. Any kind of content, be it linguistic or mental, is the content of some kind of vehicle. Linguistic expressions are the vehicles of linguistic content and mental representations are the vehicles of mental content. Vehicles have both syntactic and semantic properties. According to the syntactic strategy what is needed to account for Frege's problem are just the syntactic/formal properties of the content's vehicles: two vehicles with the same semantic properties can have different formal properties, as it happens with "Hesperus" and "Phosphorus" (both in their linguistic and in their mental realizations); those properties, or better the vehicles themselves individuated in terms of them, can account for all those issues for

which modes of presentation are required. So much for how the syntactic strategy has to be conceived at the mental level.

In the foregoing, in discussing the syntactic strategy, I shall confine my attention to the proposals put forward at that level. There are two reasons for this. The first reason is that as a consequence of the 'cognitive turn' in the theory of content there is an almost general consensus that most issues previously addressed at the linguistic level can be more fruitfully addressed when formulated at the mental level, which is taken to be prior in the order of explanation. According to this priority claim, the primary bearers of content are mental vehicles. Mental vehicles have original intentionality, whereas linguistic vehicles have only derived intentionality, in the sense that they have content or meaning only in virtue of their association with mental representations that have meaning non-derivatively (on pain of a vicious regress). If intentionality primarily concerns mental items, the same holds of its two main features: directionality and aspectuality (that feature of the intentional relation which we have claimed is at the heart of Frege's problem). As for the second reason, I prefer to discuss the syntactic strategy at the mental level because it is at that level that I think its inadequacy comes out in its clearest form.

Actually, there are different varieties of the syntactic strategy at the mental level. Even though I want to keep my discussion general, because I think that the problem I want to focus on affects all the existing varieties, let me briefly consider what distinguishes each one from the others. The pictures I take as representative, even though at different degrees, of the general strategy are: Fodor's proposal of modes of presentation as symbols of the language of thought (LOT),[5] Sainsbury and Tye's (2012) originalist theory of concepts, and Recanati's (2012) mental file theory. Of them Fodor's is the most representative of the general strategy and can be taken as its paradigm. According to Fodor, to deal with Frege's problem we need modes of presentation. Modes of presentation are required in order to distinguish co-referential concepts, that is in order to individuate the thought's constituents. In his view, modes of presentation can play this individuative role only if they are conceived as mental entities (as concrete entities in the mind/brain). It is on this

---

[5] Fodor has presented his proposal to deal with Frege's problem in different places; see in particular Fodor 1998, 2008.

ground that Fodor criticizes Frege's account of modes of presentation as abstract entities by claiming that this way of conceiving their nature cannot warrant that there be a one to one correspondence between modes of presentation and ways of thinking of an object.[6] So, according to Fodor, modes of presentation are concrete mental particulars. By combining this claim, which amounts to a sort of psychologization of Frege's senses, with Fodor's LOT hypothesis we have that what plays in his picture the mode of presentation role are the symbols of LOT, that is Mentalese words having both semantic and formal properties and a causal role. The correspondence with the syntactic strategy at the linguistic level is straightforward; that's the reason why I have said that Fodor's position is paradigmatic of the syntactic strategy at the mental level. In *LOT 2* (2008) Fodor comes back to the issue and, after having restated his claim that all that Frege's cases show is that there must be something more to the individuation of concepts than what they refer to, he concludes by saying that

> [t]here is no Frege problem for concepts that are expressed by complex Mentalese formulas (i.e. Mentalese formulas that have constituent structure). Since complex concepts can differ in their constituency, they can differ in their possession conditions [. . .]. If there is a Frege problem, it must be about how to draw the type/token relation for (syntactically) *primitive* concepts. But if there is a Frege problem about primitive concepts, then it is resolved by appeal to their form, not by reference to their content. (75)

A position very close to Fodor's is Sainsbury and Tye's originalist theory of concepts. Originalism is presented as that position which combines the best features of Fregean and Millian views. What originalism takes from Frege is the idea that to account for cognition something more than reference is needed. What it rejects is the idea that this something more needs to be epistemically individuated (as senses conceived as ways of thinking or of having knowledge of the objects are). What they take from Mill is the idea that content is referential. In their view, the extra non-semantic ingredient needed to account for cognition is provided by the vehicles of content which, as they say, "can affect cognition in ways that do not depend on their semantic properties" (2012: 26). Vehicles of cognition are concepts and "distinct concepts can, and typically will, play different roles in our cognitive activities, even if they have the same

---

[6] I have criticized this point in Sacchi 2003.

content" (53–54). Even though Sainsbury and Tye share many of Fodor's views, in particular the idea that Frege's cases can be explained without appeal to semantics, that thoughts with the same content can differ in their cognitive and causal roles, nonetheless there are some important differences between Sainsbury and Tye's position and Fodor's. The most important one concerns the individuation of the vehicles of content on which as we have said cognitive processing depends. While for Fodor the vehicles are individuated in terms of some of their intrinsic properties in the sense that, as he says, "tokens of primitive Mentalese formulas are of different types when they differ in the (presumably physical) properties to which mental processes are sensitive" (2008: 79), for Sainsbury and Tye the individuation is in terms of relational properties having to do with the historical origin of concepts (the particular event in which a given concept is introduced).

The third account I consider as representative of this general non-semantic strategy is Recanati's mental files picture. Mental files are, according to Recanati, what the cognitive system uses to refer to objects in thought; they are therefore the mental analogues of singular terms in language. Their function is to store information (or misinformation) from the object gained in virtue of standing in some direct relation with it (which Recanati calls "epistemically rewarding relations") and to represent the object in thought. The mechanism of reference determination is taken to be relational rather than satisfactional; what a given files refers, when it refers at all, is not the entity which satisfies all or most information in the file, what Recanati calls the "pile", but rather the entity which is the *relatum* of the peculiar grounding relation on which the file is based. Mental files so conceived qualify as non-descriptive and according to Recanati they are precisely the entities that play the mode of presentation role. It is worth stressing that what accounts for the way in which the object is presented is not the information contained in the file (two files can contain the same information and yet be different), but rather the file itself, that is the mental label which functions as a singular term. According to Recanati, mental labels have not only a content (a referent), when they have it, but also a functional role, which he takes

to be the mental analogue of the character of a linguistic indexical.[7] In his view, his indexical picture of mental files is able to vindicate Frege's two-layered semantics (i.e. his distinction between sense and reference) in a way which is compatible with both "singularism" (the idea that we refer directly to objects and not indirectly to them through their properties) and direct reference (taken not as the thesis that meaning coincides with reference, but as the thesis that the mechanism of reference determination is relational rather than satisfactional). On the ground of these considerations one could find inappropriate to qualify Recanati's account of modes of presentation as a variant of the syntactic account. But this initial impression dispels as soon as one considers that nothing which properly qualifies as semantic plays a role in the individuation of mental files. As in Fodor's picture, what play the mode of presentation role are the vehicles of content. As a matter of fact there are differences between the two positions: Fodor's vehicles are Mentalese words, whereas Recanati's vehicles are something like 'mental paragraphs' having not only a content but also a mental character. But in the substance, the two positions can be taken as notational variants of one and the same strategy (Fodor explicitly says that his account is compatible and can be framed in the mental files framework,[8] and Recanati stresses a similar point about his own account).[9]

---

[7] The function of a given vehicle is the storing of information gained through acquaintance and according to Recanati it is the files's function that individuates singular vehicles qua type.

[8] As regards the file picture Fodor says: "In effect, according to this story, *we think in file names*; tokens of file names serve both as the constituents of our thoughts and as the Mentalese expressions that we use to refer to things we think about [. . .]. That one thinks in file names is the best short summary I'm able to formulate of the version of RTM that I'm currently inclined to endorse" (2008: 95). It is worth stressing that, even though Fodor is willing to construe representational vehicles as mental files, he does not take them as an extra level of content. For him mental files are purely syntactic.

[9] "So what is the difference between the view I have expounded and the view, argued for by Fodor and by Sainsbury and Tye, that modes of presentation are syntactic? Not much, since I accept that mental files are representational vehicles. The difference is merely terminological [. . .]. To be sure, mental files are not purely syntactic entities, in my framework; they have a function or a role which determines their cognitive significance" (2012: 245).

Having clarified some of the main differences among the three varieties of the syntactic account in the foregoing we shall keep our discussion at a general level. What all the three theories challenge is the idea that cognitive differences must be reflected by semantic differences; Fregean data are to be explained by appeal to sameness and difference in vehicles of content, rather than sameness and difference in content.[10] Ultimately, on their view, *what Fregean arguments entail is just that a difference is needed*, not that a difference of content is needed, or that any semantic difference is needed.

This claim will be the starting point of my critical assessment of the syntactic strategy. The issue I want to raise is the following: let us concede that people who endorse it are right in claiming that in order to address Frege's problem there is no need to introduce any difference at the level of semantic content. Does it follow from this that any difference which is able to affect cognition will do?

In the foregoing I shall try to argue that at least one more requirement is needed, namely: that the suggested difference be able to account for how the object referred to is taken by the subject (this was the fourth requirement in my list). For, if this last requirement is not met the account provided is doomed to be circular or explanatorily vacuous (that's why I claim that this requirement is the most basic one).

Let us consider how people adhering to the syntactic approach account for the difference between episodes of thinking which are about the same object such as for example the episode of thinking that *Cicero is Roman* and the episode of thinking that *Tully is Roman*. According to them, what individuates a given thought is, inter alia, the vehicles that get mobilized in thinking the thought. Therefore, what one has to appeal to in order to explain Frege's data is precisely the vehicles of thought. According to Fodor the situation is explained by saying that on the two occasions the subject is related to different Mentalese words, where what makes them different are some physical properties.[11] For Sainsbury and Tye, the two episodes of thinking mobilize concepts which, by having been introduced on different occasions, count as distinct and this in their view provides a sufficient explanation. For as they claim, what is

---

[10] This is a point that has been stressed also by Millikan (1991, 1993).

[11] "Tokens of Mentalese formulas are of different types when they differ in the (presumably physical) properties to which mental processes are sensitive" (2008: 78).

needed to account for Frege's data is just finding two of something and "two concepts generate two thoughts and different thoughts may have different properties" (2012: 125). According to Recanati the situation is explained by saying that the two episodes of thinking mobilize two distinct files which are not linked (in the sense that the information in one file is insulated from information in the other file).

At first stake these accounts seem to explain the situation at hand, but a more careful consideration reveals that the explanation provided does not satisfy a crucial requirement. Let me put my point in the following way. What has to be explained is how a given subject may (rationally) take different/conflicting attitudes towards one and the same object (i.e. believing that Cicero is Roman and not believing/being agnostic about Tully's being Roman). This is the *explanandum*. Intuitively, the reason why a given subject may (rationally) take different/conflicting attitudes towards one and the same object is that she does not take the object to be the same. I take this intuition to be very strong and I think that any good explanation ought to conform to it. Let us therefore consider it as a requirement on any good explanation (this was the fourth requirement on modes of presentation I listed in the second section). Let us now consider the *explanans* provided by the advocates of the syntactic strategy. Does it satisfy this requirement? That is: is it possible to explain the fact that the subject does not take what she is thinking at to be the same in terms of vehicles by saying that the reason why the subject does not take the object to be the same is that different vehicles get mobilized? Let us consider this option. We have to distinguish two cases: either the vehicles appealed to in the answer are the same as those appealed to in the *explanandum* so that they are taken to play double duty (that is: they are taken both to account for how the subject takes different attitudes towards the same object and to account for how the subject does not take what is in fact the same object as the same) or they are different vehicles more basic than the former. In the first case the answer is circular for it would explain the subject's mobilizing different vehicles in terms of her mobilizing those very same different vehicles. In the second case it wouldn't be circular (for it would explain the subject's mobilizing some kind of vehicles in terms of her mobilizing some other kind of vehicles) but it would engender a regress. For one could then raise the further question of what explains the fact that those (more basic)

different vehicles are mobilized. It seems that the fact that a subject may mobilize different co-referential vehicles on different occasions does not explain why she does not take the object to be the same. Rather the right order of the explanation seems to be the opposite: different vehicles are mobilized because the subject does not take the object to be the same.

Recanati shows acknowledgement of this problem in his discussion of what he labels the "circularity objection". Actually he discusses two versions of this objection: one concerning internal co-reference (which concerns the information within a single file) and the other concerning identity judgements. The first version of the objection goes as follows: "what is it for information to cluster into a single (conceptual) file? The answer appeals (*inter alia*) to the fact that the information in the file is taken to concern the same object, and that is circular" (2012: 99). As regards the second version of the objection Recanati (2012: 99) says: "If conceptual files themselves depend upon identity judgements [...], then we cannot analyze identity judgements in general in terms of a linking operation on files, as I have done, without launching a regress". As regards the first version of the objection, which is the most relevant one as to my present concern, Recanati considers a point raised by Bochner which in my view deserves close attention. According to Bochner

> Many advocates of mental files [...] acknowledge – as if this were compatible with what they claim – that a mental file is created when an object is *taken* to be one by the subject. But [...] if you already need to think of the object *in order* to determine that it is a single object deserving a single location in your syntax, then this means that you must be able to think of the object *prior* to the attribution of a vehicle or mental file. And, presumably, if some identity mistakes is made in this early process of syntactic assignment – if for instance, two different vehicles are created for a unique object taken to be two distinct objects – it will be *that* early mistake that will explain cognitive significance, not the fact that there are two vehicles [...]. All of this is incompatible with the idea [...] that it is differences in syntax that determine differences in cognitive significance, and, instead, squarely supports the opposite view that it is differences in cognitive significance that determine differences in syntax. (Bochner 2010 in Recanati 2012: 97)

I think that the problem raised by Bochner is a very serious one. Moreover, I think that it affects not only the account of modes of presentation in terms of mental files, but rather any version of the syntactic strategy. That's why I prefer not to enter into the details of Recanati's suggestion for dealing with the circularity objection which is obviously tailored to

his own explanatory schema.[12] I think that Bochner's intuition according to which the way out of the circularity objection is to account for the issue of what grounds the subject's taking the object to be the same or different in terms of modes of presentation that are more basic than syntactic vehicles is on the right track. The crucial point will be to determine what kind of basicness is appropriate. If Bochner's suggestion on this issue goes in the direction of descriptive senses mine will be neither a semantic nor a syntactic account.

Before presenting my positive proposal let me recap the main lines of my criticism. I have tried to contest the claim, which is shared among people endorsing the syntactic account, that all that is required in order to address Frege's problem is to find something which can make a difference at the cognitive level. The gist of my criticism has been that if the suggested difference does not account for how the object thought about is taken by the subject, the account provided ends up being either circular or explanatorily vacuous. Representational vehicles do not satisfy this requirement and therefore they prove unable to properly address Frege's problem. Since to be able to address it is a *conditio sine qua non* in order for something to qualify as a candidate entity for the mode of presentation role, it follows that, *pace* Fodor and the other supporters of the syntactic strategy, representational vehicles cannot be modes of presentation. In my view, the problem with the representational vehicle's strategy has ultimately to do with its sub-personal level of explanation of what affects cognitive value. I do not want to question the relevance in the explanation of cognition of what goes on at the sub-personal level. My point is rather to contest that that is the right level for the individuation of the ways of thinking of the objects which are responsible for how the subject (personally and consciously) takes the object she is thinking about. The right level is not the sub-personal but rather the personal one, that is the level of first person awareness.

---

[12] For Recanati's discussion of the objection see chp. 8 of his book.

4. From syntax to phenomenology: modes of presentations as ways of experiencing

In this section I shall introduce my non-semantic, non-syntactic solution to Frege's problem. My account qualifies as non-semantic because it neither treats modes of presentation as senses nor it equates them with some level of content. It qualifies as non-syntactic because in my view what play the mode of presentation role are not the formal/syntactic properties of the representational vehicles. In my view, even though an explanation in terms of representational vehicles considered as entities occurring at the sub-personal level of explanation could account for how two of a subject's mental episodes which are about the same object can have different causal roles, it does not explain why the subject consciously takes the object she is thinking about to be different in the two cases. What is required to that end is in my view a difference that occurs at the personal level of the explanation.

Let me illustrate my proposal by using a somewhat bizarre Frege's case which as things stand cannot be accounted for neither in terms of senses nor in terms of representational vehicles at the sub-personal level. Even though the example focuses on a perceptual Frege's case (one which involves perceptual rather than thinking episodes) I intend my subsequent considerations to apply also at the level of thought.

The example goes as follows.[13] Let us consider two sensory experiences (e_1 and e_2) of a subject named Tassandra. In the past Tassandra has been told by a friend of her while looking in t_1 at a picture of Venus on a book they were leafing through "This is Hesperus" and by the same friend in t_2 (pointing unbeknownst to Tassandra to the same picture again after having browsed the book back and forth) "Look how bright Phosphorus is". Let us now consider the situation having Tassandra looking at the planet on two different occasions. Let us stipulate that on those two occasions the illumination conditions are identical, the position of Venus in the sky is identical, Tassandra visual apparatus is identical, Tassandra is unaware of whether it is morning or evening and of whether Phosphorus/Hesperus is visible in the morning/evening. Given these stipulations it follows that there is no property our subject

---

[13] The example is adapted from Kriegel 2011.

is willing to ascribe to Hesperus that she is not willing to ascribe to Phosphorus. All this notwithstanding, Tassandra does not take what she is confronted to in the two situations as the same object. Rather she takes it as different. In e_1 she has a 'strong gut feeling' that she is looking at Hesperus (if asked she would say that it is Hesperus she is facing) while in e_2 she has a 'strong gut feeling' that she is looking at Phosphorus (if asked she would say that it is Phosphorus she is facing).

From Tassandra's own point of view, the reason why she takes the object as different is that the way in which the object impinges on her conscious experience in the two situations is different. Granted that there is a difference between the way in which Venus is given to Tassandra in the two situations, and that this difference is responsible for Tassandra's not taking what she is facing at as one and the same object, how can this difference be accounted for? According to my proposal, which, I repeat, I take to be applicable not only at the level of perceptual experience but at the more general level,[14] the difference in question has to be explained in phenomenal terms, that is by adverting to some phenomenal property of Tassandra's conscious experiences that is different in the two cases. A way in which the difference in question can be put is the following:[15] what e_1 instantiates is the property of *presenting Venus Hesperus-wise* (presenting Venus in such a way as to make Tassandra aware that she is seeing Hesperus); what e_2 instantiates is the property of *presenting Venus Phosphorus-wise* (presenting Venus in such a way as to make Tassandra aware that she is seeing Phosphorus). Another way to express this point is to make use of the somewhat technical notion of 'phenomenal character', where the phenomenal character of a conscious mental state is that property of it that is responsible for there being something it is like for the subject of the state to be in that state.[16] We can then say, by mobilizing this notion, that Tassandra's two conscious experiences have *different phenomenal characters*: e_1 has a *Hesperescent* phenome-

---

[14] Of course this implies a commitment to the claim that any occurrent mental episode has or exemplifies some phenomenal property. For a defense of this position, which I cannot discuss here, see e.g. Horgan and Tienson 2002; Kriegel 2011.

[15] This is the way in which for example Kriegel 2011 puts the difference between the aspectual shape of Tassandra's two perceptual episodes.

[16] The characterization of phenomenal character in terms of the notion of 'what-it-is-like' comes back to Nagel 1974.

nal character while e_2 has a *Phosphorescent* phenomenal character. It is this difference that accounts for Tassandra's not taking the object to be the same. So, on the ground of our initial characterization of a mode of presentation as the entity sameness/difference of which accounts for the subject's taking/non-taking what she is referring to on a given occasion to be the same as what she is referring to on another occasion, we conclude that the entities that play the mode of presentation role are phenomenal characters.

What Tassandra's example shows is that what matters for aspectuality is not how the representational system represents the objects to be, but rather how the object is *(experientially) taken* by the subject. The notion of *experiential taking* is central in my phenomenological account of aspectuality.[17] Let me provide some elucidation of this notion. I take experiential taking to be a matter of experientially identifying something (identifying something as the object one is experiencing either in perceiving the object or in thinking about it). Experientially identifying something is always relative to a given subject and to her experiences (what I experientially identify is what is presented to me in this very experience in this peculiar phenomenal way). A subject's experiential taking is what grounds the subject's awareness of the object: a subject is aware of something insofar as the object is the target of one of the subject's experiential takings.[18] Ultimately, what accounts for the fact that an object is always given in a certain way to the subject is the intrinsically perspectival nature of what grounds the subject's awareness of the object, i.e. the subject's experiential taking. My proposal is therefore

---

[17] The first occurrence of the notion of *taking* is in Chisholm 1957 where the author claims that the *act of taking* (which he characterizes as the act by which the mind takes something to be thus and so) is the more basic intentional act on which the intentionality of any other less basic intentional states depend.

[18] My notion of 'experiential taking' has strong similarities with the notion of 'phenomenal taking' which appears in Strawson 2008. Unlike Strawson, however, I stay neutral as regards the nature (cognitive or sensory) of the taking which he instead takes to be cognitive (a cognitive component of the state's phenomenal character). Another important difference has to do with the theoretical role that this notion is meant to play in our respective proposals. While for Strawson the taking is meant to fix the state's directionality, for me it rather fixes the mental state's aspectual shape. This difference has important consequences on the ongoing debate on what is called "phenomenal intentionality" (see e.g. Kriegel 2013).

that modes of presentation have to be characterized at the phenomenal level. The fundamental difference between my phenomenal modes of presentation and those employed in the several variants of the standard semantic account of aspectuality is that while the latter are properties of the objects which are involved in the mental state's content, phenomenal modes of presentation are properties of the subject's experience of the objects (properties which characterize the way in which the object is experienced).

As is well known, the idea that aspectuality is grounded on (phenomenal) consciousness has been championed by Searle (1992). Searle's argument for the claim that "only conscious intentional states are intrinsically endowed with aspectual shape" is actually a sub-argument of a more complex argument for the thesis of the asymmetric dependence of intentionality on consciousness (which is called the "Connection Principle" (CP)).[19] Searle has been criticized on the ground that it leaves the connection between aspectuality and consciousness unexplained.[20]

---

[19] Searle characterizes CP as the claim that every unconscious intentional state is potentially conscious. CP is arrived at through the train of argument that I sketch below (following the reconstructions provided by Kriegel 2003 and Shani 2007). As it can be seen, Searle's claim that aspectuality is grounded in consciousness figures in the argument as a premise (P3) in the overall argument.

P1:   Unconscious intentional states are intrinsically intentional;
P2:   All intrinsically intentional states are endowed with aspectual shape;
P3:   Only conscious intentional states are intrinsically endowed with aspectual shape.

Therefore [by P1–P3]

[Derivability]: The aspectual shape of unconscious intentional states is derivative from the aspectual shape of conscious intentional states;

P4:   The only explanation for Derivability is that every unconscious intentional state is potentially conscious, i.e., it is the sort of thing that could be, or could have been, conscious, hence could be or could have been an intrinsic possessor of aspectual shape.

Therefore CP [by Derivability and P4].

[20] The sub-argument that Searle puts forward in support of P3 (the third premise of the argument in support of CP) is the following:

According to some people,[21] even though Searle claims that aspectual shape is a basic feature of representation, nothing in what Searle says about consciousness, or about aspectual shape, explains why conscious intentional states possess aspectual shape to begin with.[22] Searle's argument (the one I sketched in note 20) is based on two assumptions:

A1. Aspectuality (intending an object under some aspect) depends on (presupposes/is grounded in) perspectuality (intending the object from a point of view);

A2. Perspectuality depends on (presupposes/is grounded in) consciousness.

Searle's critics have claimed that the idea that there is such a strong association of perspectuality with consciousness is neither an evident truth nor a well-motivated empirical hypothesis.[23] I do agree that Searle's account as it stands does not explain the connection between aspectuality and consciousness. People are right in claiming that perspectuality does

1. Unconscious intentional states, when unconscious, are nothing but neurophysiological states;
2. Neurophysiological states are completely objective;
3. Unconscious intentional states, when unconscious, are completely objective [by 1 and 2];
4. Aspectual shape is, necessarily, a subjective phenomenon;
5. So there is no manifestation of aspectual shape in unconscious intentional states when unconscious and insofar as they are unconscious [by 3 and 4];
6. Aspectual shape is an intrinsic property of conscious intentional states.

Therefore

P3: Only conscious intentional states are intrinsically endowed with aspectual shape [by 5 and 6].

[21] Kriegel (2003) and Shani (2007), among others.

[22] According to Shani, for example, "Searle seems to be relying here on the fact that the aspectuality of intentional states is evident in conscious experience but this, in itself, does little to explain how what thus becomes evident in experience is possible in the first place" (2010: 326).

[23] According to Shani, "Searle's second assumption is at least partly wrong since a minimal form of first person perspective is co-emergent with basic organic capacity for autonomous conduct, a capacity manifested in creatures that, on Searle's own terms, are far too simple to be considered conscious" (2010: 326).

not necessarily presuppose consciousness. Perspectuality *per se* cannot be the right intermediate link able to connect aspectuality and consciousness. I too believe that there is a missing link in Searle's account, but I also believe that this link is provided by experiential taking. According to my account, what presupposes/depends on/is grounded in consciousness is a particular form of perspectuality, a form that qualifies as experiential and ego-oriented. This is the kind of perspectuality brought about by experiential taking.

On the ground of these considerations my suggested revision of Searle's claim concerning the connection between aspectuality and consciousness is the following:

(i)   Aspectuality is grounded in experiential taking (there is no genuine aspectuality without experiential taking);

(ii) Experiential taking is grounded in (phenomenal) consciousness (here the 'grounding relation' is outright identity: experiential taking is identical with a component of the state's overall phenomenal character).

This provides in my view an explanation of what was unaddressed in Searle's position: what makes conscious states aspectual in the first place is a component of the state's phenomenal character. It is this component, the experiential taking, that brings about ego-oriented, experiential perspectuality.

Let me conclude with some general considerations having to do with the compatibility or incompatibility of my phenomenological proposal with the semantic and the syntactic ones. As regards the semantic account, I think that my proposal is compatible with the idea that modes of presentation belong to some kind of content of the mental intentional episode, namely to what people call phenomenal content,[24] and therefore it can be compatible with the account of Frege's problem in semantic terms. I personally feel unease with such a notion of content,[25] but if one does not share my embarrassment she is free to interpret my suggestion along these lines. As regards the syntactic account, I think that what my criticisms show, if they are correct, is that this account

---

[24] For the notion of phenomenal content see e.g. Montague 2010.

[25] I have articulated my main lines of resistance to the idea of phenomenal content in Sacchi 2011.

cannot be the ultimate answer to the issue concerning the nature of modes of presentation and that representational vehicles must depend on more basic modes of presentations. Nothing prevents one from considering those more basic modes of presentation, or better the phenomenal properties that individuate them, as being instantiated by representational vehicles, provided that representational vehicles are conceived not as entities at the sub-personal level having only semantic and formal properties, but as entities at the personal level. The idea that mental representations have not only semantic and formal properties but also qualitative/phenomenal ones is at odds with the way in which mental representations have been conceived within the traditional computational framework that has dominated until recently the debate in the cognitive sciences. But things are now changing and attempts to reject the symbolic (a-modal) format of mental representations and replace that format with a different (modal, sensorimotor) one are not lacking in the philosophical/psychological current literature.[26] Within the modal account, the idea that the mental representations that we deploy in thinking can possess phenomenal properties is as unproblematic as the parallel idea concerning the mental representations that we deploy in perceiving. The reason for this is that according to the modal account cognition involves the very same representations that are used in perception and action by a process of re-activation.

Finally, I think that the phenomenal account of modes of presentation could also be embraced by people adhering to the direct reference theory, because, as a matter of fact, it does not require to introduce any other level of meaning besides reference.

As I said at the beginning, the idea that syntactic modes of presentation are grounded in some other, non-linguistic and more basic ones is a point that also Ernesto stresses in his paper by discussing what he labels the "mission problem" (a problem which should show that knowledge of meaning/reference does not guarantee knowledge of synonymy). The case he considers is one in which an absent minded priest, engaged in baptizing a group of children in a queue, happens to administer to a child already baptized "Bob" a second baptism by using the name "Tom". What Ernesto says as regards this case is the following:

---

[26] One of the most articulated proposals of this new approach has been developed in Barsalou 1999 and Barsalou 2008.

The linguistic question: Are "Bob" and "Tom" coreferential? is rooted on the empirical question: Is the child that was in front of the missionary on such a day at such a time the same as that was in front of the missionary on such a day at such different time? The boy presentations in front of the missionary were different no matter if they were or not presentations of one and the same boy. And whether the five o'clock boy presentation is the five thirty boy presentation is an empirical matter that no convention can settle. (1997: 199)

As it turns out we have very different ideas of the kind of basicness that is needed, maybe because he considers the issue at the linguistic level while I have raised it at the mental level. But the intuition that differences at the level of expressions must be grounded in some further level is shared. And this is where our two, otherwise different proposals, meet.

References

Bach, K. (1997), "Do belief reports report beliefs?", *Pacific Philosophical Quarterly*, 78, pp. 215–241.

Barsalou, L. (1999), "Perceptual symbol systems", *Behavioral and Brain Sciences*, 22, pp. 577–609.

——— (2008), "Grounded cognition", *Annual Review of Psychology*, 59, pp. 617–645.

Bochner, G. (2010), "Cognitive significance and non-descriptive senses", MS.

Chisholm, R. (1957), *Perceiving. A Philosophical Study*, Cornell University Press, Ithaca, NY.

Crane, T. (2001), *Elements of Mind. An Introduction to the Philosophy of Mind*, Oxford University Press, Oxford.

Fine, K. (2007), *Semantic Relationism*, Blackwell, Oxford.

Fodor, J. (1998), *Concepts: Where Cognitive Science Went Wrong*, Oxford University Press, New York.

——— (2008), *LOT 2. The Language of Thought Revisited*, Clarendon Press, Oxford.

Frege, G. (1892), "Über Sinn und Bedeutung", *Zeitschrift für Philosophie und philosophische Kritik*, 100, pp. 25–50, translated as "On Sinn and Bedeutung" in *The Frege Reader*, M. Beaney (ed.), Blackwell, Oxford 1997, pp. 151–171.

Horgan, T. and Tienson, J. (2002), "The intentionality of phenomenology and the phenomenology of intentionality", in D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press, Oxford, pp. 520–533.

Kriegel, U. (2003), "Is intentionality dependent upon consciousness?", *Philosophical Studies*, 116, pp. 271–307.

——— (2011), *The Sources of Intentionality*, Oxford University Press, Oxford and New York.

Kriegel, U. (ed.) (2013), *Phenomenal Intentionality*, Oxford University Press, Oxford.

Kripke, S. (1979), "A puzzle about belief", in A. Margalit (ed.), *Meaning and Use*, Reidel, Dordrecht, pp. 239–283, reprinted in his *Philosophical Troubles. Collected Papers, Volume I*, Oxford University Press, Oxford 2011, pp. 125–161.

Mates, B. (1952), "Synonymity", in B. Linsky (ed.), *Semantics and the Philosophy of Language*, University of Illinois Press, Urbana, IL, pp. 111–138.

Millikan, R. (1991), "Perceptual content and the Fregean myth", *Mind*, 100, pp. 439–459.

——— (1993), *White Queen Psychology and Other Essays for Alice*, MIT press, Cambridge, MA.

Montague, M. (2010), "Recent work on intentionality", *Analysis Review*, 70, pp. 765–782.

Nagel, T. (1974), "What is it like to be a bat?", *Philosophical Review*, 83, pp. 435–450.

Napoli, E. (1997), "Names, indexicals, and identity statements", in M. Anduschus, A. Newen and W. Künne (eds.), *Direct Reference, Indexicality and Propositional Attitudes*, CSLI Publications, Stanford, pp. 185–211.

Perry, J. (2015), "The cognitive contribution of names", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 189–208.

Recanati, F. (2012), *Mental Files*, Oxford University Press, Oxford.

Sacchi, E. (2003), "Fodor e la psicologizzazione dei sensi fregeani", in M. Carrara, G. De Anna and S. Magrin (eds.), *Linguaggio, mente e mondo. Saggi di filosofia del linguaggio, filosofia della mente e metafisica*, Il Poligrafo, Padova, pp. 81–119.

——— (2011), "A phenomenologically oriented account of the phenomenon of aspectuality in propositional attitudes", *Phenomenology and Mind*, 1, pp. 73–91.

Sainsbury, M. and Tye, M. (2012), *Seven Puzzles of Thought*, Oxford University Press, Oxford.

Searle, J. (1983), *Intentionality*, Oxford University Press, Oxford.

——— (1992), *The Rediscovery of the Mind*, MIT Press, Cambridge, MA.

Shani, I. (2007), "Consciousness and the first person: A critical appraisal of Searle's connection principle", *Journal of Consciousness Studies*, 14, pp. 57–91.

——— (2010), "Representation and aspectual shape", *New Ideas in Psychology*, 28, pp. 324–337.

Strawson, G. (2008), "Real intentionality 3: Why intentionality entails consciousness", in *Real Materialism and Other Essays*, Oxford University Press, Oxford, pp. 279–297.

Part II
Truth and logical consequence

Truth: Some preliminary considerations*

Andrea Bianchi

Among the many things I have learned from Ernesto Napoli over about twenty years of intense intellectual exchange with him, one is certainly that *reference*, conceived as a *direct*, conventional, relation which most single words (not only proper names but also common nouns, verbs, and adjectives) have with certain worldly entities (for this construal of reference, see especially Napoli 1995), is central to the *explanation* of many linguistic phenomena and properties. One of the latter, no doubt philosophically interesting, is *truth*. Indeed, Ernesto never tired of claiming that it is not that reference depends on truth, as in the course of the twentieth century many philosophers of different tendencies (e.g., Quineans, Davidsonians, and arguably at least some Fregeans) have contended. On the contrary, it is truth that depends, in a way that obviously needs to be

specified, on reference.[1] Unfortunately (but perhaps wisely, for reasons that will become clear at the end of this paper), Ernesto did not develop this claim into a *theory* of truth. In what follows, I aim to make some preliminary, and not especially original, considerations looking ahead to such a theory, and to show why it is so difficult to complete the task.

As a starting point for my considerations, I shall use two simple, perhaps even trivial, observations. The first comes from J.L. Austin, who, in a footnote to his famous article on truth, blames "'coherence' (and pragmatist) theories of truth" for failing to appreciate "the trite but central point that truth is a matter of the relation between words and world" (1950: 130n). The second is what, in his weighty investigation on truth, Wolfgang Künne calls a *truism* that philosophers have often taken as "a preparatory step on their way towards more demanding accounts of truth" (2003: 334). As Künne writes, in fact, "[a]ll philosophers … would most cordially agree that what you say or think is true if and only if *things are as you say or think they are*" (*ibid.*). I'm not completely certain that all philosophers would really cordially agree on this – what about sympathizers with the coherence or pragmatist theories mentioned by Austin? – but Austin's and Künne's simple observations seem to me to be two firm standpoints. The challenge, then, is to make a theory out of them. Unfortunately, it is not an easy challenge, and all the attempts that have been made in this direction are, for one reason or another, unconvincing. Elsewhere I have criticized, for example, Künne's articulation of his truism into what he has called the *Modest Account* of truth – in symbolic notation, "$\forall x(x$ is true $\leftrightarrow \exists p(x = [p] \wedge p))$", with the non-standard existential quantification into sentence position being objectual over propositions and the square brackets forming, "from a sentence which expresses a particular proposition, a singular term which designates that proposition" (2003: 337).[2] And correspondence theories,

---

[1] In Ernesto's writings, this issue pops up in various places, but only in scattered remarks concerning semantics, e.g. as the claim that reference precedes and is at the basis of evaluation (Napoli 1995: 326, 334; Leonardi and Napoli 1995: 256; Bianchi and Napoli 2004: 175, 179–181, 202 n. 15, 222–224). My focus in this paper will be on the metaphysical side of the issue (which is touched upon in Napoli 2010: 295–299). For elaboration on the semantic aspect, see Joseph Almog's contribution to this volume.

[2] See Bianchi 2010. In particular, I expressed perplexities concerning the appeal to propositions in the account and the (consequential?) 'obliteration' of reference from

towards which, although with some reservations, Austin was leaning, face well-known problems. What are the entities that would correspond to true (but not to false) truth(-value) bearers? Appealing to *facts*, as many did in the twentieth century, under the influence of G.E. Moore and Bertrand Russell, seems to me to lead to pseudo-explanations only, as facts appear to be no more and no less than *shadows* of true truth(-value) bearers. As W.V. Quine famously wrote, if we say that "a sentence is true if it reports a fact",

> we have fabricated substance for an empty doctrine. The world is full of things, variously related, but what, in addition to all that, are facts? They are projected from true sentences for the sake of correspondence. (1987: 213)

Donald Davidson has put the objection in the following way:

> Truth as correspondence with reality may be an idea we are better off without . . . . The formulation is not so much wrong as empty . . . . The trouble lies in the claim that the formula has explanatory power. The notion of correspondence would be a help if we were able [to] say, in an *instructive* way, which fact or slice of reality it is that makes a particular sentence true. No one has succeeded in doing this. If we ask, for example, what makes the sentence "The moon is a quarter of a million miles away" true, the only answer we come up with is that it is the fact that the moon is a quarter of a million miles away . . . . [W]e must, I think, accept the conclusion: there are no interesting and appropriate entities available which, by being somehow related to sentences, can explain why the true ones are true and the others not. (2000: 5-6)

Thus, Davidson concludes, "correspondence theories are without explanatory content" (2000: 8; for some converging considerations, see Strawson 1950). I agree.[3] But then, what kind of theory should Austin's and Künne's simple observations incline us to accept?

---

it. I shall return to both issues here. See, however, Künne 2010: 89–98 for a reply (focused on the first perplexity more than on the second).

[3] Ernesto agrees too (see Bianchi and Napoli 2004: 189–191). I should note here that some might classify even the theory I shall look ahead to as a correspondence theory. Künne probably would, Michael Devitt certainly does (see footnote 7 below). See also Glanzberg 2014: secs. 3.1 and 3.2, and David 2015: sec. 7.2. Although there is perhaps a vague sense according to which such a theory may be said to take true truth(-value) bearers to *correspond to reality*, I believe that this classification is seriously misleading, insofar as it suggests that the theory claims that there are entities to which true, but not false, truth(-value) bearers correspond. For the same reason, I would resist the classification of the theory among realist ones, if realism about truth consists in holding that "truth involves an appropriate relation between

Before moving on, some clarification is needed. Unfortunately, "theory of truth" is an expression that has been used to cover very different "projects", as Richard Kirkham (1992: ch. 1) has convincingly argued. Thus, to avoid misunderstanding, it is important that I make clear what I take a theory of truth to be. First of all, I do not take it to be a theory about the uses or even the semantic properties of the English noun "truth" and adjective "true", although it may have some indirect bearing on them as well (especially, with regard to the semantics of certain predicative uses of the adjective). Secondly, I do not take it to be an *analysis* of the *concept* of truth, or anything that can be gotten through linguistic understanding or any other *a priori* activity.[4] Rather, I take it to be an empirical investigation into the *nature* of a *property*.[5] Certain entities – we shall discuss in a moment which ones – have the property of being true, and the aim of a theory of truth in this sense is to establish, in general, what it is for them to have it, and in virtue of what they do have it. Thus, in a nutshell, the theory must take the following form:

(1) $\forall x(x \text{ is true} \leftrightarrow Px)$,

with "x" ranging over truth(-value) bearers and "P" being a signpost for a compound predicate whose non-logical constituents express properties

---

a truthbearer and some portion(s) or aspect(s) of *reality*" (Burgess and Burgess 2011: 68). It would be far better, in my opinion, to call it a *representational theory of truth*, as Michael Lynch (2009: 22–35; 2014: 8–13) does with regard to similar ones (which he criticizes).

[4] As for the concept of truth, as opposed to the property of being true (see immediately below), my own view is that, like any other lexical concept, it is atomic, hence not susceptible to any analysis. See Bianchi 2005 for a defense of this view, and Bianchi 2011 for some reflections on its consequences on how we should conceive of *philosophical theorizing*

[5] The assumption that truth is a property and that (as such, I am tempted to say) it has a nature seems to be quite natural, and so a good starting point, but it is by no means unchallenged. That truth is a property has been denied for example by P.F. Strawson ("Truth is not a property of symbols; for it is not a property" (1949: 84)), while more recently Paul Horwich (1998: 1–2, 5, 37–40, 141–144) has advanced the startling claim that truth is a property with no "underlying nature". I shall not take issue with them here, but let me note that Strawson himself changed his mind on this and came to recognize that truth is a "genuine property" after all (see Künne 2003: 62–63). For some interesting considerations on these matters, see Devitt 2001.

and relations that are more basic than truth, and by which truth can be explained.[6] If truth is a *relational* property, as it is natural to believe ("It takes two to make a truth", says another of Austin's telling footnoted remarks (1950: 124 n. 1)), it should be expected that at least one of these constituents be a relational symbol. In order for the theory to count as an account of the *nature* of truth, it has to be *necessarily* (although not, as I have just said, *a priori*) true. Only in this case, in fact, could we say that being P *is*, or *constitutes*, being true. What is important to notice in this context is that, contrary to what Alfred Tarski (1933: 153) did and notwithstanding my sympathies for his physicalist scruples, I am not requiring that no unreduced semantic expression should occur in the compound predicate by which truth is explained. Indeed, as I wrote a moment ago, I aim to develop Ernesto's (and others', of course) claim that truth depends on reference, and any theory developing this claim will have to use the semantic verb "refer", or some equivalent of it. There can be no objection to this, provided only that there is a relation called "reference" and that this relation is more basic than truth, at least in the sense that the explanation of it does not appeal to truth. This is so if reference either is primitive or can be explained in other terms (e.g., in causal or historical ones). For obvious reasons, philosophers who have naturalistic inclinations will favour the second option (see Bianchi 2015: 95), as I do, but this is beside the point here: in both cases, in offering a theory of truth one is allowed to appeal to reference and does not need to explain it (in the same way as, say, in offering a theory of water one is allowed to appeal to oxygen and does not need to explain it).[7]

---

[6] For similar considerations with regard to theories of reference, see Bianchi 2015: 93–95.

[7] See Field 1972 for criticism of Tarski on this point. The claim that reference is "the basic semantic relation" is advanced, for example, in Leonardi and Napoli 1995: 264. In recent years, the need to appeal to reference to account for truth has been highlighted, among others, by Devitt (1997: ch. 3; 2001) and Hilary Putnam (2015). Devitt, for example, writes: "A correspondence theory of the sort I am proposing explains what it is for a sentence to be true in terms of its syntactic structure and referential relations – which each require in turn substantial theories – and, of course, the way the world is" (2001: 168 n. 17). As will become clear later, except for the qualification of the theory as a *correspondence* theory (about which, see footnote 3 above), I am in broad agreement with this, and I am sure that Ernesto is too. Unfortunately, however, Devitt did not offer any such theory.

Well, then, what is truth? Or, better, given what we have just said: What is it for certain entities to be true, and in virtue of what are they so?

In order to answer these questions and hence to reach a theory of truth by finding something to take the place of "P" in (1), it seems to me all important to be clear about the issue of the truth(-value) bearers: if we want to say what kind of property truth is, we had better say what kind of entity may have it.

In the passages by Quine and Davidson I have quoted above, truth is ascribed to *sentences*. This is hardly surprising: we are certainly all used to saying that some sentences are true and some others are false. However, in recent times a different view has become more and more popular among philosophers. According to this view, which has a long history, it is certain abstract and non-linguistic entities, *propositions*, that are the truth(-value) bearers, and as a consequence it is *propositional truth* that needs to be explained in some way. As we saw, this is, for example, what Künne does with his Modest Account.[8] But he is undoubtedly in good company.[9]

Now, to consider truth as a property of propositions seems to me to dangerously lose sight of the two simple observations by Austin (who, indeed, didn't consider truth as a property of propositions) and by Künne (who, on the contrary, has proposed a perplexing 'propositionalist' articulation of his truism). The "central point" is that "truth is a matter of the relation between words and world", says Austin. And even Künne's truism appeals to language, albeit more indirectly, since it ascribes a truth-value to *what one says*, and to say something we do need words. But where do *words* end up, in Künne's Modest Account or in any other theory of truth that takes propositions to be the truth(-value) bearers?

There is, of course, an answer that propositionalists can give to this question. We are not denying, they can say, that certain linguistic expres-

---

[8] See Künne 2003: 249–269 for his defense of the claim that propositions are the primary truth(-value) bearers. The first of the two perplexities I mentioned in footnote 4 concerned precisely this (Bianchi 2010: 66–71). See Künne 2010: 89–95 for a reply.

[9] Just to give only a few almost randomly chosen examples, see Strawson 1950: 147–149, and 1964: 166–170; Horwich 1998: 16–17, 86–103, 129–135; Soames 1999: 18–19; Lewis 2001: 276. Even a first, and otherwise excellent, introduction to the theories of truth such as Volpe 2012 takes this view almost for granted, without seriously considering any alternative (see pp. 13–19).

sions, sentences, are true, or false. But everyone should agree that they are true, or false, only *derivatively*. In fact, they are true, or false, only because they have certain *semantic properties*, namely only because, relative to a context, they *express* a proposition. Moreover, they are true, or false, only insofar as they express true, or false, propositions. It is propositions, then, that are the *primary* truth(-value) bearers. A nice dividend of this way of putting things, which is ubiquitous in the literature,[10] is a useful division of labor. A *theory of meaning* will account for the linguistic expressions' semantic properties, telling us which proposition a sentence expresses relative to a context. In contrast, a *theory of truth* will tell us what it is for a proposition to be true. By combining the two theories, the propositionalists can conclude, we get a theory of truth for sentences, thus vindicating, so to speak, Austin's and Künne's simple observations.

Everything all right, then? Not exactly. This line of reasoning could work, perhaps, if it were true that an account of the linguistic expressions' semantic properties will lead us to identify certain abstract *objects* – the propositions – as what is expressed by sentences relative to contexts. Only in this case, in fact, would we really have some candidates other than sentences for the role of truth(-value) bearers. But that an account of the linguistic expressions' semantic properties will lead us to this, is, in my opinion, quite dubious, and certainly cannot be taken for granted.

As a matter of fact, I find the insistent appeal to propositions that is fashionable nowadays among analytic philosophers extremely perplexing. Not, I hasten to add, because I endorse a more or less Quinean form of semantic skepticism. On the contrary, I take it for granted that linguistic expressions have semantic properties, and that their semantic properties induce a partition of sentences, or of sentence/context pairs,

---

[10] This is, for example, Scott Soames' version: "In addition to propositions, utterances, eternal sentences, and occasion sentences taken in contexts (sentence/context pairs) can all be construed as truth bearers. However, the truth of a sentence or utterance depends on the truth of the proposition it expresses. A sentence or utterance cannot be true if it says nothing or expresses no proposition. Rather, it is true because it expresses a true proposition" (1999: 18). And this is David Lewis': "I take our topic to be, in the first instance, the truth of propositions. Sentences, or sentences in context, or particular assertions of sentences, or thoughts, can derivatively be called true; but only when they succeed in expressing determinate (or near enough determinate) propositions" (Lewis 2001: 276).

into equivalence classes.[11] Perhaps, to simplify matters, one may even go so far as to call these classes "propositions".[12] But if this were all that propositions boil down to, then obviously their elements rather than they themselves ought to be conceived as the primary truth(-value) bearers (as the quotation from John Perry in the last footnote suggests).

Actually, there are two kinds of arguments that are appealed to by those who claim that propositions are something more robust than suggested above. The most common one consists in individuating some theoretical roles (for example, being the semantic value of a sentence relative to a context, being the object, or the content, of a propositional attitude, being the referent of a that-clause, and, of course, being the primary truth(-value) bearer) and calling the entities that play these roles, whatever they are, "propositions". Such an argument has always seemed to me quite weak. Although this is not the place to go into this, the theoretical roles that are usually called upon seem to me either ill-defined – are we sure that the that-clauses are referential expressions?[13] – or incompatible with each other – no kind of entity can play all of them. A perhaps less common, but in my opinion more interesting, argument consists in highlighting how pervasive our pre-philosophical commitment to propositions is. As Künne put it, "[t]hose who are keen to ban talk of propositions often seem not to realize how many general terms which are common coin in non-philosophical discourse do 'specialized' duty for

---

[11] Interestingly, this way of seeing things is adumbrated by Quine himself. In 1948 he wrote: "The useful ways in which people ordinarily talk or seem to talk about meanings boil down to two. The *having* of meanings, which is significance, and *sameness* of meaning, or synonymy . . . . The problem of explaining these adjectives 'significant' and 'synonymous' with some degree of clarity and rigor – preferably, as I see it, in terms of behavior – is as difficult as it is important. But the explanatory value of special and irreducible intermediary entities called meanings is surely illusory" (1948: 11–12). With a more skeptical ring, the point is repeated much later: "The doctrine of propositions seems in a way futile on the face of it, even if we imagine the individuation problem solved. For, that solution would consist in some suitable definition of equivalence of sentences; why not then just talk of sentences and equivalence and let the propositions go?" (1970: 10).

[12] More or less in this spirit, I believe, John Perry has contended that propositions are "abstract objects that we use to classify states and events by the requirements their truth (or some other form of success) impose on the rest of the world"; as such, they "are a bit analogous to weights and lengths" (2001: 20–21).

[13] For my negative answer to this question, see Bianchi 2010: 67–69.

'proposition'" (2003: 252). Künne's list includes verbal nouns such as "allegation", "belief", "conjecture", "contention", "judgement", "report", "statement", "supposition", and "thought", all of which "have readings under which they are used to refer to propositions", as well as non-verbal nouns such as "axiom", "dogma", "tenet", "theorem", and "thesis" (pp. 249–252). The point is certainly well taken: what else could phrases such as "Goldbach's conjecture" or "Pythagoras' theorem" single out, if not abstract entities that may be expressed by different sentences belonging to different languages? I must admit that I do not have an answer, but in any case this does not seem to me to be a sufficient reason for considering propositions as the primary truth(-value) bearers (cf. Bianchi 2010: 70).

This said, I know by experience that it is very difficult to induce a propositionalist to doubt the existence of propositions, and I shall not even try here. I would like, however, to suggest the adoption of a different starting point for the elaboration of a theory of truth, a starting point nearer, so to speak, to Austin's and Künne's simple observations. Since, as we saw, the propositionalists, too, are ready to acknowledge that certain linguistic expressions, sentences, can be said to be true, or false, as well, why not attempt to develop a theory that takes sentences themselves as truth(-value) bearers? If this will then force us to appeal somehow to propositions, so much the worse (and so much the better for the propositionalist).

There are at least three types of objections that can be moved against such an approach. The first is tied to considerations we have already encountered. Sentences, we saw, can be said to be true, or false, only in virtue of the fact that they have semantic properties. But then, it could be claimed, it is not sentences, but their semantic properties (or better, for the propositionalist, the propositions that the latter determine relative to a context) that are properly true, or false. This is, however, clearly a *non sequitur*.[14] Compare: Ernesto is a voter in Italy only in virtue of the fact that he is an Italian citizen, that is to say, for an

---

[14] Alas, Ernesto as well seems to have fallen prey to it. He wrote: "If sentences are true or false in virtue of what they say, it does not take much to think that it is not the sentence but what the sentence says to be true or false; if you prefer, it does not take much to think that the sentence's truth or falsity is none other than the truth or the falsity of what the sentence says" (Napoli 2010: 295, translation mine).

objectifier, only in virtue of the fact that he possesses the thing that we call "Italian citizenship". Nonetheless, the voter is Ernesto, not his civil properties (his being an Italian citizen), or, worse, that strange abstract object his Italian citizenship would be.

A second objection that is very often raised (see for example the passages quoted in footnote 10) against the idea that sentences are the primary truth(-value) bearers is that sentences are not true, or false, *absolutely*, but only *relatively*. Indeed, they are true, or false, only relative to a language, and/or an interpretation, and/or a context, and/or what have you. There are, however, at least two replies that can be given to this objection. On the one hand, that something has a property only relative to something else implies neither that it does not have the property nor that something else (say, an abstract object) has that same property absolutely rather than relatively (as is the case, according to the propositionalist, with propositions in relation to truth). On the other hand, and this is what I want to stress here, it is not at all clear that the linguistic expressions of which we can predicate truth, or falsity, are true, or false, only relatively. In fact, the objection depends on considering linguistic expressions in abstract, as *types*, that can be used in various contexts and with varying interpretations and whose tokens can even belong to different languages. But it is not mandatory for someone aiming at developing Austin's and Künne's simple observations into a theory of truth to look at linguistic expressions from this perspective. On the contrary, what I would like to suggest is that we should develop a theory of truth that takes as truth(-value) bearers a subclass – the subclass of *sentence tokens* – of what I like to call "linguistic particulars" (see Bianchi 2015): physical entities of some type (usually, sounds or marks) that are produced by someone at some particular time, usually to communicate something to someone else. Similarly, Austin takes as the truth(-value) bearer the *statement*, namely "the words or sentence *as used by a certain person on a certain occasion*" (1950: 119). Indeed, he takes care to specify that the making of a statement is "an historic event, the utterance by a certain speaker or writer of certain words (a sentence) to an audience with reference to an historic situation, event or what not" (119–120).[15] It seems to

---

[15] There is actually some ambiguity in Austin's way of phrasing his point, so that it is not completely clear whether he takes statements to be *acts* or the 'products' of acts of a certain kind. By "linguistic particulars" I mean the latter.

me that to consider those linguistic particulars that are sentence tokens, or what Austin calls "statements", as the (primary) truth(-value) bearers is quite reasonable. On the one side, it is commonsensical (think once again of Austin's and Künne's simple observations);[16] on the other, it does not appeal to entities that are ontologically dubious. And, to come back to the objection we were discussing, we can say of a sentence token, or of a statement in Austin's sense, that it is true, or false, absolutely, not only relative to something else (a language, an interpretation, a context, or what have you), although it is still true that it is true, or false, only in virtue of its semantic properties.

But haven't I been too hasty in arriving at this conclusion? Künne gives us some grounds to think so:

> Suppose a speaker is talking on the phone to his worst enemy while looking at his best friend: in a single utterance of 'You are my best friend' he might address both persons simultaneously and thus express two propositions . . . with different truth-values. Surely confusion would result if we were to call the utterance (or the token) true and not true. Or suppose you utter a grammatically and/or lexically ambiguous sentence, intending your utterance to be understood both ways. (Perhaps you are making a joke, and the point of the joke depends on the sentence being given both readings by the person you are addressing.) Then it may very well be the case that you express a truth and a falsehood at one stroke. Again, confusion would result if we were to call the utterance (or the token-sentence) true and not true. (Künne 2003: 266–267)

> Let me round this off by adding an interlingual example . . . . Annabella, a business woman in Milan, has two telephones on her desk . . . . An American colleague and a British friend rang her simultaneously wanting to know how much

---

[16] Or at least *I* find it commonsensical. Horwich and Künne think otherwise. This is Horwich: "I shall follow ordinary language in supposing that truth is a property of propositions. Thus, if we agree with Oscar, we attribute truth to *what he said*, to the proposition he asserted. Evidently the *sentence-type of English* that he used is not true; for that very sentence-type is used on other occasions to make false statements. Nor would one normally characterize the noises he made, or his belief state, as true" (1998: 16). This is Künne: "In our everyday employment of 'true', we normally, if not exclusively, take propositions to be the things that are susceptible of truth . . . . An account that aims to be faithful to our workaday concept of truth cannot afford to turn its back on propositions: they are the *primary truth-value bearers* . . . . But don't we ascribe truth (without relativization) to *utterances* in our daily transactions? I don't think we do" (2003: 263–264). However, as a teacher I know by experience how difficult it is for a (philosophy!) student to even entertain the thought that there are abstract entities that are the content of our mental acts and states, what is expressed by sentences in context, and what truth and falsity are properties of.

profit her firm made last year. She wanted only her friend to know the truth. So picking up both receivers she said, 'One billion lira. But excuse me, I have a visitor in my office. Let's talk tomorrow.' And then she hung up. Annabella intended her American colleague to understand that the profit amounted to 10^9 lira, and her British friend to understand that it amounted to 10^{12} lira. A falsehood as well as a truth were conveyed by just one utterance . . . . Therefore, if utterances . . . were themselves truth-value bearers, some bearers would be both true and not true. All this is certainly unbearable. (268)

What these, admittedly ingenious but undoubtedly far-fetched, cases might suggest is that sentence tokens are true (at most) only relative to a context ("You are my best friend"), an interpretation (the ambiguous sentence), or a language (the "one billion lira" example). However, it seems to me that there are good reasons to resist this conclusion, which depends on a bad understanding of what a sentence token is.[17] Linguistic particulars are not *mere* sounds or marks, to which an interpretation needs to be somehow attached. On the contrary, they have their semantic properties absolutely, so to speak: their *origin* – the history of their production – makes them have them (see Bianchi 2015: 100–103 for a development of this point in relation to referring linguistic particulars). From an epistemological point of view their semantic properties may sometimes be difficult to discern, and this is a fact that the speakers exploit in Künne's cases, but they are metaphysically determinate nonetheless. Consider the "one billion lira" case, for example. Either on that occasion Annabella is speaking British English, in which case by producing a true sentence token she probably succeeds in leading her American colleague to acquire a false belief, exploiting the fact that he will probably take her to be speaking American English and as a consequence misunderstand the token; or she is speaking American English, in which case by producing a false sentence token she probably succeeds in leading her British friend to acquire a true belief, exploiting the fact that he will probably take her to be speaking British English and as a consequence misunderstand the token. And which language Annabella is speaking is perfectly determinate, by her cognitive history, although in this case irrelevant from a practical point of view, since no matter which language she is speaking she will probably achieve what she wants

---

[17] Just for the record, let me register here that Kirkham (1992: 67–69) discusses an interlingual case similar to Künne's, but concludes that it does not *force* us to accept that sentence tokens are true only relative to a language.

to achieve, having "only her friend ... know the truth" (but notice that only in the second case could Annabella be charged with *lying* to her American colleague). In my opinion, Künne's other cases can be dealt with in a similar way, *mutatis mutandis*.[18]

There remains to be considered a third objection to the claim that sentences (or, better, given what we have just said, sentence tokens) are the primary truth(-value) bearers.[19] In fact, isn't the claim too reductive? Even Künne's truism may suggest that this is so, insofar as in it a truth-value is ascribed not only to what one says, but also to *what one thinks*. And in our discussion of the "one billion lira" case we talked of acquiring a *true*, or a *false*, *belief*. Now, to say something we do need words, but at least *prima facie* we do not need them to think or believe something. If this were so, a theory of truth taking sentence tokens as the truth(-value) bearers would not be able to account for certain significant truth-ascriptions we are all used to making. Consider, however, that we are all trained in the linguistic game of ascribing truth and falsity in relation to certain simple linguistic particulars (what is uttered by mom, or suchlike). If we come to extend our ascriptions to certain mental acts and states as well, we probably do so because they are relevantly similar to what we have been trained to ascribe truth or falsity to. This might seem to be grist for the propositionalists' mill, since they may claim that the mental acts and states at issue are similar to utterances precisely in that like them they have a proposition as their *content*. It is also because of this, they might conclude, that we need to develop a theory of truth for propositions rather than for sentence tokens. But it can also be the case that the points of similarity are different. For example, it is possible

---

[18] As Giuseppe Spolaore has pointed out to me, however, even if, in contrast to what I have said, it were maintained that sentence tokens such as those involved in these cases are actually *equivocal*, no interesting general conclusion would follow. In fact, it would still be the case that most of the sentence tokens that we produce are not equivocal. Hence, the resulting relativity would not pose insurmountable difficulties to the claim that truth is a property primarily of sentence tokens.

[19] Actually, there is a further objection that is sometimes raised against the claim. It is difficult, it is contended, to account for logical truth and logical consequence if we take sentence tokens as the pertinent truth(-value) bearers. See Kaplan 1989: 522, 546; Kaplan 1989a: 585 n. 40, 586–587; Künne 2003: 265–266. I cannot deal with this here, as the discussion would take us too far afield. As should be clear, my focus in this paper is plain truth, not logical truth.

that the mental acts and states to which we ascribe truth or falsity involve *vehicles* that are sufficiently similar to linguistic expressions, as the language of thought hypothesis may suggest.[20] It is also to shed light on these issues, then, that it seems useful to me to try and develop a theory of truth for sentence tokens. If we understand what it is for a sentence token to be true, it is possible that we gain a better understanding of why we ascribe truth or falsity to certain mental acts and states as well.

As for the issue of the truth(-value) bearers, I shall stop here. Now, the really difficult part. What is it for a sentence token to be true? That is, what should we put in place of "*P*" in our theory of truth? Let's go back once again to the two simple observations that we have chosen to use as our starting point. Austin tells us that "truth is a matter of the relation between words and world", while Künne appeals to how things are. But how can we develop this generic mention of the *world* or of the *way in which things are* into a theory without 'fabricating' strange entities – *facts* – that would correspond to those sentence tokens that are true? Künne gives us a little help when he writes that "what you say or think is true if and only if *things are as you say or think they are*". When by producing a linguistic particular – a sentence token – we make a statement, to use Austin's terminology, we *say that things are in a certain way*. And this is so because the linguistic particular we produce *represents* things as being that way. What's more, in order to say that things are in a certain way, we have to *talk about* certain things. There is no *saying that* without *talking about*. And we cannot *talk about* without using linguistic particulars that *refer* to something. Briefly, we can say that things are in a certain way only because we can combine in a certain way linguistic particulars that refer to certain things, so as to obtain linguistic particulars that represent things as being that way. (Moreover, if my speculations in the last paragraph are on the right track, the same holds for thinking: we can think that things are in a certain way only because we can combine in a certain way (quasi-)linguistic particulars that refer to certain things, so as to obtain (quasi-)linguistic particulars that represent things as being that way). Here, then, emerges the crucial link between truth and reference, which has more than once

---

[20] On the language of thought hypothesis, see Fodor 1975 and Field 1978. In Bianchi 2005 and 2007 I have argued for the claim that we think by means of the language that we speak. This, obviously, would make the similarity even more marked.

been highlighted in the course of the twentieth century (by Ernesto as well, as I said) but which those who take truth to be a property of propositions tend to lose sight of.[21] In a nutshell, the truth-value of a sentence token depends on the way things are with regard to those entities that certain linguistic particulars by which it is constituted refer to.

Let's consider an example to clarify things. If I now say that Ernesto is insightful, I produce a structured linguistic particular – a sentence token. The sentence token that I produce is true, because it is constituted by a linguistic particular (a token of the name "Ernesto") that refers to a specific individual and by a linguistic particular (a token of the adjective "insightful") that refers to a specific property, and the individual that the first particular refers to, Ernesto, has the property that the second particular refers to, insightfulness.[22] In this way, we have explained the truth of a certain structured linguistic particular – a sentence token – in terms of the *reference* of the linguistic particulars by which it is constituted and the *way in which things are.* Moreover, we did this, so it seems, in line with Austin's simple observation that "truth is a matter of the relation between words and world" and, in the end, with Künne's that "what you say ... is true if and only if *things are as you say ... they are*" as well. And, last but not least, we did so without postulating the existence of ontologically dubious entities such as propositions or facts.

Was it all so easy? Not at all, of course. We neglected a fundamental aspect: the truth of a sentence token depends not only on the reference of the linguistic particulars that constitute it and the way in which things are, but also on its structure, or *form.* If all true sentence tokens had the same form as the one that I produced when I said that Ernesto is insightful, it would not be difficult to offer a satisfactory theory of truth. Simplifying a bit, here it is: $\forall x(x$ is true $\leftrightarrow$ the individual that x's first constituent refers to has the property that x's second constituent refers to). By such a theory, we could for example account for the truth of the

---

[21] This was my second perplexity concerning Künne's Modest Account (see footnote 4).

[22] That adjectives (as well as verbs and common nouns) are referential expressions and that they refer to entities other than individuals are basic tenets of Ernesto's philosophy of language. See for example Napoli 1995: 329; Leonardi and Napoli 1995: 263–264; Bianchi and Napoli 2004: 223–225.

linguistic particular that I produce when I say that Paolo is generous, of the one that I produce when I say that Diego is austere, and of the one that I produce when I say that I am male. But obviously, as soon as we consider the linguistic particular that I produce when I say that Ernesto is younger than Paolo, we are in trouble. Actually, it would not be difficult to extend our theory so as to cover linguistic particulars having the same form as this as well. But, again, it would not be difficult to then find linguistic particulars with a different form that put us in trouble. And so on and so forth.[23]

In the first half of the twentieth century, Ludwig Wittgenstein's *Tractatus* on the one hand and Tarski's meta-mathematic work on the other highlighted the role of form in accounting for truth.[24] However, neither of their proposals seems to me satisfactory, for different reasons. Wittgenstein demanded too much of form, so to speak. In his picture theory, he postulated the existence of non-linguistic entities (facts!) *isomorphic* to certain linguistic expressions (his "elementary propositions") to account for the truth of the latter. Tarski, instead, in his truth definitions focused on linguistic forms (although only in relation to certain very simple formal languages) but failed to take reference seriously, as Hartry Field has convincingly shown in his 1972 article.

It is this dependence of truth on form that makes the development of a theory of truth a peculiarly complex, if not impossible, task: unfortunately, there is no compact way to characterize the dependence of truth on reference, form and the way in which things are. There are no shortcuts (except for *recursion*): as Tarski realized, we need to produce a *complete* catalogue of possible sentential forms, and then to explain how the truth of the linguistic particulars having each of these forms is determined by the reference of their constituents and the way in which things are. Sad to say, until we have this catalogue and all the relative explanations, we shall not have a theory of truth. In the meantime, we

---

[23] This is basically a version of what Künne (2003: 111) has called "the Procrustes Problem".

[24] It can be argued that they were somehow anticipated by Aristotle, who famously claimed that "to say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true". According to Davidson, in fact, "Aristotle's characterization ... makes clear ... that the truth of a sentence depends on the inner structure of the sentence, i.e., on the semantic features of the parts" (1997: 23).

have to content ourselves with some fragments, relative to linguistic particulars having simple sentential forms, which will at least give us an idea of the way to go, and with a couple of simple observations (which, unfortunately, are ways too often blatantly ignored).

References

Austin, J.L. (1950), "Truth", *Aristotelian Society Supplementary Volumes*, 24, pp. 111–129, reprinted in his *Philosophical Papers*, edited by J.O. Urmson and G.J. Warnock, Clarendon Press, Oxford 1979, 3rd edition, pp. 117–133.

Bianchi, A. (2005), "Words as concepts", in J. Acero and P. Leonardi (eds.), *Facets of Concepts*, Il Poligrafo, Padova, pp. 83–108.

——— (2007), "Speaking and thinking (Or: A more Kaplanian way to a unified account of language and thought)", in M. Beaney, C. Penco and M. Vignolo (eds.), *Explaining the Mental: Naturalist and Non-Naturalist Approaches to Mental Acts and Processes*, Cambridge Scholars Publishing, Newcastle, pp. 13–32.

——— (2010), "I limiti della modestia. Alcune riflessioni su *Conceptions of Truth* di Wolfgang Künne", in M. Carrara and V. Morato (eds.), *Verità. Annuario e Bollettino della Società Italiana di Filosofia Analitica*, Mimesis, Milano, pp. 65–74.

——— (2011), "What do philosophers do? A few reflections on Timothy Williamson's *The Philosophy of Philosophy*", in R. Davies (ed.), *Analisi. Annuario della Società Italiana di Filosofia Analitica*, Mimesis, Milano, pp. 117–125.

——— (2015), "Repetition and reference", in A. Bianchi (ed.), *On Reference*, Oxford University Press, Oxford, pp. 93–107.

Bianchi, A. and Napoli, E. (2004), "Poche parole", in M. Carrara and P. Giaretta (eds.), *Filosofia e logica*, Rubbettino, Soveria Mannelli, pp. 175–225.

Burgess, A.G. and Burgess, J.P. (2011), *Truth*, Princeton University Press, Princeton.

David, M. (2015), "The correspondence theory of truth", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, fall 2015 edition, `http://plato.stanford.edu/archives/fall2015/entries/truth-correspondence/`.

Davidson, D. (1997), "The folly of trying to define truth", *Journal of Philosophy*, 94, pp. 263–278, reprinted in his *Truth, Language, and History*, Clarendon Press, Oxford 2005, pp. 19–37.

——— (2000), "Truth rehabilitated", in R.B. Brandom (ed.), *Rorty and His Critics*, Blackwell, Oxford, pp. 65–73, reprinted in D. Davidson, *Truth, Language, and History*, Clarendon Press, Oxford 2005, pp. 3–17.

Devitt, M. (1997), *Realism and Truth*, Princeton University Press, Princeton, 2nd edition with a new afterword.

——— (2001), "The metaphysics of truth", in M.P. Lynch (ed.), *The Nature of Truth*, MIT Press, Cambridge, MA, pp. 579–611, reprinted in M. Devitt, *Putting*

Metaphysics First: Essays on Metaphysics and Epistemology, Oxford University Press, Oxford 2010, pp. 155–181.

Field, H. (1972), "Tarski's theory of truth", *Journal of Philosophy*, 69, pp. 347–375, reprinted with a postscript in his *Truth and the Absence of Fact*, Clarendon Press, Oxford 2001, pp. 3–29.

——— (1978), "Mental representation", *Erkenntnis*, 13, pp. 9–61, reprinted with a postscript in his *Truth and the Absence of Fact*, Clarendon Press, Oxford 2001, pp. 30–82.

Fodor, J.A. (1975), *The Language of Thought*, Harvard University Press, Cambridge, MA.

Glanzberg, M. (2014), "Truth", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, fall 2014 edition, `http://plato.stanford.edu/archives/fall2014/entries/truth/`.

Horwich, P. (1998), *Truth*, Clarendon Press, Oxford, 2nd edition.

Kaplan, D. (1989), "Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals", in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford University Press, Oxford, pp. 481–563.

——— (1989a), "Afterthoughts", in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford University Press, Oxford, pp. 565–614.

Kirkham, R.L. (1992), *Theories of Truth: A Critical Introduction*, MIT press, Cambridge, MA.

Künne, W. (2003), *Conceptions of Truth*, Clarendon Press, Oxford.

——— (2010), "Risposte a due critici", in M. Carrara and V. Morato (eds.), *Verità. Annuario e Bollettino della Società Italiana di Filosofia Analitica*, Mimesis, Milano, pp. 89–109.

Leonardi, P. and Napoli, E. (1995), "On naming", in P. Leonardi and M. Santambrogio (eds.), *On Quine*, Cambridge University Press, Cambridge, pp. 251–266.

Lewis, D. (2001), "Forget about 'the correspondence theory of truth'", *Analysis*, 61, pp. 275–280.

Lynch, M.P. (2009), *Truth as One and Many*, Clarendon Press, Oxford.

——— (2014), "What ever happened to the correspondence theory of truth?", in J. Dutant, D. Fassio and J. Meylan (eds.), *Liber Amicorum Pascal Engel*, Université de Gèneve, Gèneve, pp. 3–14.

Napoli, E. (1995), "(Direct) reference", *Journal of Philosophical Logic*, 24, pp. 321–339.

——— (2010), "Realtà, verità e linguaggio", in M. Carrara and V. Morato (eds.), *Verità. Annuario e Bollettino della Società Italiana di Filosofia Analitica*, Mimesis, Milano, pp. 295–302.

Perry, J. (2001), *Reference and Reflexivity*, CSLI, Stanford.

Putnam, H. (2015), "Naturalism, realism, and normativity", *Journal of the American Philosophical Association*, 1, pp. 312–328.

Quine, W.V. (1948), "On what there is", *Review of Metaphysics*, 2, pp. 21–38, reprinted in his *From a Logical Point of View*, Harvard University Press, Cambridge, MA, 1953 (2nd revised edition 1961, pp. 1–19).

——— (1970), *Philosophy of Logic*, Prentice Hall, Englewood Cliffs (2nd edition Harvard University Press, Cambridge, MA, 1986).

——— (1987), *Quiddities: An Intermittently Philosophical Dictionary*, Belknap Press, Cambridge, MA.

Soames, S. (1999), *Understanding Truth*, Oxford University Press, New York.

Strawson, P.F. (1949), "Truth", *Analysis*, 9, pp. 83–97.

——— (1950), "Truth", *Aristotelian Society Supplementary Volumes*, 24, pp. 129–156, reprinted in his *Logico-Linguistic Papers*, Ashgate, Aldershot 2004, 2nd edition, pp. 147–164.

——— (1964), "A problem about truth – A reply to Mr. Warnock", in G. Pitcher (ed.), *Truth*, Prentice Hall, Englewood Cliffs, reprinted in P.F. Strawson, *Logico-Linguistic Papers*, Ashgate, Aldershot 2004, 2nd edition, pp. 165–179.

Tarski, A. (1933), "Pojęcie prawdy w językach nauk dedukcyjnych", Warsaw, translated as "The concept of truth in formalized languages" in his *Logic, Semantics, Meta-mathematics*, edited by J. Corcoran, Hackett, Indianapolis 1983, 2nd edition, pp. 152–278.

Volpe, G. (2012), *La verità*, Carocci, Roma.

Napoli sulla dipendenza della verità dal linguaggio

Diego Marconi

In "Realtà, verità e linguaggio" (2010), Ernesto Napoli rivolge alcune critiche a un mio argomento antiheideggeriano (presentato in Marconi 2006). Nel far questo, Napoli sostiene tesi di portata molto generale e meritevoli di discussione indipendentemente dalla loro applicazione al mio argomento. Come molti suoi scritti, anche questo è profondo, acuto, intensamente polemico e abbastanza intricato. Molto probabilmente, Napoli sosterrà che, come sempre, non ne ho capito nulla e che di conseguenza le mie critiche mancano il bersaglio di qualche chilometro. Tuttavia ci proverò lo stesso, perché evitare di discutere con Ernesto è quasi impossibile. Il suo stile filosofico, perentorio e al tempo stesso incline a mettere in dubbio quasi ogni (apparente) ovvietà,[1] non sollecita la discussione: la impone.

1. L'argomento antiheideggeriano era il seguente. In *Essere e tempo*, Heidegger asserisce:

> Prima che le leggi di Newton fossero svelate, non erano "vere"; non ne consegue che fossero false e neppure che esse, se non se ne rendesse più possibile alcuno stato di svelamento ontico, diventerebbero false. [...] Che le leggi di Newton non fossero, prima di lui, né vere né false, non può significare che l'ente da esse svelato e mostrato prima non sia stato. Quelle leggi divennero vere grazie a Newton, grazie a esse l'ente si rese in sé accessibile all'esserci. Una volta svelato, l'ente si mostra proprio come quell'ente che già prima era. (1927: 643)

---

[1] Nella Prefazione alla sua *Filosofia del linguaggio*, Paolo Casalegno ringrazia Ernesto Napoli, "per il quale nulla è ovvio e che spesso vede ciò che agli altri sfugge" (Casalegno 1997: 10).

Da questa presa di posizione sembra lecito evincere che, ad esempio, Heidegger avrebbe accettato entrambe le tesi (1) e (2):

(1)   Prima di Newton, le orbite dei pianeti erano ellittiche.
(2)   Prima di Newton, non era vero che le orbite dei pianeti erano ellittiche.

Quindi, secondo lui, ci sono casi in cui vale che Pp ma non vale che PTp (dove p è una proposizione, "T" è l'operatore di verità ("è vero che") e "P" è l'operatore di passato ("si è dato il caso che"; "PTp" = "si è dato il caso che è vero che p" = "era vero che p")). Ma i contesti della forma "P[...]" sono intensionali: in questi contesti è possibile sostituire proposizioni con proposizioni necessariamente equivalenti conservando l'intensione. Dunque la posizione di Heidegger implica o la non equivalenza necessaria di "p" e "Tp" (che è solitamente ammessa), o che i contesti della forma "P[...]" *non* siano intensionali. In quest'ultimo caso si avrebbe che, anche se "p se e solo se Tp" è una verità necessaria (per qualsiasi p), non per questo per qualsiasi p è una verità necessaria che "Pp se e solo se PTp": per qualche p, diciamo q, è possibile che – ad esempio – si sia dato il caso che q ma non che Tq. In altre parole, ci sarebbero verità necessarie alcuni esempi delle quali non sono necessariamente verità eterne.

   Napoli contesta alcuni dettagli della mia argomentazione, ma non nega che le premesse, se fossero sensate, implicherebbero la conclusione. La sua critica è più radicale: secondo lui, enunciati del tipo di (2) non hanno senso, e quindi non possono fungere da premesse in nessuna argomentazione. L'esempio che Napoli discute (e che è usato anche da me) è (3):

(3)   Era vero 70 milioni di anni fa che c'erano dinosauri.

Secondo lui, (3) non segue da (4):

(4)   È vero che 70 milioni di anni fa c'erano dinosauri.

Infatti se (3) seguisse da (4), nella seguente argomentazione il passo da (6) a (7) sarebbe accettabile:

(5)   Nel 1943 Diego non era ancora nato
(6)   È vero che nel 1943 Diego non era ancora nato

?(7) Era vero nel 1943 che Diego non era ancora nato.

Ma, dice Napoli, il passo da (6) a (7) dev'essere fallace, perché "Diego nel 1943 non esisteva e quindi nel 1943 niente poteva esser vero di lui. Niente Diego, niente verità su di lui" (2010: 298). Ma il passo da (6) a (7) ha la stessa forma del passo da (4) a (3) – è basato sulla stessa equivalenza generale, che è palesemente invalida, perché consente di derivare l'insensato (7). Quindi (3) non segue da (4).

Qui, il punto teoricamente più 'pesante' è il principio "Niente Diego, niente verità su di lui" (che sostituirò con "Niente Ugo, niente verità su di lui" per evitare sospetti di narcisismo. Supporrò che anche Ugo, come Diego, sia nato nel 1947). Lo discuterò nella seconda parte di questo intervento. Per adesso supporrò che il principio sia vero: quali sarebbero le conseguenze per la mia argomentazione antiheideggeriana? Napoli dice che (3) non segue da (4) perché il passo è basato sulla stessa equivalenza generale che si è visto non esser valida: quella tra "è vero che a t p" e "è vero a t che p". Ma ciò che si è visto è, al massimo, che l'equivalenza non vale *in generale*; e se la ragione per cui il passo da (6) a (7) è invalido è che Ugo non esiste nel 1943, e quindi non ci sono verità su di lui nel 1943, questa non può essere la ragione per cui è invalido il passo da (4) a (3), dato che 70 milioni di anni fa i dinosauri esistevano, e quindi c'erano verità sui dinosauri; in particolare, *era vero* che esistevano dinosauri. Lo stesso nel caso delle orbite dei pianeti: prima di Newton le orbite dei pianeti c'erano, quindi c'erano verità su di esse: in particolare, *era vero* che erano ellittiche.

Dunque l'enunciato (3) non è insensato, o comunque non lo è in base al principio "Niente Ugo, niente verità su di lui". Ma, supponendo che (3) sia sensato, è derivabile da (4)? Ammettendo che in questo caso "p se e solo se Tp" sia vera e necessariamente tale, e supponendo che i contesti della forma "P[...]" siano intensionali, certamente sì.

Indubbiamente, in questo ragionamento si presuppone che formulazioni come "Era vero a t che p" o "Non era vero a t che p" abbiano senso, cioè che sia sensato dire che una proposizione è vera, o non lo è, in un certo momento del tempo, e quindi dire, di alcune proposizioni, che sono vere in *ogni* momento del tempo. Napoli non apprezza questa concezione (2010: 298), e tuttavia concede che "l'eternità della verità [possa essere concepita come] stabilità attraverso il tempo" (ib.), e perciò non tratta le formulazioni di cui sopra come semplicemente insensate. Anzi, ne pro-

pone lui stesso una analoga: "Era vero [sulla Terra] 70 milioni di anni
fa che c'erano dinosauri [sulla Terra] 70 milioni di anni fa". Ma – osser-
va – la formulazione che io uso nell'argomento antiheideggeriano non ha
questo formato.

Avrei potuto usare la pesante formulazione proposta da Napoli – cioè
una formulazione in cui sia il contenuto proposizionale, sia il predicato di
verità sono esplicitamente parametrizzati rispetto al luogo e al momento
del tempo – ottenendo lo stesso risultato? A me pare di sì. Le tesi di
Heidegger diventano

(1′) Le orbite dei pianeti sono ellittiche prima di Newton, nel sistema
      solare
(2′) Non è vero prima di Newton, nel sistema solare che le orbite dei
      pianeti sono ellittiche prima di Newton, nel sistema solare.

Il problema non cambia. Se prima di Newton nel sistema solare le orbite
dei pianeti sono ellittiche, allora è vero prima di Newton, nel sistema
solare, che prima di Newton nel sistema solare le orbite dei pianeti sono
ellittiche; e viceversa. Se piove a Torino il 15 agosto 2015, allora è vero a
Torino, il 15 agosto 2015, che piove (a Torino il 15 agosto 2015); e vice-
versa. Napoli potrebbe obiettare che l'equivalenza non vale in generale,
perché consente la derivazione

(5′) Nel 1943 a Torino Ugo non era ancora nato
(6′) È vero nel 1943 a Torino che nel 1943 a Torino Ugo non era ancora
      nato

che non è valida per il principio "Niente Ugo, niente verità su di lui" (che
rende insensato l'enunciato (6′)). Su questo, come ho detto, tornerò in
seguito. In ogni caso, per le orbite dei pianeti il problema non si pone:
l'equivalenza è vera. Ed è anche necessaria: non c'è un mondo in cui è
vero in un certo giorno che *quel giorno* piove, ma quel giorno non pio-
ve, o in cui piove un certo giorno, ma quel giorno non è vero che (quel
giorno) piove. Potrebbe essere così solo se l'espressione temporale ("il 15
agosto 2015") designasse giorni diversi in "Piove il 15 agosto 2015" e in
"È vero il 15 agosto 2015 che...", ma non c'è ragione alcuna di ammet-
terlo. Dunque, anche nella formulazione accolta da Napoli la posizione
di Heidegger contraddice un'equivalenza che c'è ragione di ritenere vera,
e necessariamente vera.

Si noti che la difficoltà *non* ha a che fare con l'alternativa tra permanentismo ("In ogni tempo, ogni cosa è sempre qualcosa") e temporaneismo ("Non in ogni tempo ogni cosa è sempre qualcosa") (Williamson 2013: 4). Se si è temporaneisti, si può pensare che ci sia stato un tempo in cui i pianeti, non esistendo, non erano nulla e quindi non avevano orbite ellittiche; a quel tempo non era vero che le orbite dei pianeti a quel tempo erano ellittiche. L'equivalenza non è in discussione, e nemmeno la sua necessità.

2. A questo punto, lasciamo Heidegger (appropriatamente) al suo destino e veniamo al punto centrale. Napoli sostiene che se qualcosa non esiste ancora (ad esempio perché non è ancora nato, come Ugo nel 1943), allora non ci sono verità su di lui. Per simmetria, dovrebbe essere altrettanto d'accordo che se qualcosa non esiste *più* (per esempio perché è morto), non ci sono verità su di esso. Eppure, sembrano esserci proposizioni vere (e proposizioni false) su cose che non esistono più: "Queen Anne is dead. My father became a grandfather only after his death" (Williamson 2013: 150). È l'ovvia verità di queste proposizioni ad aver convinto molti, tra cui Richard Montague, che "sarebbe eccessivamente restrittivo" richiedere che l'estensione di un predicato includa soltanto individui che esistono rispetto al punto di riferimento a cui è valutato l'enunciato in cui compare il predicato, o che una costante individuale abbia come denotazione un individuo esistente rispetto al punto di riferimento in questione. Il suo esempio – "C'è chi rimpiange l'ultimo re d'Inghilterra"[2] – tratta come punti di riferimento proprio i momenti del tempo (Montague 1970: 124). In altre parole, la scelta semantica di Montague è coerente con il permanentismo: ci sono proposizioni vere (oggi) che riguardano oggetti che (oggi) non esistono, almeno come oggetti concreti.

Del resto, anche Napoli tratta come sensati (e veri) enunciati su cose che non esistono più, per esempio i dinosauri:

> Sulla Terra 70 milioni di anni fa c'erano dinosauri. E questo è quanto. Se poi vogliamo essere enfatici possiamo anche dire: è vero che sulla Terra 70 milioni di anni fa c'erano dinosauri. (Napoli 2010: 298)

Eppure, sembrerebbe, niente dinosauri, niente verità sui dinosauri. I dinosauri non esistono (oggi) esattamente nel senso in cui Ugo non esisteva

---

[2] L'esempio di Montague è "C'è chi ricorda il precedente Papa", che le attuali circostanze storiche, del tutto eccezionali, rendono inutilizzabile.

nel 1943: non c'è nessun oggetto concreto che sia un dinosauro, così come nel 1943 non c'era nessun oggetto concreto che fosse Ugo; d'altra parte, esistono – presumibilmente – gli atomi di cui era composto ciascun dinosauro, così come nel 1943 esistevano gli atomi di cui sarebbe stato composto Ugo; i dinosauri sono possibili, tant'è vero che sono esistiti, così come nel 1943 Ugo era possibile, infatti qualche anno dopo è esistito. E così via. Perché, allora, Napoli accetta che ci siano oggi verità sui dinosauri, mentre nega che ci fossero verità su Ugo nel 1943?[3] Evidentemente la sua posizione non deriva dalla scelta del temporaneismo contro il permanentismo: il temporaneista, infatti, non fa differenza tra oggetti che non esistono ancora (come Ugo nel 1943) e oggetti che non esistono più (come i dinosauri o la regina Anna). Ciò che conta, per il temporaneista, è che né i dinosauri né Ugo esistono a t (rispettivamente, oggi e il 1943), e quindi a t non ci sono proposizioni su di loro, né vere né false.

Il fatto è che per il temporaneista, ma non per Napoli, la verità (e falsità) di una proposizione a t è indipendente dal fatto che quella proposizione sia espressa o esprimibile nel linguaggio a t; in effetti, è indipendente dal fatto che a t *esistano* linguaggi. Napoli la pensa in modo radicalmente diverso:

> Quello che un enunciato dice [= la proposizione espressa] non dipende per la sua esistenza dall'esistenza di questo piuttosto che quell'enunciato che lo dica. Tuttavia [...] l'esistenza di quello che un enunciato dice dipende dall'esistenza di un qualche enunciato che lo dica. Ossia, non solo un enunciato, ma anche quello che un enunciato dice è linguaggio dipendente. (Napoli 2010: 296)

Napoli conclude che, essendo la verità e la falsità proprietà di proposizioni, anche verità e falsità dipendono dal linguaggio (ib.).

Su queste premesse, il trattamento asimmetrico del caso di Ugo e di quello della regina Anna può essere giustificato nel modo seguente. Per un certo intervallo di tempo nel passato la regina Anna è esistita, e quindi è stato possibile riferirsi a lei come "Anna". Da allora è stato possibile,

---

[3] Nel suo (2011) Napoli, mentre sostiene – contro Wittgenstein – che è del tutto corretto dire che, quando N muore, muore il referente (quindi il significato) di "N", sottolinea peraltro che così come un libro conserva il suo autore anche dopo la morte dell'autore, allo stesso modo anche dopo la morte di N il suo nome, "N", conserva il referente che aveva, cioè continua a riferirsi a N (p. 194). Di conseguenza ci saranno anche allora proposizioni (vere o false) su N.

e lo è tuttora, esprimere in varie lingue proposizioni, vere o false, che vertono sulla regina Anna. Invece nel 1943 Ugo non esisteva, quindi non era possibile riferirsi a lui con "Ugo", o con un altro nome, in nessuna lingua. Non esistendo enunciati che parlassero di Ugo, non esistevano nemmeno proposizioni su Ugo, vere o false. Il fatto che *oggi*, e da un po' di anni, sia invece possibile parlare di Ugo ed esistano proposizioni vere che vertono su di lui (ad esempio, la proposizione che nel 1943 Ugo non era ancora nato) non autorizza a retrodatare la verità di queste proposizioni a prima che fosse possibile parlare di Ugo (per esempio al 1943), perché *allora* quelle proposizioni non esistevano e perciò non potevano essere vere. Dunque, il caso della regina Anna e dei dinosauri è ben diverso da quello di Ugo: ci sono oggi proposizioni vere (e false) sui dinosauri, anche se i dinosauri non esistono, perché ci sono enunciati sui dinosauri; mentre non c'erano nel 1943 proposizioni vere (né false) su Ugo perché non c'erano, in nessuna lingua, enunciati su Ugo.

C'è un dibattito, ormai tradizionale, sul rapporto tra proposizioni e linguaggio, imperniato sulla possibilità di pensieri non linguistici. Le proposizioni, oltre a essere espresse da enunciati, sono pensieri o contenuti di pensieri: dato che esistono sia pensieri non linguistici, sia esseri capaci *soltanto* di pensieri non linguistici (come certi animali superiori e i bambini prelinguistici), ci sono proposizioni – contenuti di pensieri – che non sono espresse da nessun enunciato. Non entrerò nel merito di questo dibattito, che è tuttora in corso, perché non mi sembra decisivo rispetto a ciò che è in gioco nella discussione di cui ci stiamo occupando.[4] Come vedremo fra poco, il punto non è se ci sono proposizioni che non sono di fatto espresse linguisticamente (sembra ovvio che sia così, se le proposizioni possono essere soltanto pensate), ma se ci sono proposizioni che non sono *esprimibili* linguisticamente (a un dato momento t).

Non intendo dar peso nemmeno a un'altra considerazione, peraltro ben nota: ad alcuni è parso che sia possibile esprimere, nel momento t, proposizioni che vertono su qualcosa che non esiste a t ed esisterà solo in un momento successivo t'. Per esempio, secondo Nathan Salmon e Francesco Berto l'esempio di Newman 1 proposto da David Kaplan dev'essere interpretato così. Ricordo che, in "Quantifying In", Kaplan aveva immaginato che qualcuno dichiarasse di battezzare "Newman 1" il

---

[4] McGrath (2014) dà conto di molti aspetti del dibattito sullo status metafisico delle proposizioni e contiene un'utilissima bibliografia.

primo bambino nato nel XXII secolo. Contestualmente, Kaplan esprimeva scetticismo nei confronti di qualsiasi teoria dei nomi che contemplasse la possibilità di un tale battesimo: "Non sono propenso ad adottare una teoria dei nomi propri che mi permetta di eseguire un battesimo (*dubbing*) *in absentia*" (Kaplan 1968: 200–201).[5] Al contrario, secondo Salmon chiunque risulterà essere il primo bambino nato nel XXII secolo *è già* il referente di "Newman 1": la non esistenza attuale di un individuo siffatto "non preclude questo fatto, non più di quanto la non esistenza attuale di Socrate precluda il fatto che 'Socrate' si riferisca a lui" (1998: 65). Berto, che cita esplicitamente Salmon, è sulla stessa posizione e immagina altri esempi di nomi di oggetti al momento inesistenti ma univocamente determinati sia da una descrizione, sia da connessioni causali: come "Giorgina", che designa la libreria IKEA che costruirò a partire dai pezzi staccati che in questo momento ho davanti (Berto 2013: 211). Se Salmon e Berto hanno ragione, non è vero in generale che finché un individuo non esiste non ci sono proposizioni vere o false su di lui perché non è possibile riferirsi a quell'individuo. Ma, come si è visto, il punto è controverso e non voglio investirci più di tanto.[6]

Le considerazioni principali che militano contro la posizione di Napoli sono altre. Una, molto generale, ha a che fare con la preferibilità del permanentismo rispetto al temporaneismo. L'altra ha a che fare con la difficoltà di concepire in modo convincente la dipendenza delle proposizioni dagli enunciati.

Cominciamo dalla seconda. La dipendenza ontologica delle proposizioni dagli enunciati che le esprimono può essere concepita in due modi, forte e debole:

(F)   A ogni t, la proposizione p esiste a t sse esiste a t un enunciato S di un linguaggio L tale che S esprime p.

(D)   A ogni t, la proposizione p esiste a t sse per qualche t' esiste a t' un enunciato S di un linguaggio L tale che S esprime p.

Affinché p esista in un dato momento t, (F) richiede che esista a t un enunciato che esprime p, mentre il più debole (D) richiede soltanto che

---

[5] Nei suoi "Afterthoughts" Kaplan sembra aver cambiato nettamente opinione a questo riguardo: cfr. Kaplan 1989: 309–313.

[6] Una discussione molto approfondita dell'esempio di "Newman 1" si trova in Adams 1986.

esista *in qualche momento* t' un enunciato che esprime p. Se p esiste in base a (F) esiste anche in base a (D), ma non viceversa. Per (D), una proposizione esiste (in ogni tempo) se è esprimibile in un linguaggio, senza ulteriori restrizioni; per (F), esiste quando (e soltanto quando) è esprimibile in un linguaggio. Dico "esprimibile" e non "espressa" perché sembrerebbe esageratamente restrittivo far dipendere l'esistenza delle proposizioni dai proferimenti. Sembra chiaro, ad esempio, che si può far riferimento a una proposizione anche in un momento t in cui nessun enunciato la esprime ("Il comunicato dell'ISIS di ieri è particolarmente scandaloso": la proposizione in questione esiste oggi, tant'è vero che possiamo far riferimento a essa, anche se – supponiamo – nessun enunciato emesso oggi la esprime).

Se la dipendenza è concepita come in (D), la proposizione ⟨Nel 1943 Ugo non era ancora nato⟩, che è esprimibile oggi, esiste in ogni momento del tempo e quindi anche nel 1943; dunque può essere vero, nel 1943, che nel 1943 Ugo non era ancora nato. Supponiamo allora che la dipendenza debba essere concepita come in (F). Alcune conseguenze sembrano poco plausibili. Per esempio, dopo l'estinzione della specie umana e delle sue lingue non esisterebbe più la proposizione che il Sole dista 4,22 anni luce da Proxima Centauri (e quindi non sarebbe più vero che il Sole dista 4,22 anni luce da Proxima Centauri), nonostante che esistano ancora il Sole e Proxima Centauri e la loro distanza sia immutata; a meno che una civiltà aliena, dotata di linguaggio, non sia al corrente del Sole e di Proxima Centauri. Peraltro, nell'ipotesi che siamo noi la sola specie dell'Universo dotata di linguaggio, le proposizioni esisterebbero soltanto da 60.000 anni circa: i pensieri dei nostri antenati prelinguistici non avevano contenuto proposizionale, e non erano quindi né veri né falsi. Forse si può convivere con queste conseguenze, ma a me non pare che incoraggino a difendere (F).

Si potrebbe pensare che la dipendenza delle proposizioni dagli enunciati sia in realtà espressa in modo ottimale da una terza condizione, intermedia fra (F) e (D) e, almeno a prima vista, più consona alla concezione di Napoli, e cioè (I):

(I) A ogni t, la proposizione p esiste a t sse per qualche t', $t' \leq t$, esiste a t' un enunciato S di un linguaggio L tale che S esprime p.

Ma a ben vedere anche (I) ha implicazioni bizzarre. Se è esistito, nel più remoto passato, un qualche linguaggio ormai estinto e dimenticato dalla storia che ha espresso la proposizione p, allora p esiste oggi (ed è sempre esistita da allora a oggi), anche se – supponiamo – nessuna delle lingue esistenti è in grado di esprimerla, per esempio perché nessuna lingua ha nomi per le entità di cui la proposizione tratta. Si noti: non è particolarmente bizzarro ritenere che quella proposizione esista. Quel che è poco plausibile è pensare che esista *perché* è stata a suo tempo espressa nella lingua dimenticata. Insomma, un conto è sostenere che una proposizione deve poter essere espressa in qualche linguaggio, e un altro conto è pensare che le proposizioni vengano all'esistenza via via che nascono linguaggi in cui sono esprimibili.

Veniamo ora alla prima considerazione a cui si è accennato poco fa, sulla preferibilità del permanentismo rispetto al temporaneismo. La posizione di Napoli sembra implicare una forma di temporaneismo: ci sono cose che, in certi momenti del tempo, non sono nulla, quindi non è vero che in ogni tempo ogni cosa è sempre qualcosa, come asserisce il permanentismo. La ragione per cui la posizione di Napoli implica la negazione del permanentismo è la seguente. Secondo lui, quando Ugo non esiste – prima della sua nascita – non ci sono proposizioni su Ugo, mentre ci sono proposizioni su Ugo quando Ugo esiste, e da allora in poi. Questo implica o che prima che Ugo esista (come oggetto concreto) Ugo non è nulla, o che non ci sono proposizioni che vertono su oggetti che non sono oggetti concreti esistenti: per esempio, non ci sono proposizioni su oggetti possibili ma non esistenti. Nel primo caso, ci sono momenti del tempo in cui qualcosa, Ugo, non è nulla, quindi il permanentismo è falso. Il secondo caso è in sé problematico, perché sembra ovvio che ci siano proposizioni che vertono su cose che non sono oggetti concreti esistenti, per esempio sul numero 4. Dunque il problema deve riguardare in particolare le proposizioni su oggetti possibili non esistenti (Ugo prima della sua nascita). Sembra che, se non ci sono proposizioni che vertono su oggetti possibili non esistenti, debba essere perché *non ci sono* oggetti possibili non esistenti. Ma questo è difficile da difendere, perché capita che parliamo di oggetti siffatti: per esempio, parliamo di Ugo prima della sua nascita o di un libro che non abbiamo ancora scritto, citandone il titolo. In ogni caso, le proposizioni sono qualcosa (ad esempio, sono possibili oggetti di riferimento): se ci sono momenti del tempo in cui

qualcosa, una proposizione, non è nulla, nuovamente il permanentismo è falso.

Tim Williamson ha scritto un grosso libro (Williamson 2013) per sostenere le ragioni del necessitismo contro il contingentismo, e indirettamente (ma spesso anche direttamente) le ragioni del permanentismo contro il temporaneismo, che sono analoghe. Questo non vuol dire che non sia lecito essere temporaneisti. Williamson non dimostra che il contingentismo e il temporaneismo siano posizioni incoerenti: come ripete più volte, i suoi argomenti sono in massima parte abduttivi, cioè volti a dimostrare che contingentismo e temporaneismo hanno conseguenze più difficilmente accettabili del necessitismo e del permanentismo. Se i suoi argomenti sono convincenti (e molti sembrano esserlo), la palla è nel campo dei temporaneisti e delle posizioni, come quella di Napoli, che implicano il temporaneismo.

Istintivamente, non ho particolare simpatia per la proliferazione ontologica: oggetti possibili non esistenti e proposizioni su oggetti possibili mi mettono a disagio. D'altra parte, sono un tarskiano irriducibile. Ci sono anni in cui il mondo contiene Ernesto Napoli (per esempio il 2015) e anni in cui non lo contiene (per esempio il 1943). Negli anni del primo tipo è vero che il mondo contiene Ernesto Napoli, negli anni del secondo tipo è vero che non lo contiene. Le cose starebbero così anche se Ernesto Napoli non avesse un nome; l'unica differenza è che, in quel caso, non potremmo *dire* che stanno così; allo stesso modo, Giulio Cesare non poteva dire che il sale è cloruro di sodio, pur essendo vero, anche ai suoi tempi, che lo era. Se per salvaguardare la permanenza di "p se e solo se Tp" devo turare il mio naso metafisico, lo farò.

Riferimenti bibliografici

Adams, R. (1986), "Time and thisness", *Midwest Studies in Philosophy*, 11, pp. 315–329, rist. in J. Almog, J. Perry e H. Wettstein (a cura di), *Themes from Kaplan*, Oxford University Press, Oxford 1989, pp. 23–42.

Berto, F. (2013), *Esistence as a Real Property: The Ontology of Meinongianism*, Springer, New York.

Casalegno, P. (1997), *Filosofia del linguaggio*, La Nuova Italia Scientifica (Carocci), Roma.

Heidegger, M. (1927), *Sein und Zeit*, Niemeyer, Halle, trad. it *Essere e tempo*, Mondadori, Milano 2006.

Kaplan, D. (1968), "Quantifying in", *Synthese*, 19, pp. 178–214.

——— (1989), "Afterthoughts", in J. Almog, J. Perry e H. Wettstein (a cura di), *Themes from Kaplan*, Oxford University Press, Oxford, pp. 565–614, trad. it. "Ripensamenti", in S. Raynaud (a cura di), *Tu, io, qui, ora*, Guerini, Milano 2006, pp. 279–316.

Marconi, D. (2006), "On the mind dependence of truth", *Erkenntnis*, 65, pp. 301–318.

McGrath, M. (2014), "Propositions", in E.N. Zalta (a cura di), *The Stanford Encyclopedia of Philosophy*, edizione spring 2014, `http://plato.stanford.edu/archives/spr2014/entries/propositions/`.

Montague, R. (1970), "Pragmatics and intensional logic", *Synthese*, 22, pp. 68–94, rist. in *Formal Philosophy: Selected Papers of Richard Montague*, Yale University Press, New Haven 1974, pp. 119–147.

Napoli, E. (2010), "Realtà, verità e linguaggio", in M. Carrara e V. Morato (a cura di), *Verità. Annuario e Bollettino della Società Italiana di Filosofia Analitica*, Mimesis, Milano, pp. 295–302.

——— (2011), "Reference fixing and the stiffness of reference. Or three (would be) puzzles concerning names", *Rivista di Estetica*, 47, pp. 179–195.

Salmon, N. (1998), "Nonexistence", *Noûs*, 32, pp. 277–319, rist. nel suo *Metaphysics, Mathematics, and Meaning, Philosophical Papers Vol. 1*, Oxford University Press, Oxford 2005, pp. 50–90.

Williamson, T. (2013), *Modal Logic as Metaphysics*, Oxford University Press, Oxford.

Revisable truths*

Andrea Bonomi

0. The problem

Napoli (2000) correctly argues that in his work on mental models Johnson-Laird offers no serious reason to maintain that ordinary reasoning is non-monotonic. This is a very interesting point as concerns natural language semantics, because we can wonder whether, in general, there are expressions, in everyday language, whose interpretation calls for a non-monotonic modelization.

In what follows I will try to show that there are future-oriented statements conceived of by speakers as intrinsically *revisable* and which require a non-monotonic characterization of the changing backgrounds of information selected by the time flow.

1. The future as the mirror image of the past

The term "factivity" is often used by linguists to characterize the behaviour of some operators. In this sense, an operator **O** is factive if, for any sentence ϕ, the truth of "**O**ϕ" entails the truth of ϕ.

It is immediate to see that, if past and future are associated to suitable operators (e.g., "**P**" and "**F**" respectively[1]), these operators cannot be factive *tout court*: of course, the truth, at a time t, of "it was (will) be the case that ϕ" does not entail the truth, at t, of ϕ. But a notion of 'shifted' factivity is available in these cases. In particular, it is quite natural to associate the simple past to 'backward factivity', that is to the idea that the truth of "**P**ϕ", at a given time t, entails the truth of ϕ at a time s preceding t. Such a requirement is made explicit in the classical truth conditions for the (simple) past tense, where $\mathcal{M}$ is the given model, w is a possible world and t is a time:[2]

(1) $\llbracket \mathbf{P}\phi \rrbracket^{\mathcal{M},w,t} = 1$ iff there is a time s such that $s < t$ and $\llbracket \phi \rrbracket^{\mathcal{M},w,s} = 1$.

In other terms, the sentence "**P**ϕ" is true in the model $\mathcal{M}$, in the world w, at the time t, if and only if there is a time s preceding t such that ϕ is true in w at the time s.

In most versions of formal semantics the same kind of reasoning is applied to the future tense, whose truth conditions are often expressed by definitions like:

(2) $\llbracket \mathbf{F}\phi \rrbracket^{\mathcal{M},w,t} = 1$ iff there is a time s such that $s > t$ and $\llbracket \phi \rrbracket^{\mathcal{M},w,s} = 1$.

(2) is obtained from (1) by a simple inversion in the time direction. To use a metaphor exploited by Prior: the semantic representation of the future is here the mirror image of the semantic representation of the past. The idea is, of course, that the former is symmetrical with respect to the latter. Let us call such an assumption *symmetry hypothesis*.

This approach is nicely pictured by a familiar way of representing the time flow: an oriented line which stands for a *single* course of events (i.e., the world where the utterance event occurs) and where events are located before or after a given point (which in the simple cases is the utterance time = now).

---

[1] "**P**ϕ" = "It was the case that ϕ"; "**F**ϕ" = "It will be the case that ϕ".

[2] In the so-called $W \times T$ framework, a model is a structure $\mathcal{M} = \langle W, T, <, D, I \rangle$ where W is a set of worlds, T is a set of times, $<$ is a *linear* order on T ("$x < y$" means that x is earlier than y), D is the domain of individuals and I is a function from $P^n \times W \times T$ to sets of n-tuples of individuals (P^n is the set of n-place predicate letters).

This kind of representation is often associated to a *stability principle* which can be expressed as follows:

(SP) Let u be an utterance of a sentence[3] S and t_u the utterance time. The proposition expressed by S at t_u must be evaluated as true or false at t_u. If this proposition is evaluated as true (false) at t_u, then it must be evaluated as true (false) at any moment $t \geq t_u$.

In what follows, I will try to show that, with suitable qualifications, there are linguistic data showing that the stability principle (SP) is not always applicable and that a flexible notion of propositional content can help to account for the cases in which it fails.

2. Settledness

A principle like (SP) is implicitly questioned by Prior when he discusses what I call the multiple-choice paradox,[4] illustrated by the following example:

> (MCP) Suppose A and B are being pushed towards the edge of a cliff, and there will be no stopping this process until there is only room for one of them.

---

[3] The implicit assumption, here, is that S is no *deviant* sentence, in any plausible sense of the term.

[4] I use this term because Prior's example is a future-tensed version of the 'multiple-choice paradox' discussed in Bonomi 1997: 181–184 with respect to the progressive.

> Then we may be able to say truly that it will definitely be the case that A or B will fall over, even though we cannot say truly that A will definitely fall or that B will definitely fall over. (Prior 1957: 85)

The issue, here, is the nature of the statements concerning *contingent* future events (such as being pushed towards the edge of a cliff and falling over), and the solution proposed by Prior is based on the idea that the existence of *present facts* is an appropriate criterion to distinguish, among the future-tensed statements, those that are *definitely* true (at the utterance time) from those that are not.[5] As we have just seen, in his example this point is illustrated by the statement

(3)   A or B will fall over

which, according to Prior, turns out to be definitely true in the circumstances described above, whilst *neither*

(3a) A will fall over

nor

(3b) B will fall over

is definitely true. (This is the *apparent* paradox.)

In other words, Prior assumes that the evaluation of future-tensed statements as definitely true or false depends on the existence of *present* facts or circumstances. A statement like "x will ϕ" is now true if the truth, in the future, of "x is ϕing" is already *settled*. Let us call *settledness* condition such a requirement.[6]

It is also clear, from his example, that settledness is a property of statements that *depends on time* in this sense: what is not settled at time t can *become* settled at a later time t' in view of new facts. (In the

---

[5] See Prior 1968: 38 for a further illustration of the principle of settledness: "Nothing can be said to be truly 'going-to-happen' (futurum) until it is so 'present in its causes' as to be beyond stopping; until that happens neither 'It will be the case that p' nor 'It will not be the case that p' is strictly speaking true."

[6] In Øhrstrøm and Hasle 1995: 265, Prior's point of view is associated to the following principle:

(P)   The proposition "**F**p" is true now if and only if there exist now facts which make it true (i.e., which will make it true in due course).

original example: at the beginning of the process, that A or B will fall over is not settled, but it *becomes* settled at some point in the process.) This point is made explicit, in connection with the so-called Peircean approach, in Prior 1967: 129:

> "Will" here means "will definitely": "It will be that p" is not true *until* it is in some sense settled that it will be the case, and "It will be that not p" is not true *until* it is in some sense settled that not-p will be the case. (Italics mine.)

From this point of view, forward factivity (and the symmetry hypothesis in general) is no longer a sufficient condition for the truth of a future-tensed statement: the fact that A eventually falls over is *not* a sufficient condition for the truth of the statement (3a) in the given scenario, where *only* the truth of (3) is settled.

One way to implement this idea is to reason in terms of the *possible* courses of events that are compatible with the current state of the world (i.e., with past and present events): a statement is settled, at a time t, if it turns out to be true, at t, in *all* these courses of events, i.e., no matter what the future is like.

Is this a plausible requirement as concerns the truth conditions of the future tense in natural languages? To be sure, there are special situations in which the existence of present (and past) facts makes the truth of a future-tensed statement settled in the sense analyzed by Prior. For example, if I say

(4)   It will be my birthday tomorrow,

the fact that it is April 4th today and that I was born on April 5th makes (4) *definitely* true at the utterance moment. In the theoretical framework sketched above, this means that (4) is true in *every* course of events compatible with the current state of the world.

The problem, in general, is that for most future-tensed statements in our everyday language, that is for most statements about future, but *contingent*, issues, there are no facts whose presence would entail the future occurrence of a particular event as something *already* settled, at least in the sense that any possible course of events in which that event does not occur is definitely ruled out.

Indeed, consider Prior's example again, where the definite truth of (3) is explained by resorting to the existence of facts that are supposed to

settle the issue. Yet, although these facts are taken for granted, how can we rule out the possibility that some unexpected event will prevent *both* A and B from reaching the edge of the cliff? For example, an earthquake might destroy the cliff itself before the event of pushing is completed: however remote this possibility may be, it is nonetheless compatible with the scenario described by Prior. Of course, one might try to point out further restrictions in order to make the given scenario definite enough to make the truth of (3) already 'settled' at the utterance moment. But further unexpected events might be mentioned to contrast the idea that the outcome at issue (i.e., the fact that A or B will fall over) is already 'settled'.

This kind of argument would apply to most situations in which contingent events are concerned. Thus, with the exception of logical truths or statements like (4), if *historical*[7] 'settledness' is assumed as a necessary condition for the truth of a future-tensed sentence, then some statements that we are willing to consider as intuitively true (e.g., the statement expressed by a sentence like "The next Olympic Games will be held in Brazil") would be evaluated as false (or neither true nor false).

It should be remarked that, in discussing Prior's example, my task was *not* to question his idea that forward factivity is not a sufficient condition for the truth of a future-tensed statement, but his idea that it should be replaced by the notion of *objective* or *historical* settledness (as a necessary and sufficient condition). Thus, the problem I have in mind can be expressed in this form: how should we characterize the notion of settledness in order to make it relevant in the truth conditions for future-tensed statements?

To answer this question notice that Prior's approach sounds much more plausible if, instead of considering the *mere* facts and the totality of future possibilities compatible with them, we consider the facts *with respect to* a suitable state of information.

---

[7] The truth of a statement P is historically settled at t if P is true in *every* course of events compatible with the state of the world at t. 'Historical' necessity is formally defined in Thomason 1984.

3. Monotonicity

As we have just seen, on Prior's analysis settledness depends on time, because the truth of a statement may be unsettled at time t, but settled at a time $t' > t$. The reverse is not possible, of course: the truth of a statement cannot be settled at t but unsettled at t', if $t' > t$.

In Thomason's formalization, such an approach is still conservative enough to meet the following requirement of stability:

> *If* the statement expressed by an utterance of sentence S is settled as true (false) at any time t, then it is settled as true (false) at every time $t' > t$.

Let us see why.

As shown in Fig. 2, in the branching time (BT) framework associated to this analysis of tensed statements, the past moments, but *not* the future ones, are linearly ordered: given any moment t, there is only one course of events stemming from t towards the past, whilst there is a plurality of courses of events stemming from t towards the future. This
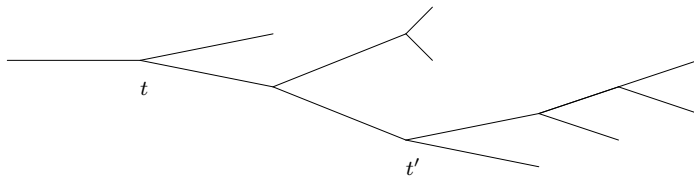

Fig. 2

is so because when you proceed from t towards the future, i.e., when you pass from t to a moment $t' > t$, new information gets available. Which means that the historical alternatives decrease (the 'branches' stemming from t' are fewer than those stemming from t). In other terms, a BT model *à la* Thomason is monotonic in this sense:

(Mon) $t < t' \rightarrow H_{t'} \subseteq H_t$

where, for any moment x, H_x is the set of courses of events passing through x, that is the set of courses of events that are 'historically' possible at x.

An immediate consequence of (Mon) is that in this framework stability of evaluation is respected in the following (weak) sense:

(WSP) Let u be an utterance of a sentence S and t_u the utterance time. The proposition expressed by S at t_u may fail to be evaluated as true or false at t_u or later. But if this proposition is evaluated as true (false) at any moment t, then it must be evaluated as true (false) at any moment $t' \geq t$.

This characteristic is inherited by the semantic system adopted in MacFarlane 2003, 2008, where the only possible transition is from neither true nor false to true (or false), but not from true to false or from false to true. (Actually, as far as I can judge, this kind of semantics is not designed to provide a unified treatment of the multiple interpretations that the future tense has in a natural language like English.) As in Thomason's approach, settledness, for future-oriented statements, is defined in terms of what happens in all the historical alternatives that are live options at the time of evaluation (or assessment). Once more, thanks to the monotonicity of the model, stability of evaluation is not questioned (starting from the moment at which an evaluation is possible).

Yet, there is a preliminary question which should be addressed if we are concerned with the semantics of the temporal markers in natural languages (of the future tense, in particular):

> Are we justified in assuming that the evaluation of the proposition expressed by an utterance is stable (even in the weak sense stated in (WSP))? Does such an assumption conform to the intuitions (if any) of the speakers?

A negative answer is in order here. To see this, imagine the following scenario.

(i)   On June 27 the Republican National Convention nominates Sarah Palin the official candidate for the 2012 Presidential Election.

(ii)   On July 27 Sarah Palin is forced to give up because of her last hunting fiasco (she shot 285 times at a wandering caribou and missed).

(iii)   On October 27, at the end of a new Republican Convention, Michael Moore is nominated the official candidate (and wins the Presidential Election).

Now consider the following sentences:

(5a)   The person who will run for President in the 2012 Election is a woman (uttered on June 28).
(5b)   The person who will run for President in the 2012 Election is *no longer* a woman (uttered on October 28).

From an intuitive point of view, (5a) would be judged as simply *true*, at the utterance moment u, by any competent speaker. This is so because, at u, the definite description "the person who will run for President" refers to Sarah Palin, not to Michael Moore. The obvious idea is that in such cases truth and reference do not depend upon the way the world will *actually* be, but upon the current (appropriate) information, for instance, about the relevant nominations.

The point is that this kind of information can change over time: this is why an utterance of (5b) does not mean, of course, that the candidate has changed sex (as predicted by the usual interpretation of "no longer"), but that something that was true in the past is no longer true at the utterance moment.

As concerns definite descriptions, there is a clear asymmetry between past and future, for the reference of a future-oriented definite description can change over time, as shown by the fact that by uttering (5a) on June 28 we would make a true statement, whilst by uttering it on October 28 we would make a false one. On the contrary, the only natural interpretation of a statement like (6) is that this statement entails a change of sex, not a change of truth value:

(6)   The person who ran for President in the 2008 Election is *no longer* a woman.

This contrast between past and future as concerns definite descriptions can be expressed by the following generalizations:[8]

---

[8] The obvious assumption, here, is that the referent of the definite description does not depend on the presence of indexical expressions, for in such cases a past-oriented

(RefVar) It may happen that the referent of a future-oriented definite description (like "The person who will run for President in the 2012 Election") turns out to be the individual x at a given time t and the individual y ($y \neq x$) at a time $t' > t$.

(RefStab) If, at moment t, x is the referent of a past-oriented definite description (like "The person who ran for President in the 2008 Election"), then x is the referent of that description at any moment t' such that $t' > t$.

4. No longer true

The variability of truth value in the case of future-oriented statements is shown by a very peculiar use of the expression "no longer", witness the following example:

> According to the program of the tomorrow concert, Bill Evans will play in a duo with Jim Hall. Leo, who has heard some vague rumours, asks:

> (7)   What about the tomorrow concert? Is it *true* that Bill Evans is playing with Jim Hall?

> Since Lea is well informed, she promptly answers:

> (8)   (a)   Yes, it is *true*.
>   (b)   (Tomorrow) Bill Evans is playing with Jim Hall.

There is no doubt that such an answer testifies a quite intuitive use of the predicate "true" when it is applied to future-oriented statements and that it would be unnatural to object that, if the event at issue does not take place in the end, such a predicate is misplaced here. Once more, using this predicate in relation with a background of *current* information concerning a *planned* sequence of events (in the sense analyzed in

definite description can have different referents at different times. (Think of a definite description like "the person who bought me a drink yesterday night" which can designate the individual a at time t and the individual b at time t'.) Crucially, the contrast between (5b) and (6) concerns definite descriptions whose referents are fixed by dates.

Copley 2009) is a fact that seems to mirror the speaker's intuitions, *independently* of philosophical speculations about the future and the debate on indeterminism.

Indeed, suppose that tomorrow, before the concert, the program is modified because of some unexpected events. According to the new program, Bill Evans will play with his trio. So, at this point Leo (who has been informed by the organizers of the concert) can call Lea before the concert and say:[9]

(9)   Bill Evans is *no longer* playing with Jim Hall.

As I said, this is a very peculiar use of the phase adverb "no longer".

In a different, and more familiar, kind of context an utterance of (9) would presuppose the existence of a *past* time at which an *event*[10] of Bill Evans' playing with Jim Hall was going on, and would assert that such an event is not going on at the present time.

But, since *no past event* of Bill Evans' playing with Jim Hall is involved in the scenario described above, what does Leo's utterance of (9) presuppose here? And what does it assert?

Roughly speaking, the idea is that this utterance of (9) presupposes that a *planning* about a certain kind of event was in force at some point in the past, whilst it asserts that such a planning is not in force at the utterance time.

There is an interesting relationship between (8b) and (9). Indeed, (9) can be analyzed as follows:

(i)   *Presupposition* (triggered by "no longer"): the proposition expressed by Lea's utterance of (8b) (i.e., the proposition that Bill Evans will play with Jim Hall tomorrow night) *was true until* some moment in the past; it was true, in particular, at the moment of Lea's utterance (in the light of the original program);

(ii)   *Assertion*: this proposition is *not* true at the present moment (considering the new program).

Intuitively, the reason why the statement made by Lea's utterance of (8) is true at the utterance moment u but false at a moment $t > u$

---

[9] This kind of example is discussed in Dummett 2004. See Del Prete 2010 for a similar discussion about examples suggested by Benoît de Cornulier and Orin Percus (p.c.).

[10] Or series of events, on the habitual reading.

(witness the truth of (9)) is that these two moments are associated to two different backgrounds of information (based, respectively, on the original program and the modified program). In other words, the adverb "no longer" signals a change of the truth value which is to be assigned to the statement made by the utterance at issue, *depending on the moment* at which this statement is evaluated. And since a transition from truth to falsehood (and vice versa) is always possible in the case of future-oriented statements, *there is no reason to stick to the stability principle* (not only in its stronger version, but also in the weaker one, according to which the only admissible transition is from neither-true-nor-false to a definite truth value).

As a matter of fact, the content expressed by an utterance of (9) might also be expressed by an utterance of

(9′)   It is no longer true that Bill Evans will play with Jim Hall

where it is evident that what we are evaluating *now* is the statement made by uttering (8b) at some past moment.

So a non-trivial consequence of this short excursus through the "no-longer" clauses is that, as concerns *future-oriented* statements, there are clear cases of *variable* truth values:

(TruthVar) It may happen that the statement made, in an appropriate context,[11] by uttering a future-tensed sentence turns out to be true (false) at a given time t, but no longer true (false) at a time $t' > t$.

This is what happens with statement (8b), witness (9) (or (9′)).

Significantly, nothing similar happens in the case of past-tensed sentences, as stated by the following principle:

(TruthStab) It cannot happen that the statement made, in an appropriate context, by uttering a past-tensed sentence turns out to be *true* at a given time t, but *no longer* true at a later time $t' > t$.

---

[11] The obvious assumption, here, is that there are no gaps in the information which is contextually required and that all the contextual coordinates have been fixed. For example, in the case of (8b), or (9), it must be clear from the context that we are speaking of the tomorrow concert.

5. Contexts

The data we have just discussed lead us to the following conclusions:

(i)   As shown by the way the predicate "true" is used by speakers in re-
        lation to some future-oriented statements, the statement expressed
        by an utterance of a sentence like (5a) or (8b) is evaluated as true,
        at the utterance moment u, by referring to a relevant background
        of information (let us call it VIEW for brevity), whatever course
        of events may be actualized in the end.
(ii)   There is an asymmetry between past and future, in the sense that
        while a single course of events is referred to for the evaluation of a
        past-oriented statement, in the case of a future-oriented statement
        a plurality of alternative courses of events is made relevant: it is
        the set of courses of events that are compatible with VIEW.
(iii)   Some peculiar uses of phase adverbs like "no longer" show that
        different evaluations of a future-oriented statement are possible at
        different moments (because of the variability of VIEW). In other
        terms, this kind of statement is intrinsically *defeasible*, for the vari-
        ability of truth value is not limited to the transition from an in-
        definite truth value to a definite one, but allows for the transition
        from truth to falsehood (and vice versa).

I will skip the details of the formal semantics which is presupposed
here and which is presented in Bonomi 2010. From now on I will concen-
trate on the temporal constraints provided by the context.

For the sake of simplicity, I will ignore the features that are not rele-
vant here by reducing a context c to the triple $\langle u, \text{TT}, \text{RT} \rangle$, where:

– u is the *utterance moment*;
–   TT is the *target time*, i.e., time interval which is contextually deter-
        mined.[12] It is *the interval which is spoken about* in the utterance

---

[12] This is *the time we aim at* in order to locate an event *from a given perspective
point*, located at the reference time. I use the term *target time* (TT), by resorting
to the metaphorical distinction between the aiming device, associated to a temporal
observation point, and the target aimed at by that device, associated to another
temporal location. At least to some extent, the latter notion overlaps with Klein's
notion of *topic time*, that is the time span "to which the speaker's claim is confined"

event: it can be narrowed down by information retrievable from previous discourse, tacit assumptions and so on;

– RT is the reference time *function*, which maps temporal situations to informational backgrounds. Intuitively, the role of RT is to associate to any temporal situation v the VIEW holding at v. In other words, v is the temporal situation where the observation point is located.

Notice that, due to the presence of the reference time *function* RT, an utterance context c has an inherently *dynamic* character, for different informational backgrounds can be associated to different times *in the same context*. In particular, $RT(u)$ is the informational background associated to the utterance moment. Yet, as time goes by, *other* moments become available as moments which feed function RT: for example, moment v, or, later on, moment z, and so on, so that, by suitable *revisions*, other backgrounds of information can become available in the *same* utterance context. As we saw above, the intuition, here, is that, once an utterance event e has taken place, *the effects of this event stretch far along the time line*, where different times can be associated to different backgrounds of information.[13] As a consequence, the new perspective points may involve possible courses of events that were previously ruled out. (While the historical possibilities may only change by decreasing when passing from time t to time $t' > t$, the universe of possibilities associated to $RT(t')$ is not necessarily included in the universe associated to $RT(t)$.)

Given a context $c = \langle u, TT, RT \rangle$, for any expression α, $[\![\alpha]\!]^c$ is the content expressed by α in the context c. $[\![\alpha]\!]^{c,v}$ is the semantic value of α with respect to a temporal situation v. If ϕ is a sentence, $[\![\phi]\!]^{c,v}$ is the

---

(Klein 1994: 4). But since this notion is often seen as replacing Reichenbach's concept of reference time, it is safer to use an alternative term.

[13] What Kratzer (2012) writes about modalized sentences seems to apply to the treatment of future-oriented sentences proposed in this paper: "We might wonder why there should be a unique conversational background for a modalized sentence to express a proposition. This seems too strong. More often than not, conversational backgrounds for modals remain genuinely *underdetermined* and what speakers intend to convey is compatible with several choices of conversational backgrounds" (Kratzer 2012: 32, italics mine). In the case of the future tense, my idea is that there is a sort of *announced indeterminacy* as concerns the background of information which is to be fixed by the context, in the sense that, as time goes by, *different* backgrounds can be associated to *different* moments in the *same* utterance context.

truth value of ϕ in context c with respect to v. The idea underlying the formal truth conditions (omitted here) is that:

(TN) $[\![\phi]\!]^{c,v} = 1$ iff the propositional content expressed by ϕ at u is settled as true at v, that is if and only if ϕ is true on every course of events compatible with the informational background associated to v.

6. Back to non-persistent truths

As we saw when discussing this example, the statement made by an utterance of (8b) can be true if evaluated at the utterance moment u, in view of the original program for the concert, but false at a time v, such that $u < v$ and v is later than the moment at which the program is modified (but earlier than the time at which the concert takes place). That is why

(9)   Bill Evans is *no longer* playing with Jim Hall

or

(9′)   It is no longer true that Bill Evans will play with Jim Hall

can be truthfully uttered at v.

   On the proposal under discussion, this is possible because there is a *change* of perspective when passing from moment u to moment v, and such a change is formally accounted for by the fact that function RT can associate different backgrounds of information to u and v, respectively. In other words, to account for the change of evaluation expressed by (9) or (9′) we can simply say that the proposition expressed by (8b) in the given utterance context $c = \langle u, \mathrm{TT}, \mathrm{RT} \rangle$ turns out to be true with respect to u, but false with respect to v:

(10) $[\![(8b)]\!]^{c,u} = 1$
       $[\![(8b)]\!]^{c,v} = 0$.

   We can have different truth values because the intended proposition $[\![(8b)]\!]^c$ is evaluated relative to different times (u and v, respectively), which in turn correspond to different backgrounds of information. As I

have just recalled, in the formal framework presupposed here this pecu-
liarity is accounted for by resorting to the reference time *function*, which
picks out different backgrounds depending on the time flow. More ex-
actly, given an utterance context $c = \langle u, \text{TT}, \text{RT} \rangle$, this task is achieved
by its third coordinate, function RT, which represents the dynamic side
of c, for it makes different temporal situations available in order to eval-
uate the propositional content with respect to *that* utterance context.
The point is that such a context determines not only the temporal loca-
tion of the utterance event itself (fixed by the first coordinate) and the
time which is spoken about (fixed by the second coordinate), but also
the alternative temporal situations (with the associated backgrounds of
information) which are relevant to the evaluation process.

Specifically, the change of perspective justifying the change of truth
value is explained as follows:

(11) $\text{RT}(u) \neq \text{RT}(v)$,

where $\text{RT}(u)$ is the set of courses of events compatible with the original
program for the concert (which is *still* valid at u), while this program
is *no longer* valid at v, so that $\text{RT}(v)$ is the set of courses of events in
which Bill Evans does not play with Jim Hall but with his trio.

Thus, we have detected an important source of contextual dependency,
because the truth of an *utterance* is relative to the background of infor-
mation selected by the reference time function RT. *Stretching the utter-
ance context* in order to cover different temporal positions makes new
backgrounds of information relevant to evaluating the content of that
utterance and allows for a principled explanation of the transition from
a definite truth value to its opposite.

As desired, the sequence of informational backgrounds is non-monotonic
in the following sense:

> for any two moments x and y:
> if $x < y$
> then $H_y \subseteq H_x$
> but it may happen that $\text{RT}(y) \not\subseteq \text{RT}(x)$.

7. Contexts and propositional profiles

A more careful account of the role of time in fixing the relevant truth conditions is in order at this point.

As an illustration, consider two possible utterances of (12):

(12) Bill Evans is playing with Jim Hall.

In the first scenario, (12) is a very natural answer to a question (concerning the identity of Bill Evans' partner) asked by a person during a concert at the Montreux Jazz Festival. What is involved here is an event *which is occurring at the utterance moment*. But, as we saw above, (12) can be used, in a different context, in order to speak of a *planned* event, whose occurrence is located in the *future* of the utterance moment. *Both* these readings are accounted for by the kind of truth conditions discussed above, thanks to different presuppositions concerning the target time TT. Indeed, if we look at the time which is spoken about (i.e., TT) we see that it coincides with the utterance time in the first scenario, but *not* in the second scenario, where the situation is more complex: utterance time and perspective point still coincide (for it is in the light of the *information available at the utterance time* that we judge the statement at issue as true, if it is used, for instance, as an answer to a question like "*Is* it true that ...?"), but they do *not* coincide with the target time (which is the time of the tomorrow concert).

TT can be fixed by contextual factors such as a previous discourse (as suggested by the second scenario) or current evidence (our presence at the concert, in the first scenario). The idea is that an utterance of a sentence like (12) *concerns* a particular time interval, which can be located in the present or in the future of the utterance time. This interval which is spoken about has a crucial role to play in defining the content of an utterance.

This is the role Frege has in mind when in a famous passage he explains how the utterance time contributes to determining the time we refer to by using a tensed sentence:

> If a time indication is needed by the present tense, one must know when the sentence was uttered to apprehend the thought correctly. Therefore, *the time of utterance is part of the expression of the thought*. If someone wants to say the same today as he expressed yesterday using the word "today", he must replace

this word by "yesterday" [...] The mere wording, as it is given in writing, is not the complete expression of the thought, but the knowledge of certain accompanying conditions of utterance, *which are used as a means of expressing the thought*, are needed for its correct apprehension. (Frege 1918: 24; italics mine)

So, the 'complete' expression of a thought or proposition must contain a specification of the time the statement at issue is about, and thanks to such a specification (made possible by the implicit or explicit reference to the utterance moment) the evaluation of the thought or proposition at issue is fixed once and for all. And the stability of evaluation follows.

The contrast, here, is between a complete expression of the thought or proposition and an incomplete one. According to eternalism, the latter has *no* semantic relevance: there is no intermediate entity, no temporally *neutral* proposition, which accounts for the dependency of the evaluation on a temporal parameter. This is so for the simple reason that such a parameter is *incorporated* into the expression of the thought.

I will not address here Kaplan's well-known argument against this line of thought, an argument based on the role of temporal operators: applying these operators, so runs the objection, to propositions where the temporal information is completely specified would be tantamount to using them vacuously.[14] I will turn instead to the role that Prior attributes to temporally neutral propositions to account for some peculiar uses of tensed sentences.

Interestingly enough, his starting point is the same as Frege's: the time a proposition *is about* (which in many cases coincides with the utterance time) is an essential ingredient to determine the full content expressed by an utterance event: "[A tensed language] *implicitly refers* to the time of utterance, and by tensing what is implicitly said of the time of utterance it can indirectly characterise other times also [...]. In at least the most elementary tensed languages instants or times are not mentioned, but tensed propositions are understood as directly or indirectly characterising the *un*mentioned time of utterance" (Prior and Fine 1977: 30). So, on this account, the time which is spoken about, with its anchoring effect, plays a crucial role in determining the full content expressed by an

---

[14] See Recanati 2007 for a reconstruction of the debate between eternalists and temporalists.

utterance. Still, we can insist that there are plausible reasons to isolate a notion of content which is *independent* of that kind of anchoring.

To see this, consider the following situation.[15]

On September 27, 2014, Leo, a famous economist, says in an interview:

(13)   Italy is facing a severe crisis.

As everyone knows, this a true statement.
One year later, after reading the old interview, he comments:

(14)   Thank Goodness, what I said one year ago is no longer true.
        (Italy is out of the crisis.)

Now, consider (13) and suppose that, as required by the kind of temporal anchoring which is necessary in order to get 'eternal' propositions, the time which is spoken about (*and which coincides with the utterance time in this case*) is incorporated into the content expressed by Leo's utterance.

If we look at the content expressed by Leo when he utters (13) to speak of the Italian crisis, we observe the following:

> *utterance time = reference time (perspective point) $\neq$ time which is spoken about* (TT).

It is the utterance time that Leo has in mind when he utters (13) in order to locate the relevant event or state (Italy's crisis) and *it is with respect to this very moment, and the informational background associated to it, that his utterance is to be evaluated as true or false.* But if it is true, of the utterance moment u, that Italy is facing a severe crisis, then there is *no* moment t, such that: $t \geq u$ and it is false at t that Italy is facing a severe crisis at u.

In this case, to determine the content no parameter is abstracted over and what we get is *eternalism*:

---

[15] This example is reminiscent of Prior's "Thank Goodness, it's over". Notice that the situation depicted by a sequence like (13)–(14) can be more complex. Imagine two economists, A and B, who speak different languages. For example, one of them utters (13), whilst the other, who speaks Italian, utters "L'Italia sta attraversando una crisi molto seria". Supposing that these utterance events take place at the same time, one year later an observer C might comment: "Thank Goodness, what A and B said one year ago is no longer true".

(A) $\lambda v [\![(13)]\!]^{\langle x,x,\text{RT}\rangle,x}$, where $x = $ the situation corresponding to the utterance time (presupposition).

This is a *constant* function (proposition), since, for any temporal situation s, such that $x \leq s$:

$\lambda v [\![(13)]\!]^{\langle x,x,\text{RT}\rangle,x}(s) = 1$, where $x = sit_1$ (the temporal situation corresponding to the interview).

By sticking to the time which is originally spoken about (= the utterance time) we get a propositional profile whose truth value is fixed once and for all, for any s it applies to. Time flow has no relevance here, because the propositional content has incorporated all the necessary information.

But the eternalist propositional content in (A) cannot be referred to if we want to account for the comment made by uttering (14). This comment can be plausible (and true) only by associating the expression "what I said" to a propositional content which is *not* anchored to the utterance time of (13), and which includes a *shiftable* component. In other terms, we have to isolate a temporally *neutral content* that can be obtained by abstracting over the parameter represented by the utterance time of (13), which coincides with the evaluation time and with TT.

Indeed, by uttering (14), Leo does *not* intend to *revise* his original statement, which was, is and will be true: the expression "what I said", in (14), denotes a content that is not temporally anchored to the time which is spoken about in (13), i.e., the utterance time of (13). More exactly, such a content can only be obtained by *abstracting over* that contextual parameter (which coincides with TT), for there is a sense in which (14) might be paraphrased as follows:

(14′) If I should say at the present moment what I said one year ago, I would say something false,

where the propositional content referred to by the expression "what I said" is not anchored to the situation Leo had in mind when he uttered (13).

So the appropriate content corresponding to the expression "what I said" one year ago in (14) can be represented as the following propositional profile, where the time which is spoken about is shiftable (is λ-bound):

(B) $\lambda v [\![(13)]\!]^{\langle v,v,\mathrm{RT} \rangle,v}$.

Indeed, as shown by the discussion about (13)-(14), the comment made by uttering

(14) Thank Goodness, what I said one year ago is no longer true

can be plausible (and true) only by associating the expression "what I said" to a propositional profile which is *not* anchored to the time which is originally spoken about and which includes a *shiftable* component. We have here a *temporally neutral* content which, if applied to different temporal situations, yields the intended result:

$$\lambda v [\![(13)]\!]^{\langle v,v,\mathrm{RT} \rangle,v}(sit_1) = 1$$
$$\lambda v [\![(13)]\!]^{\langle v,v,\mathrm{RT} \rangle,v}(sit_2) = 0,$$

where sit_1 is the temporal situation corresponding to the interview, sit_2 is the temporal situation in which the economist comments his old statement.

Similar remarks apply to the Sarah Palin's case discussed above. But there are some interesting differences.

When, on June 28, Leo utters

(15) A woman will run for President

he says something intuitively true.

But, on October 28, after Sarah Palin's withdrawal and Michael Moore's nomination, he might comment:

(16) What I said three months ago is no longer true.
      (A man will run for President.)

There is a strong similarity, of course, between (14) and (16), because both of them raise a problem of truth value variability. But the Sarah Palin story has a peculiarity which deserves a short reflection.

What distinguishes the sequence (13)–(14) from (15)–(16) is the fact that the expression "what I said", in (16), denotes a content temporally *anchored* to the time which is spoken about in (15): as shown by the second sentence in (16), the speaker is still referring to the time of the next Presidential election. What he means, by uttering (16), is that the

anchored proposition associated to the utterance of (15) is no longer true if evaluated at a different time. In other terms, the reference time (where the perspective point is located) is made shiftable by abstraction, but the time which is spoken about (i.e., the time of the Presidential Election) remains *unchanged*.

The difference, with respect to (13)–(14), is that (15), unlike (13), has the following profile:

> *utterance time = reference time (perspective point) ≠ time which is spoken about* (TT).

That is why we can speak of a *revisable* statement made by uttering (15): what the speaker said *about* a given time located in the future is judged to be true at t, but no longer true at t', with a different perspective point.

This means that if, by lambda abstraction, we want to determine an appropriate content (corresponding to the expression "what I said" in (16)), the value which should be assigned to TT is *not* shiftable: unlike the utterance time (which coincides with the perspective point), it cannot be λ-bound. So, what we get is the following propositional function:

(C) $\lambda v [\![(14)]\!]^{\langle v, \text{TT}, \text{RT} \rangle, v}$ [TT $= t$ (presupposition), where t is the time of the Presidential Election.]

This peculiarity of future-oriented statements comes as no surprise in the theoretical framework adopted here: the passing of time modifies not only the state of the world, but also the state of the relevant information, which is an essential ingredient of the truth conditions for this kind of statements.

The moral to be drawn, after this short survey, is that the same utterance may be associated to different propositional profiles, depending on which contextual parameters stay fixed and which are abstracted over.

This is clear, as we have just seen, in the case of the utterance of (13) discussed above:

(13) Italy is facing a severe crisis.

Indeed, this utterance event can be associated to an 'eternal' proposition (of type $\lambda v [\![(13)]\!]^{\langle x, x, \text{RT} \rangle, x}$) if we are concerned with what Leo,

the famous economist, said *about a given temporal situation* (September, 2014). But it can also be associated to a different proposition (of type $\lambda v[\![(13)]\!]^{\langle v,v,\mathrm{RT}\rangle,v}$) if we abstract from that temporal location and we focus on alternative time spans, as shown by the comment "Thank Goodness...".

In these cases different propositional contents are available for the same utterance event, since *what we abstract from* in order to determine *what was said* depends on the conversational situations we are engaged in when we talk about that event.

References

Bonomi, A. (1997), "The progressive and the structure of events", *Journal of Semantics*, 14, pp. 173–205.

——— (2010), "Imperfect propositions", `http://filosofia.dipafilo.unimi.it/bonomi/impbonomi.pdf`.

Copley, B. (2009), *The Semantics of the Future*, Routledge, New York.

Del Prete, F. (2010), "Non-monotonic futures", MS.

Dummett, M. (2004), *Truth and the Past*, Columbia University Press, New York.

Frege, G. (1918), "Der Gedanke. Eine logische Untersuchung", *Beiträge zur Philosophie des deutschen Idealismus*, 1, pp. 58–77, engl. transl. "The thought", *Mind*, 65 (1956), pp. 289–311.

Klein, W. (1994), *Time in Language*, Routledge, London and New York.

Kratzer, A. (2012), *Modals and Conditionals*, Oxford University Press, Oxford.

MacFarlane, J. (2003), "Future contingents and relative truth", *Philosophical Quarterly*, 53, pp. 321–336.

——— (2008), "Truth in the garden of forking paths", in M. Garcìa-Carpintero and M. Kölbel (eds.), *Relative Truth*, Oxford University Press, Oxford, pp. 81–102.

Napoli, E. (2000), "Inferenze non monotòne?", in P. Cherubini, P. Giaretta and A. Mazzocco (eds.), *Ragionamento: Psicologia e Logica*, Giunti, Firenze, pp. 69–73.

Øhrstrøm, P. and Hasle, F.V. (1995), *Temporal Logic*, Kluwer, Dordrecht.

Prior, A. (1957), *Time and Modality*, Clarendon Press, Oxford.

——— (1967), *Past, Present, and Future*, Oxford University Press, Oxford.

——— (1968), *Papers on Time and Tense*, Clarendon Press, Oxford.

Prior, A. and Fine, K. (1977), *Worlds, Times, and Selves*, Duckworth, London.

Recanati, F. (2007), *Perspectival Thought*, Oxford University Press, Oxford.

Thomason, R.H. (1984), "Combinations of tense and modality", in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic, Vol. 2*, Reidel, Dordrecht, pp. 135–165.

Marginalia sulla nozione di conseguenza logica

Mauro Mariani ed Enrico Moriconi

Alla fine degli anni Ottanta del secolo scorso, il secondo degli autori di questo lavoro scrisse in collaborazione con Ernesto Napoli un paio di articoli dedicati alla critica neoverificazionista della concezione classica del significato.[1] I protagonisti della contesa erano, da una parte, principalmente Michael Dummett e Dag Prawitz e, dall'altra, i promotori della semantica formale classica, alla cui base ci sono fondamentalmente i lavori di Alfred Tarski.[2] Il tema centrale dei due articoli prima ricordati erano le caratteristiche della nozione classica, *realista*, sia di verità sia di verità logica, nonché alcune difficoltà insite nel progetto di erigere una teoria del significato alternativa, imperniata sul concetto di dimostrazione, o verificazione; o, come si dice spesso, su una nozione *anti-realista* di verità. In qualche modo connesse a quei due lavori degli anni Ottanta sono le considerazioni, diciamo storico-teoriche, che intendiamo svolgere in queste pagine, e che riguardano quella generalizzazione del concetto di verità logica che è la nozione di *conseguenza logica*. O meglio, e per

---

[1] Moriconi e Napoli 1987, 1988.

[2] Tra le sue fondamentali memorie degli anni Trenta, faremo riferimento in particolare a Tarski 1936. Questo testo epocale, di cui Tarski approntò più o meno contemporaneamente una versione tedesca e una polacca, ha avuto una prima traduzione in inglese in Tarski 1956, con il titolo "On the Concept of Logical Consequence", traduzione poi rivista nella seconda edizione che il libro ebbe nel 1983. Più recentemente, nel 2002, l'articolo in questione ha avuto una nuova e più accurata traduzione inglese, a cura di Magda Stroińska and David Hitchcock, condotta tenendo conto sia della versione tedesca sia di quella polacca, e pubblicata con il titolo "On the Concept of Following Logically". È a quest'ultima edizione che faremo riferimento in queste pagine.

restare al dettato dell'ultima traduzione del saggio tarskiano, il concetto di 'seguire per logica'. Questo concetto ha una lunga storia, che in larga misura coincide con la storia della logica stessa; e non è ovviamente nostra intenzione ripercorrerne qui anche solo le tappe principali. Più modestamente, il nostro scopo è individuare i tratti salienti di due modi di caratterizzare questa nozione, da parte rispettivamente di Aristotele e Bolzano, e di metterli a confronto con quello tarskiano allo scopo di farne risaltare la peculiarità. Le ragioni profonde di questa scelta, che per certi versi appare fin troppo ovvia, saranno chiarite da quello che segue.

1. La caratterizzazione aristotelica

Iniziamo ricordando la definizione aristotelica di sillogismo:

> Il sillogismo è un discorso in cui, poste certe cose, segue di necessità qualcosa di diverso dalle cose che sono state poste per il fatto che queste cose sono. (*Analitici Primi* I 1, 24b18-20)[3]

Questa definizione, che troviamo anche in *Topici*, I 1, 100a25-27 (con l'unica differenza che, al posto di "per il fatto che queste cose sono" troviamo "attraverso queste cose") e, con maggiori ma inessenziali varianti, in *Confutazioni Sofistiche* 1, 164b27-165a2, rappresenta la prima caratterizzazione a noi nota della nozione di seguire logicamente.

In questa sede, come abbiamo detto, siamo interessati principalmente al confronto con la nozione di conseguenza logica come sarà caratterizzata da Tarski.[4] Alcune differenze saltano subito agli occhi: il sillogismo deve essere produttivo, ossia la conclusione deve essere diversa dalle premesse, e le premesse devono essere causa della conclusione: la clausola "per il fatto che queste cose sono" equivale, infatti, a "a causa di queste cose", che a sua volta, come risulta dalla definizione dei *Topici*, significa "attraverso queste cose". Questo requisito può essere inteso in due modi:

---

[3] Questa e le altre citazioni dalle opere di Aristotele sono traduzioni nostre. Nelle successive citazioni il riferimento a *Analitici Primi* verrà omesso.

[4] Ci sentiamo autorizzati a dare per note le caratteristiche generali della definizione tarskiana, riservandoci poi di fare alcune osservazioni al proposito nel paragrafo 3 di questo lavoro.

(i) le premesse devono essere sufficienti per derivare la conclusione;
(ii) non ci devono essere premesse inutili ai fini della derivazione della conclusione.

Aristotele lo intende in entrambi i modi: (i) è esplicitamente asserito in I 1, 24b21-22 ("Segue a causa di queste cose" significa "non c'è bisogno di nessuna premessa estranea perché si produca la necessità"); (ii) è prescritto invece in *Topici*, VIII 11, 161b28-30 ("Ancora [l'argomentazione può essere criticata] se è possibile eliminare alcune premesse: talvolta assumono infatti più del necessario, cosicché *non per il fatto che queste cose sono* si genera il sillogismo"). Inoltre (i) e (ii) sono richiamati in I 32, 10-22, in particolare alle righe 18-20:

> Dobbiamo considerare se è stato assunto qualcosa di superfluo e se qualcosa di necessario è stato omesso, e porre il secondo eliminando il primo fino a giungere alle due premesse.

(ii) ha suscitato però qualche perplessità per via della discussione aristotelica della cosiddetta *falsa causa* (cfr. II 17 e *Confutazioni Sofistiche* 5, 167b21-36). Questa fallacia viene commessa da chi in una dimostrazione per assurdo utilizza in maniera non corretta l'ipotesi d'assurdo: ad esempio, se "Ogni A è B" è assunta come ipotesi d'assurdo, "Ogni B è C" e "Ogni C è D" sono le altre premesse e l'assurdo è rappresentato da "Ogni B è D", la fallacia consiste nell'inferire da ciò la falsità di "Ogni A è B", da cui l'assurdo ovviamente non dipende. Ora, è stato sostenuto che parte di questa dimostrazione consiste nella derivazione sillogistica di "Ogni B è D" dalle premesse "Ogni A è B", "Ogni B è C" e "Ogni C è D", dove evidentemente "Ogni A è B" è inutile rispetto alla conclusione. Tutto ciò, però, non dimostra l'ammissibilità di derivazioni con premesse superflue, anzi la fallacia nasce proprio dal fatto che nel derivare sillogisticamente l'assurdo si sono assunte premesse superflue. Un'ulteriore conferma, dunque, di quanto esplicitamente asserito in *Topici*, VIII 11, 161b28-30.

La diversità della conclusione deve essere intesa in senso lato, non basta che la conclusione non coincida con una delle premesse. In particolare le conversioni valide delle premesse sillogistiche ("Nessun A è B" equivale a "Nessun B è A"; "Qualche A è B" equivale a "Qualche B è A"; da "Ogni A è B" segue "Qualche B è A") non sono sillogismi – Aristotele afferma ripetutamente che non c'è sillogismo con una sola

premessa – anche se la conversa segue di necessità dalla premessa convertita. Ora, conversa e convertita sono ovviamente diverse non solo come *item* linguistici, ma anche per il loro contenuto, come lo stesso Aristotele riconosce (cfr. II 1, dove si osserva che un sillogismo può avere più conclusioni perché, ad esempio, se la conclusione è un'universale negativa il sillogismo avrà come altra conclusione la sua conversa). In che senso, dunque, la conversa non costituisce "qualcosa di diverso"? A nostro avviso la risposta più plausibile, anche se, come tutte, puramente congetturale, è che la conclusione è diversa perché connette termini che le premesse, prese singolarmente, non mettono in relazione tra loro: ad esempio nel sillogismo in BARBARA "Ogni A è B, Ogni B è C, quindi Ogni A è C" il medio B fa da ponte tra i termini A e C che non erano in relazione in nessuna delle due premesse.

Da queste osservazioni risulta che il sillogismo aristotelico esprime una nozione di seguire logicamente per certi versi analoga a quella delle logiche rilevanti, ma sicuramente più stretta, ad esempio, dell'*entailment* di Anderson e Belnap[5] (che ha in comune con il sillogismo la condizione necessaria che conclusione e premesse abbiano un termine in comune, ma che ammette la conseguenza banale di una premessa da se stessa). Detto ciò, è abbastanza inutile cercare di istituire parallelismi più stretti, se non altro per l'assenza in Aristotele di ogni trattazione formale del calcolo proposizionale.

L'altro tratto saliente della definizione di sillogismo è l'aspetto modale: la conclusione segue *di necessità* da ciò che è stato posto. Viene naturale chiedersi se "seguire di necessità" vada inteso come un primitivo non analizzabile, oppure se sia almeno possibile delucidarne in qualche modo il significato. In effetti Aristotele non fornisce mai un'*esplicita* riduzione della nozione di necessità, in nessuna delle sue accezioni, ad altri concetti più fondamentali; o meglio, la necessità è ricondotta, in generale, all'impossibilità di essere altrimenti (*Metafisica* V 5, 1015a33-35). Per un logico modale moderno questa spiegazione non esprime che l'interdefinibilità degli operatori modali; ma Aristotele ha una concezione metafisica della potenza come principio del mutamento e quindi, tutto sommato, la sua caratterizzazione della necessità ha un certo valore esplicativo. Sulla scorta della distinzione in I 10 tra necessità assoluta e necessità "sussistendo certe cose", potremmo raffigurarci il "seguire di necessità" come

---

[5] Anderson e Belnap 1967.

inibizione, da parte di ciò che è stato posto, di potenzialità che di per sé sarebbero inerenti ai termini contenuti nella conclusione. Data la coppia di premesse (in seconda figura) "Ogni uomo è animale" e "Nessun bianco è animale", Aristotele dice infatti:

> Anche uomo non apparterrà a nessun bianco, ma non di necessità. È possibile infatti che l'uomo diventi bianco, non tuttavia fino a quando animale apparterrà a nessun bianco. Cosicché la conclusione sarà necessaria stando così le cose, ma non in senso assoluto. (I 10, 30b36-40)

In altre parole, la potenzialità di divenire bianco inerente agli uomini in quanto tali viene annullata dall'assunzione della premessa che nessun bianco è animale e quindi avremo la conclusione che nessun bianco è uomo a partire dalla premessa che ogni uomo è animale.[6] Questa, naturalmente, è solo una delucidazione, ed è inoltre poco probabile che possa funzionare in tutti i casi, per cui non può essere considerata una riduzione ad altro della necessità sillogistica.

Aristotele procede dunque alla costruzione della sillogistica non modale (categorica), ossia di una *macchina dimostrativa* che, per così dire, implementi la definizione data, innanzitutto regimentando in maniera rigorosa la nozione di seguire di necessità. Una macchina del genere deve essere 'puramente formale', nel senso di essere applicabile a qualsiasi contenuto: da qui l'uso delle lettere come segnaposto per i termini concreti, ossia per le parti non logiche degli enunciati (le parti categorematiche, per usare la terminologia medioevale).[7] Negli *Analitici* la distinzione tra

---

[6] Poiché "Ogni uomo è animale" è necessaria in senso assoluto, per questa premessa non può essere questione di 'fino a quando' e perciò l'inibizione della possibilità che gli uomini diventino bianchi è imputata alla sola premessa che nessun bianco è animale.

[7] In seguito soprattutto ai lavori di Barnes (cfr. soprattutto Barnes 2007), si è molto discusso recentemente sul significato dell'uso delle lettere da parte di Aristotele e soprattutto della loro indispensabilità. Ora, se, seguendo le orme di Łukasiewicz (cfr. Łukasiewicz 1957), riteniamo che i modi sillogistici siano enunciati della forma "se-allora" con una quantificazione universale almeno implicita sui termini che vi compaiono, le lettere sarebbero variabili terministiche e quindi il loro uso sarebbe praticamente indispensabile. In caso contrario sono solo un utilissimo mezzo per mostrare la struttura logica degli enunciati e per sottolinearne il carattere formale, in linea di principio equivalente ad altri (e in effetti Aristotele formula i modi sillogistici anche usando, invece delle lettere, "il primo, il secondo e il terzo" oppure "il primo, il medio e l'ultimo"). Dovrebbe essere chiaro che la nostra posizione è più vicina a questa seconda interpretazione.

parti logiche e parti non logiche è sufficientemente netta. Gli enunciati che vengono presi in considerazione sono infatti solo quelli che costituiscono il cosiddetto quadrato aristotelico ed esprimono l'appartenenza o la non appartenenza, in forma universale o non universale, di un termine B a un termine A, dove i termini logici sono quanto è necessario per esprimere questa appartenenza. Al contrario nei *Topici*, dove l'uso delle lettere segnaposto è del tutto assente, non esiste una regimentazione di questo tipo e, di conseguenza, nemmeno la possibilità di distinguere in maniera non arbitraria ciò che è logico da ciò che non lo è; e inoltre gli schemi di argomentazione hanno portata più o meno ampia, ma in linea di massima non pretendono di essere applicabili universalmente a qualsiasi contenuto.

Ora, una conclusione esprime, come abbiamo appena detto, l'appartenenza o la non appartenenza, in forma universale o non universale, di un termine B a un termine A. Questa connessione deve essere ricondotta all'esistenza di un termine medio che faccia da ponte tra B e A; ma per fare ciò il termine medio deve comparire in una coppia di enunciati che contengono l'uno il termine A e l'altro il termine B e dai quali, assunti come premesse, la conclusione segua di necessità. Diventa ora possibile selezionare tutti i modi sillogistici validi, ossia tutti e soli i tipi di coppie di premesse da cui un dato tipo di conclusione segue di necessità. Alcuni modi, detti perfetti, risultano intuitivamente validi sulla base di cosa significa appartenere o non appartenere universalmente, mentre la validità di altri, detti imperfetti, è provata riducendoli ai primi mediante conversione delle premesse o dimostrazione per assurdo. Infine, tramite la costruzione di controesempi, Aristotele dimostra che tutti gli altri modi non sono universalmente validi.

Abbiamo detto che la macchina dimostrativa della sillogistica regimenta in maniera rigorosa la nozione di seguire di necessità. Ma questo rischia di essere un semplice *fiat* arbitrario se non si fornisce qualche argomento per mostrare che la nozione di seguire sillogisticamente è, almeno estensionalmente, equivalente a quella intuitiva (e, come abbiamo visto, in qualche modo suscettibile di delucidazione metafisica) di seguire di necessità. Sappiamo che in linea di principio non è possibile dimostrare rigorosamente l'equivalenza tra una nozione formale e una intuitiva,

tuttavia si possono trovare argomenti che ne mostrino la plausibilità.[8]
E ciò è proprio quello che cerca di fare Aristotele. Innanzitutto, le considerazioni del capoverso precedente riassumono quanto egli dice in I 23 per dimostrare che ogni conclusione può essere ricavata da opportune coppie di premesse mediante applicazione di opportuni modi sillogistici. In secondo luogo, in I 23-25 si considerano tre casi in cui la derivazione della conclusione non è immediatamente riconducibile alla semplice applicazione di un modo sillogistico:

(a)   La conclusione deriva da più di due premesse. Aristotele cerca di dimostrare che questa derivazione può essere ricondotta a una serie di applicazioni dei modi sillogistici validi (a partire da coppie appartenenti all'insieme delle premesse date si ricavano sillogisticamente conclusioni intermedie, e così via fino a giungere alla conclusione).[9]

(b)   La conclusione è derivata per assurdo. L'intera derivazione non può ovviamente essere ridotta alla macchina sillogistica, tuttavia quest'ultima è sufficiente per derivare dalle premesse date e dall'ipotesi d'assurdo una conclusione assurda: argomentare da ciò alla falsità dell'ipotesi d'assurdo è una mossa per così dire standard, per cui la macchina sillogistica è sufficiente per controllare la validità formale della derivazione.

(c)   Nel caso, invece, in cui l'ipotesi non sia d'assurdo, ma sia, ad esempio, l'accordo di accettare q una volta dimostrata p, quest'ultima viene derivata sillogisticamente; ma l'accettazione di q non richiede, a sua volta, che q segua logicamente da p, è solo una mossa del gioco dialettico basata sull'accordo dei partecipanti alla discussione.[10]

---

[8] L'esempio classico è la tesi di Church (che afferma l'equivalenza tra la nozione intuitiva di computabilità e quella definita formalmente, ad esempio ricorrendo alle macchine di Turing), a sostegno della quale si adduce, tra l'altro, l'equivalenza tra le varie definizioni formali di computabilità.

[9] Questa dimostrazione è per lo più considerata carente, ma non è questa la sede per procedere a un'analisi dettagliata; quello che c'interessa mettere in luce è l'obiettivo che Aristotele vuole raggiungere.

[10] Naturalmente le cose non sono così semplici. Si potrebbe infatti dire – e ai nostri occhi sembrerebbe senz'altro più corretto – che q segue di necessità da p (derivato sillogisticamente) e dall'*ipotesi* "Se p, allora q". Da qui la lunga polemica tra gli Stoici, che vedevano sostanzialmente le cose in questo modo, e gli Aristotelici, che non accettavano invece questo tipo di ipotesi tra le premesse di una derivazione formalmente corretta.

Ma questo non basta. La regimentazione di "seguire logicamente" richiede, come abbiamo visto, che le premesse abbiano una certa forma standard. Aristotele ha argomentato che dato un enunciato, sempre in forma standard, è possibile trovare premesse da cui derivarlo sillogisticamente. Ma come escludere che lo stesso enunciato non segua di necessità anche in maniera non sillogistica, magari da premesse non immediatamente riconducibili a una delle forme previste dal quadrato?

In I 32 Aristotele considera due casi di conclusioni non derivate sillogisticamente in forma rigorosa: il primo (47a10-22) in cui alcune premesse mancano o viceversa ve ne sono di superflue; il secondo (47a22-40), in cui, nonostante il fatto che le premesse non abbiano forma sillogistica, la conclusione segue da esse di necessità (situazione che corrisponde appunto alla difficoltà cui abbiamo accennato nel capoverso precedente). Aristotele non cita esempi del primo caso, ma può darsi che abbia in mente, tra l'altro, il sillogismo retorico, in cui spesso le premesse che stabiliscono qualcosa di ovvio o universalmente noto vengono omesse (cfr. *Retorica* I 2, 1357a7 e sgg.).[11] Gli esempi del secondo caso sono i seguenti:

> Se una non sostanza viene distrutta, non viene [in seguito a ciò] distrutta una sostanza.
>
> Se ciò di cui una cosa è composta viene distrutto anche il composto viene distrutto.
>
> ∴ La parte di una sostanza è sostanza.
>
> Se essendo l'uomo, di necessità l'animale è e se essendo l'animale, di necessità la sostanza è, allora essendo l'uomo, di necessità la sostanza è.[12]

A proposito di questi Aristotele osserva che "il necessario ha maggiore estensione del sillogismo: infatti ogni sillogismo è necessario, ma non tutto il necessario è sillogismo" (33-35).

In un certo senso questa osservazione è banale: la trattazione aristotelica delle conversioni mostra, infatti, che "seguire di necessità" non si

---

[11] Sillogismi di questo tipo sono detti comunemente "entimemi". In realtà per Aristotele l'entimema è il sillogismo retorico in genere, ma, appunto perché in questo genere di sillogismo spesso le premesse troppo ovvie vengono tralasciate, entimema ha finito per assumere, per metonimia, il significato di sillogismo in cui sono state omesse delle premesse.

[12] Cfr. *De Interpretatione*, 10, dove gli enunciati di questo tipo, che i medioevali chiamavano *de secundo adiacente*, sono considerati la 'prima affermazione'.

riduce sempre alla necessità sillogistica. Ma le conversioni non sono una parte essenziale della macchina dimostrativa: è vero, certamente, che per loro tramite si dimostra metateoricamente la validità dei sillogismi imperfetti, ma questi, a loro volta, rendono superflue[13] le conversioni come mezzo di dimostrazione sillogistica. Quindi la necessità delle conversioni non mette in gioco, come in questo caso, la riduzione del seguire di necessità alla necessità sillogistica. Proviamo dunque a ricostruire (anche se la mancanza di indicazioni precise rende la cosa piuttosto congetturale) i precedenti argomenti. Per quel che riguarda il primo argomento si può ragionare così: se per assurdo una parte X di una sostanza Y non fosse sostanza e questa parte fosse distrutta, per la prima premessa la distruzione di X non dovrebbe comportare quella di Y; ma per la seconda premessa la distruzione della parte X dovrebbe comportare quella del tutto, ossia quella di Y. Otteniamo così una contraddizione, quindi l'ipotesi d'assurdo è falsa e non è possibile che le parti di una sostanza non siano esse stesse sostanze.[14] Il secondo argomento appare ancora più facile da formalizzare: infatti, da $\exists x A(x) \to \exists x B(x)$ e $\exists x B(x) \to \exists x C(x)$ segue immediatamente $\exists x A(x) \to \exists x C(x)$. Questi argomenti hanno quindi tutta l'apparenza non solo di essere validi, ma anche di contenere tutte e sole le premesse necessarie per derivare la conclusione. Tuttavia Aristotele, anche in questo caso, parla di "qualcosa che manca" e dice che ελλειπουσι premesse. I commentatori sono perciò per lo più inclini a pensare che anche questi argomenti siano privi di alcune premesse indispensabili e che l'apparente necessità della conclusione dipenda dall'ovvietà delle premesse mancanti. In realtà, da un lato non appare chiaro perché allora Aristotele abbia distinto i due casi, dall'altro trovare che cosa manca in questi argomenti, soprattutto nel secondo, non sembra facile. Ma anche per quel che riguarda il primo, proposte esplicative come quella di Alessandro, per il quale (cfr. *In Analyticorum Priorum librum I Commentarium*, CAG II, 1, 347.5–7) la premessa mancante è che un tutto è composto dalle sue parti (in effetti la conclusione parla di parti, mentre la seconda premessa di ciò di cui una cosa è composta), a nostro avviso non sono troppo convincenti. D'altra parte la mancanza può riferirsi a una deficienza formale, mentre il verbo ελλειπειν può indicare generica-

---

[13] O per lo meno marginali (cfr. nel già citato II 1 l'osservazione che alcuni modi hanno più di una conclusione perché la conclusione standard può essere convertita).

[14] Per una ricostruzione che utilizza la logica del primo ordine cfr. Mignucci 2002.

mente un'insufficienza strutturale delle premesse, non necessariamente la loro mancanza (nell'Index di Bonitz la prima accezione di questo verbo è *deficere, inferiorem esse*).

Sia come sia, Aristotele prescrive (47a35 e sgg.) di assumere le due premesse e di analizzarle in maniera tale da formare con esse una coppia sillogistica da cui derivare la conclusione data. Purtroppo si dilunga in una serie di prescrizioni (già abbondantemente trattate nei capitoli precedenti) su come formare questa coppia, mentre non ci dice praticamente nulla sulla relazione tra le nuove premesse e quelle originali, in particolare se il contenuto delle prime è equipollente o meno a quello delle seconde. La vaghezza di Aristotele è comprensibile: postulare l'esistenza di nuove premesse equipollenti da cui dedurre la conclusione data non è che una *petitio*,[15] mentre prescrivere la ricerca di premesse non equipollenti non è che un'applicazione di quanto già osservato in I 23, e ha perciò come risultato non che ogni derivazione logicamente valida può essere ridotta alla forma sillogistica, ma solo che ogni conclusione può essere derivata sillogisticamente. Bisogna quindi giudicare poco convincente il tentativo di Aristotele di mostrare che, dopo tutto, è possibile ridurre al seguire sillogisticamente anche quel seguire di necessità che in prima istanza appare invece irriducibile.

Resta il fatto che Aristotele vuole convincerci dell'equivalenza estensionale tra la nozione intuitiva di seguire di necessità e quella formalizzata di seguire sillogisticamente. Più che a Tarski Aristotele appare perciò idealmente vicino a Frege. È vero, infatti, che in una derivazione sillogistica valida la sostituzione uniforme dei termini occorrenti in essa mantiene la validità, ma questa è solo una condizione necessaria che non definisce in alcun modo né la nozione di seguire sillogisticamente né quella intuitiva di seguire di necessità. Invece Aristotele è vicino a Frege nel progetto di costruire, su basi puramente sintattiche, un sistema formale nel quale la struttura dimostrativa sia del tutto trasparente e le premesse di una derivazione siano tutte e sole quelle necessarie per ottenere la conclusione su basi puramente logiche e senza ricorso all'intuizione; anche se, rispetto

---

[15] Non basta infatti assumere che ogni contenuto sia esprimibile nelle forme proposizionali del quadrato, ma bisogna anche supporre che se qualcosa segue di necessità da premesse in forma non standard segua sillogisticamente dalle premesse in forma standard a esse equivalenti; ma questo è esattamente quello che doveva essere dimostrato.

a Frege, appare più consapevole della natura problematica del rapporto tra nozione intuitiva e nozione formalizzata di seguire logicamente.

2. L'impostazione di Bolzano

Si dice comunemente che Bernard Bolzano sia stato il primo a fornire una trattazione adeguata della nozione di seguire logicamente, o derivabilità.[16] O, per meglio dire, il primo a fornire una trattazione di questa nozione da un lato senza ridurla a una qualche nozione sintattica di derivabilità in base a regole; dall'altro prescindendo dall'uso di concetti modali considerati primitivi. In effetti nelle definizioni medioevali di *consequentia formalis* (*grosso modo* il condizionale logicamente valido) è spesso in opera un criterio sostituzionale che ricorda quello di Bolzano, ad esempio in quella di Buridano:

> Conseguenza formale significa che vale per tutti i termini mantenendo la forma comune. Ossia una conseguenza formale è quella che, per ogni proposizione simile quanto alla forma che può essere formulata, ci sarà una *bona consequentia.*[17]

Tuttavia Buridano non prescinde dal ricorso ai concetti modali e applica il criterio della sostituibilità solo quando già risulta impossibile che le premesse siano vere e la conclusione falsa, ma resta ancora da stabilire se si tratta o meno di una *consequentia formalis* (ad esempio "La rosa è rossa, quindi è colorata" supera il test modale, ma non quello della sostituibilità, e quindi non è una *consequentia formalis*).

Ma torniamo a Bolzano. Lo strumento logico fondamentale di cui si serve è il metodo della variazione, che consiste nel considerare tutti i possibili risultati della sostituzione uniforme di idee con altre idee all'interno di un'idea o di una proposizione. Bisogna innanzitutto osservare

---

[16] Il termine usato da Bolzano è "*Ableitbarkeit*", la cui traduzione più corretta è appunto "derivabilità". "Conseguenza" rimanda piuttosto a "*Abfolge*", che è invece una nozione più ristretta, che ricorda piuttosto il sillogismo del '*dioti*', ossia quello in cui il medio è il fondamento scientifico dell'appartenenza del predicato al soggetto nella conclusione.

[17] Cfr. Hubert 1976: 22–23. Per tutta la questione cfr. Gabbay e Woods 2008: cap. 8, sez. 3.

che le idee non sono contenuti mentali e che le proposizioni non sono *item* linguistici (a differenza che in Buridano),[18] ma che in entrambi i casi si tratta di entità astratte che ricordano i sensi di Frege (in particolare esiste un parallelismo abbastanza stretto tra i pensieri di quest'ultimo e le proposizioni di Bolzano). Quindi, detto *en passant*, la critica secondo cui questo metodo è insufficiente a causa dei limiti del nostro linguaggio non appare del tutto fondata, dal momento che Bolzano parla di idee e non di termini e che non è necessario che tutte le idee trovino espressione nel linguaggio.[19]

La nozione fondamentale definita con questo metodo non è tuttavia quella di derivabilità, ma quella di compatibilità. Nel §154 del secondo libro della sua opera filosofica principale, *Wissenschaftslehre* (Dottrina della Scienza),[20] Bolzano definisce in questo modo la compatibilità:

> Se noi confrontiamo diverse proposizioni A, B, C, D,... e scegliamo arbitrariamente certe idee i, j,... occorrenti in queste proposizioni, si presenta la questione se possono esistere idee tali che, se sostituite [uniformemente] a i, j,..., rendono tutte le proposizioni precedenti vere nello stesso tempo. Se la risposta è affermativa chiamerò la relazione tra A, B, C, D,... una relazione di *compatibilità*, *concordanza* o *consistenza* e le proposizioni stesse *compatibili*, *concordanti* o *consistenti*.

È opportuno sottolineare alcuni aspetti di questa definizione:

(i)   Si tratta di una nozione relativa, perché decidere se un insieme di proposizioni Σ è compatibile o meno dipende dalla scelta *arbitraria* delle idee occorrenti in Σ: Σ, infatti, può essere compatibile in relazione a certe idee e non compatibile in relazione a certe altre. Ad esempio "Caio è alto" e "Caio è basso" non sono compatibili in relazione all'idea Caio, ma lo sono rispetto alle idee Alto e Basso perché possono essere sostituite con idee entrambi appartenenti all'idea Caio (ad esempio Grasso e Calvo). Questo significa che, in realtà, la compatibilità riguarda, più che le proposizioni

---

[18] Bar-Hillel (1950), allo scopo di rendere la successiva definizione di derivabilità più simile a quella che Tarski attribuirà a Bolzano, propone di sostituire il riferimento alle idee e alle proposizioni con quello ai termini e agli enunciati.

[19] Tuttavia Simons (1987) presenta un argomento diagonale che dovrebbe dimostrare l'impossibilità di un'idea distinta per ogni oggetto matematico.

[20] Bolzano 1837 Curiosamente l'opera principale di Bolzano ha lo stesso titolo di quella di Fichte (da ciò la traduzione "Dottrina della scienza").

stesse, gli schemi proposizionali: gli schemi "x è alto" e "x è basso" non sono compatibili perché non possono essere saturati entrambi dando luogo a proposizioni vere; "Caio è x" e "Caio è y" sono invece compatibili perché c'è modo di saturarli ottenendo proposizioni vere.

(ii)   Tutte le proposizioni vere sono compatibili perché sono le idee stesse che vi compaiono (considerate sostitute di loro stesse) che le rendono simultaneamente vere.

(iii) Da (ii) segue che possono essere compatibili in relazione a un insieme I di idee anche proposizioni che non condividono la totalità delle idee appartenenti a I: infatti per (ii) un insieme qualunque di proposizioni vere è compatibile in relazione all'insieme di tutte le idee che vi compaiono, ma in generale proposizioni diverse non hanno in comune esattamente le stesse idee.

(iv)   Le proposizioni logicamente contraddittorie sono definite dall'incompatibilità rispetto a ogni insieme di idee. Se invece la contraddizione dipende dal significato delle idee (come nell'esempio di (i)) esiste un insieme di idee rispetto alle quali le proposizioni sono compatibili.

Bolzano può ora definire la nozione di derivabilità (§155):

> Se esiste una relazione tra le proposizioni compatibili A, B, C, D, . . . , M, N, O, . . . tale che tutte le idee che rendono vere le proposizioni A, B, C, D, . . . se sostituite a i, j, . . . rendono vere anche alcune delle proposizioni M, N, O, . . . , allora diremo che M, N, O, . . . sono derivabili da A, B, C, D, . . .[21]

Come avevamo anticipato, la derivabilità è dunque ricondotta alla compatibilità. Ne consegue, innanzitutto, che sia la derivabilità sia la derivabilità esatta (cfr. sotto) sono, per (i), relativizzate a un insieme di idee variabili. Per avere invece una nozione assoluta bisognerebbe individuare un tipo di idee variabili e definire la derivabilità assoluta come derivabilità relativamente a tutte le idee di questo tipo che compaiono nella derivazione stessa. La cosa più ovvia sarebbe distinguere tra idee logiche e non logiche e definire la variabilità assoluta come variabilità rispetto alle prime. In effetti, nel §148 Bolzano ammette l'opportunità di distinguere tra i concetti che appartengono alla logica e quelli che non

---

[21] Detto *en passant*, il fatto di non rendere necessariamente vere tutte le proposizioni M, N, O, . . . ricorda il calcolo classico dei sequenti.

le appartengono, in modo tale da poter definire l'analiticità in termini di invarianza rispetto a tutte le idee non logiche; ma, ancora più di quanto farà Tarski, riconosce la difficoltà di tracciare una netta distinzione tra ciò che è logico e ciò che non lo è e finisce, di fatto, per non applicare una tale distinzione.

In secondo luogo, la regola *ex falso* (o meglio *impossibili*) *quodlibet* non è valida, dal momento che, per (iv), una proposizione logicamente contraddittoria non è compatibile con nessun insieme di idee; e quindi neppure la derivabilità di Bolzano coincide estensionalmente con la conseguenza nella logica classica. La derivabilità è dunque non monotona, nel senso che in alcuni casi non è possibile aggiungere premesse (ad esempio, se la premessa aggiunta contraddice logicamente quelle date); ma in altri casi sembra possibile farlo. Infatti, se le proposizioni aggiunte all'insieme delle premesse non contengono le idee 'variabili' i, j,... (in base a (iii) questo non inficia la compatibilità), allora la conclusione segue anche dall'insieme delle premesse così ampliato. Bolzano, tuttavia, ha anche una nozione più ristretta di *derivabilità*, quella di *derivabilità esatta* o *esattamente proporzionata* (cfr. il punto (26) del §155). In pratica una conclusione M è esattamente derivabile da un insieme di premesse A, B, C, D,... (sempre in relazione a un insieme I di idee 'variabili') se non è possibile omettere una delle premesse o uno dei loro componenti[22] senza che M cessi di essere derivabile. È quindi ovvio che non è possibile aggiungere premesse a una derivazione esatta mantenendone nello stesso tempo l'esattezza. Questo avvicina la derivabilità esatta alla nozione di derivabilità propria delle logiche rilevanti: la questione è dibattuta,[23] ma noi non ce ne occuperemo ulteriormente, limitandoci a sottolineare che la nozione bolzaniana di derivabilità differisce comunque da quella classica.

Allo stesso modo sarebbe abbastanza futile discutere punto per punto in cosa sono simili e in cosa differiscono la definizione di derivabilità di Bolzano e quella di conseguenza logica di Tarski. Quello che vogliamo sottolineare è, lo ripetiamo, il fatto che Bolzano abbia, in anticipo su Tarski, fornito una definizione che prescinde totalmente dall'uso dei con-

---

[22] Ad esempio, da "Ogni x è y" e "Ogni y è z e u" è derivabile (in relazione a x, y, z e u) "Ogni x è z", ma in questo caso il termine u può essere omesso: non si tratta perciò di una derivazione esatta.

[23] Cfr. per una discussione Siebel 2002.

cetti modali; e che, a differenza di Tarski e, in un certo senso, anche di Aristotele, non sembra essersi posto il problema dell'adeguatezza della sua definizione di derivabilità rispetto all'intuizione pre-teorica secondo cui ogni conseguenza logica valida, oltre che formale, deve anche essere necessaria.

3. La definizione tarskiana

La memoria tarskiana del 1936, una autorevole e influente pietra miliare della tradizione analitica, è rimasta sostanzialmente al riparo da critiche e contestazioni per decenni fino a quando nel 1990 John Etchemendy pubblicò *The Concept of Logical Consequence*.[24] Come è noto, il saggio di Tarski non è di natura *tecnica*, ma si impegna nel tentativo di fornire una definizione precisa e adeguata, un'*analisi*, di una nozione intuitiva fondamentale per il pensiero scientifico in generale. Le conclusioni cui arrivò Etchemendy sono però distruttive: l'analisi tarskiana è un completo fallimento poiché è inadeguata *estensionalmente*, e poiché anche quando riesce a restituire porzioni della nostra nozione intuitiva di conseguenza lo fa comunque per ragioni sbagliate. Sia quel che sia, l'intervento di Etchemendy ha avuto il merito di ravvivare fortemente l'interesse per la memoria tarskiana e da allora la letteratura pertinente è cresciuta moltissimo.[25] Le questioni in discussione sono molte, da quella concernente la variabilità dei domini delle interpretazioni considerate al trattamento fallace delle nozioni modali coinvolte: in queste pagine, tuttavia, ci vogliamo limitare a qualche considerazione di tipo più *marginale* su alcuni passaggi del testo tarskiano.[26]

---

[24] I temi del volume erano stati in parte anticipati da alcuni articoli dello stesso autore pubblicati nel corso degli anni Ottanta.

[25] Qui ci limitiamo a ricordare due testi ricchi di indicazioni bibliografiche: Sher 1991 e Gomez-Torrente 1996.

[26] Su alcuni di questi temi, peraltro, gli autori di queste pagine sono già intervenuti in Moriconi e Mariani 1997.

3.1   Il rifiuto delle spiegazioni sintattiche

Tarski inizia il saggio (§0.1) proponendosi di rendere preciso il concetto di 'seguire per logica' in modo conforme all'*everyday 'pre-existing' way it is used*. Tuttavia, nel corso della definizione non fa riferimento al linguaggio naturale, ma, seppure in maniera implicita, a un'ampia classe di linguaggi *formalizzati* di ordine non precisato. È stato osservato, sulla base dell'esame dei suoi testi tecnici dello stesso periodo, che Tarski si riferisce fondamentalmente a una teoria dei tipi semplici, dove ci sono infinite variabili per ogni tipo. Le *regole di formazione* hanno le doverose restrizioni di tipo; gli *assiomi*, oltre a quelli ovvii, contengono l'assioma di comprensione per ogni tipo, l'assioma di estensionalità per ogni tipo, e un assioma di infinito che garantisce l'esistenza di infiniti oggetti del tipo più basso (gli individui). Le *regole di inferenza* sono quelle ovvie. Come s'è detto, tuttavia, Tarski non esplicita il tipo di linguaggio (o meglio: il tipo di teoria) cui fa riferimento e conduce le sue considerazioni a un livello del tutto *informale*. Lo spunto iniziale (§§1.1.1-6) è costituito dall'insoddisfazione per la trattazione sintattica della nozione in questione, e il riferimento polemico è chiaramente, data anche la sede della conferenza, all'impostazione neopositivista in generale e carnapiana in particolare; cioè, al tentativo di caratterizzare precisamente ed esaustivamente la nozione in questione attraverso quella di derivabilità entro una teoria formale (un numero finito di applicazioni di assiomi e di regole di derivazione). Tuttavia, è opportuno ricordare[27] che forse il riferimento è anche a se stesso, allo studio dei sistemi deduttivi sviluppato da Tarski tra la fine degli anni Venti e l'inizio degli anni Trenta, e nei quali la nozione di 'seguire per logica' riceveva un trattamento sintattico, in termini di teoria della dimostrazione.[28] Dato un sistema formale **S**, con sistema di assiomi **A** e di regole **R**, l'insieme delle conseguenze logiche in **S** di un insieme di enunciati X è il più piccolo insieme di enunciati ben formati del linguaggio di **S** che include X e **A** ed è chiuso sotto le regole di **R**. Su questa base, era poi proposta una caratterizzazione delle varie nozioni *metalogiche*: enunciato logicamente vero, enunciati logicamente equivalenti, assiomatizzabilità di un insieme di enunciati, non-contraddittorietà, completezza, ecc. A questo proposito, può essere opportuno ricordare

---

[27] Come viene fatto in Sher 1991.

[28] Il riferimento principale è a Tarski 1930, 1930a.

che una analoga mappa concettuale era stata sviluppata da Rudolf Carnap in un testo solo recentemente pubblicato, il cui sottotitolo, molto significativamente, era *Metalogik*;[29] un testo che era comunque circolato ed era stato uno dei principali punti di riferimento dello stesso Kurt Gödel[30] al tempo della preparazione della sua tesi sulla completezza semantica delle teorie del primo ordine (il cosiddetto "calcolo funzionale ristretto"). Si può ipotizzare che negli anni immediatamente successivi, l'esistenza, segnalata dalla memoria gödeliana del 1931, di uno scarto fra nozioni sintattiche (dimostrabilità) e semantiche (verità),[31] nel caso di teorie dell'aritmetica, spinse Tarski a ritenere che il quadro offerto con gli strumenti forniti dalla teoria della dimostrazione fosse insufficiente per esprimere adeguatamente la nozione di conseguenza logica. Forte delle sue epocali indagini sul *Concetto di verità nei linguaggi formalizzati*,[32] sviluppate all'inizio degli anni Trenta, Tarski ritenne che il quadro concettuale adatto per catturare le nostre intuizioni alla base della nozione informale, pre-teorica, di conseguenza logica era di tipo semantico: quelle intuizioni, infatti, riguardavano le connessioni tra gli elementi linguistici, o teorici,[33] e gli oggetti e le strutturazioni di oggetti presenti in quelle che possiamo, in maniera neutrale, chiamare configurazioni del mondo. L'ipotesi è dunque che all'inizio degli anni Trenta Tarski sia giunto alla convinzione dell'impossibilità di procedere entro il quadro teorico offerto dalla teoria della dimostrazione poiché non era possibile definire la conseguenza logica (e i concetti connessi, a cominciare da quello di verità *logica*) senza passare prima attraverso la caratterizzazione della nozione stessa di *verità*. Cioè, nella definizione *modellistica* proposta da Tarski il concetto di conseguenza logica, e quello correlato di verità logica, sono definiti in relazione al concetto di verità sotto una interpretazione, cioè

---

[29] Carnap 2000.

[30] Ed era probabilmente noto anche a Tarski. Si veda per una prima informazione Goldfarb 2005.

[31] Per un primo esame di questa complessa situazione si può vedere Moriconi 2014.

[32] Pubblicate in polacco nel 1933, da lui stesso tradotte in tedesco e pubblicate nel 1935. La traduzione inglese è disponibile in Tarski 1956.

[33] C'è effettivamente un'ambiguità nell'uso che Tarski fa del termine "linguaggio", spesso intendendo anche ciò che più correttamente, e usualmente, si intende parlando di "teorie".

al concetto di verità *materiale* (specifica di un dominio, non-formale).[34]

3.2   Il legame con l'intuizione pre-teorica

Questa costellazione concettuale compare nel testo tarskiano (§§1.3.1–
1.4.4) in un certo senso *in disguise*, attraverso la considerazione della
possibilità di teorie ω-incomplete dell'aritmetica; possibilità determinata
dai risultati gödeliani del 1931, e indagata autonomamente dallo stesso
Tarski.[35] Ora, l'utilizzazione di tali teorie per sostenere che:

> the formalized concept of following, which until now was generally used in the
> construction of deductive theories, by no means coincides with the everyday
> concept. (§1.2.3)

costituisce uno dei punti più enigmatici di questo pur problematico te-
sto tarskiano.[36] Facendo (evidentemente) riferimento al modello inte-
so, o standard, $\boldsymbol{N}$ dell'aritmetica di Peano,[37] Tarski considera che in
questo contesto vale ovviamente che se $\boldsymbol{N} \vDash \{\mathrm{A}(\mathbf{n}) \mid n \in Nat\}$, do-
ve $\mathbf{n}$ è il numerale per n, allora $\boldsymbol{N} \vDash \forall x\mathrm{A}(x)$. Da ciò segue quindi
che $\{\mathrm{A}(\mathbf{n}) \mid n \in Nat\} \vDash \forall x\mathrm{A}(x)$. In questo caso, in sostanza, è fat-
to rientrare tra le costanti logiche il quantificatore "per ogni numero
naturale". Tuttavia, per il primo teorema di incompletezza di Gödel,
$\{\mathrm{A}(\mathbf{n}) \mid n \in Nat\} \nvdash \forall x A(x)$. La sfasatura tra i due livelli, e quindi
l'*inadeguatezza* del livello derivazionale sintattico, è evidente; meno evi-
denti, però, sono i termini del ragionamento tarskiano. Molteplici sono
stati i tentativi di individuare il particolare sistema di teoria dei tipi cui
Tarski poteva far riferimento. Tuttavia, ferma restando l'inderivabilità

---

[34] In questo legame con la verità è forse da vedere la base reale della critica di
Etchemendy, incentrata sul cosiddetto *Principio di Riduzione*. Secondo tale princi-
pio la *verità logica* di un enunciato è equiparata da Tarski (1936) alla *verità tout
court* della chiusura universale della funzione enunciativa associata all'enunciato in
questione. È dalla combinazione fra questo Principio e la concezione *interpretazionale*
attribuita a Tarski che, per Etchemendy, derivano fondamentalmente le inadeguatezze
della caratterizzazione tarskiana.

[35] Ad es. in Tarski 1933.

[36] Questo tema è trattato da quasi tutti i testi che si sono occupati della memoria
tarskiana. Particolarmente attenti a esso sono Sagüillo 1997 e Bay 2001.

[37] È *la* struttura il cui dominio è l'insieme *Nat* dei naturali, dove la costante indivi-
duale $\mathbf{0}$, la funzione s di successore, le operazioni $+$ e $\bullet$ di addizione e moltiplicazione,
e la relazione $=$ di identità hanno il significato usuale.

sintattica sanzionata dal risultato gödeliano del 1931, si è concluso che, pur pensando che l'ω-regola[38] valesse nella sua versione della teoria dei tipi semplice, Tarski non poteva ritenerla logicamente valida: la definizione che egli darà nel §2.6.1, infatti, fa riferimento a una molteplicità di interpretazioni (ancorché, probabilmente, relative a uno stesso dominio). E in questo caso, variando la denotazione attribuita a qualcuno degli elementi designati nella struttura standard $\boldsymbol{N}$, non è difficile far sì che A($\mathbf{n}$) sia vero di ogni n, ma non sia invece vero $\forall x$A(x). Ma allora, ci si può chiedere, qual è il significato dell'esempio fatto da Tarski? Inoltre, Tarski dedica ben quattro pagine di questo peraltro non lungo saggio a discutere la possibilità di ottenere – tramite l'aritmetizzazione della metateoria – versioni sintatticamente dominabili della regola infinitaria, con le quali riuscire a catturare

> the 'essential' content of the concept of following, which has by no means been exhausted by the rules used until now, (§1.4.1)

per poi concludere bruscamente che in ogni caso il risultato di Gödel sta lì a ricordarci che

> no matter how we enrich the stock of rules of inference – we shall be able to construct sentences which follow in the everyday sense from the theorems of the deductive theory under consideration, but which cannot be proven in this theory on the basis of the accepted rules. (§1.4.2)

Senza voler contribuire a arricchire ulteriormente il panorama delle interpretazioni del testo tarskiano, crediamo che un dato da tenere presente sia il fatto che la definizione modellistica cui Tarski puntava sarà fornita solo molto più avanti, nel §2.6.1. Pensiamo che sia quindi inappropriato chiedersi se la regola infinitaria stabilisca per Tarski una relazione di conseguenza logica ragionando sulla base di *quella* definizione. Per sua stessa ammissione, egli fa riferimento all'*everyday concept*. E indubbiamente, dal punto di vista delle *everyday intuitions*, se tutti gli enunciati A($\mathbf{n}$), per $n \in Nat$, sono veri, allora anche l'enunciato $\forall x$A(x) deve essere vero. In maniera molto semplice, possiamo allora avanzare l'ipotesi che il significato di questo esempio iniziale fatto da Tarski sia fondamentalmente di tipo metodologico: ci dice quale direzione sia da

---

[38] È la regola che per l'appunto permette di passare dalle infinite premesse A($\mathbf{n}$), per $n \in Nat$, alla conclusione $\forall x$A(x).

prendere, e quale no, [i]n order to obtain the proper concept of following, essentially close to the everyday concept (§1.4.3). E questa direzione è quella che guarda all'apparato concettuale semantico, ai fili che collegano gli elementi linguistici con gli oggetti e gli stati di cose presenti in quelle che abbiamo chiamato configurazioni del mondo. Ciò detto, è tuttavia opportuno ricordare che, in chiusura del §1.4, Tarski sente il dovere di avvertire che la trattazione del problema in termini di teoria della dimostrazione

> by no means loses its importance: this concept, possibly widened with the help of new rules of inference, will probably always play a decisive role in practice, in the construction of deductive theories, as an instrument which allows one to prove or refute individual sentences of the theories being constructed. (§1.4.4)

anche se subito ribadisce che, comunque,

> it seems on the other hand that one should put the proper concept of following in the foreground in considerations of a general theoretical character.

3.3   La caratterizzazione semantica

Nel §2.2.1 Tarski procede

> to construct a formally correct and materially adequate definition of the concept of following for an extensive category of formalized languages.

Analizzando il contenuto intuitivo della nozione di 'seguire per logica', Tarski formula due condizioni ritenute imprescindibili per poter dire in maniera *formalmente corretta e materialmente adeguata* che un dato enunciato X segue logicamente da una classe di enunciati K:[39]

(M) *Non può accadere* che tutti gli enunciati della classe K siano veri e X sia falso

e

(R) Questa relazione di conseguenza non può andare persa *rimpiazzando* le designazioni degli oggetti cui si fa riferimento in questi enunciati con la designazione di altri oggetti.

---

[39] Anche in questo caso, non intendiamo esaminare i temi e i passi della definizione tarskiana in maniera esaustiva e analitica. Esistono già molti testi in cui questa operazione è fatta e a essi rimandiamo senz'altro. Ci limitiamo a qualche osservazione su aspetti senz'altro marginali, ma, crediamo, non trascurabili del testo tarskiano.

La prima condizione esprime il vincolo modale, secondo Tarski inestricabilmente proprio della nostra nozione intuitiva: se tutti gli enunciati della classe K sono veri, anche X *deve* essere vero. La seconda esprime la caratteristica 'logica'; cioè, precisa Tarski, *formale*. Sembra, cioè, che nel concetto di conseguenza logica Tarski scomponga i due elementi: il seguire *e* la logica. "Formale" vuol dire che la relazione del 'seguire' deve essere completamente determinata dalla forma degli enunciati coinvolti. Se solo la forma conta, 'seguire' non può dipendere dalla conoscenza del mondo esterno. In particolare, dalla nostra conoscenza degli oggetti di cui si parla negli enunciati coinvolti; il che vuol dire che il 'seguire' è un nesso che mantiene la sua validità rimpiazzando i nomi di questi oggetti con nomi di altri oggetti.[40] È d'uso notare che (R) pone un limite a (M): non tutte le conseguenze *necessarie* sono logiche. Lo sono solo quelle in cui la relazione di conseguenza fra gli enunciati della classe K e l'enunciato X si fonda su relazioni *formali* fra gli enunciati in questione. Gli esempi non mancano: "Il piano A non è completamente blu" è una conseguenza necessaria, ma non formale (e quindi non logica), di "Il piano A è completamente rosso". Non è formale perché dipende dalla 'grammatica dei colori', la quale a sua volta dipende dalle caratteristiche identificative degli oggetti di cui si parla. Rimpiazzando "blu" con "liscio", infatti, il rapporto di conseguenza cade: nulla esclude che un piano sia al contempo rosso e liscio.[41]

Il passo successivo della costruzione tarskiana è costituito dalla proposta – nel §2.3.5 – di un principio in cui *dovrebbero* trovare espressione entrambe le precedenti condizioni (M) e (R). Si tratta del principio (F) – per *Folgerung* – con il quale viene tacitamente pagato il debito alla concezione bolzaniana, decretandone nello stesso tempo l'insufficienza teorica:

---

[40] Da notare che in questa occasione si parla solo di oggetti e non di predicati. Tuttavia, poiché sta idealmente lavorando in una qualche variante della logica del secondo ordine (o anche di ordine superiore), Tarski intende probabilmente comprendere fra gli 'oggetti' anche le proprietà e relazioni 'designate' dai simboli predicativi e relazionali del linguaggio in questione. Sempre in nota, ci limitiamo a segnalare che naturalmente tra le caratteristiche di fondo dell'approccio tarskiano c'è il fatto che si assume come data la struttura logica degli enunciati, e quindi in particolare la decisione su quali segni debbano essere considerati *logici* e quali *non-logici*.

[41] Cfr. Sher 1991: 43.

If in the sentences of the class K and in the sentence X we replace the constant terms which are not general-logical terms correspondingly by arbitrary other constant terms (where we replace equiform constants everywhere by equiform constants) and in this way we obtain a new class of sentences K′ and a new sentence X′, then the sentence X′ must be true if only all sentences of the class K′ are true.

Come è noto, Tarski osserva subito che la condizione (F) è necessaria, ma non sufficiente: la condizione potrebbe infatti essere soddisfatta solo perché il linguaggio della teoria formalizzata non contiene abbastanza termini costanti. E aggiunge, nel §2.4.3, che (F) sarebbe *anche* sufficiente a garantire che X segua formalmente dagli enunciati della classe K se si assumesse che il linguaggio che stiamo considerando contiene in se stesso i nomi di tutti i possibili oggetti; un'assunzione che, aggiunge subito, non potrà mai essere realizzata.

Questi rilievi tarskiani sono abbastanza sorprendenti e rivelativi. Cioè, ci fanno capire che pur volendo analizzare l'*everyday concept*, Tarski in realtà ragiona con riferimento al *linguaggio di una teoria formalizzata*, il cui dominio di riferimento è fissato e il cui insieme di costanti non logiche ha già ricevuto un'interpretazione. Ed è opportuno, a questo proposito, ricordare quanto aveva detto pochi anni prima nel *Wahrheitsbegriff*:

Resta forse da aggiungere che qui non siamo interessati a linguaggi e scienze "formali", nel senso speciale di scienze cui non è attribuito alcun significato. Per scienze di questo tipo il problema qui discusso non ha alcun rilievo, non è neanche dotato di significato. Noi attribuiremo sempre significati del tutto concreti e, per noi, comprensibili ai segni che occorrono nei linguaggi che considereremo. [...] Gli enunciati che sono scelti come assiomi ci sembrano materialmente veri, e nello scegliere le regole di inferenza siamo sempre guidati dal principio che applicando tali regole a enunciati veri dobbiamo sempre ottenere enunciati che sono ancora veri. (Tarski 1956: 166–167)

Quelli che Tarski considera sono dunque linguaggi formalizzati e non linguaggi formali. Senza tener conto di ciò non si capisce bene la critica della condizione (F): le risorse espressive del linguaggio? Se il riferimento fosse stato al linguaggio naturale – come poteva far pensare il voler analizzare l'*everyday concept* di conseguenza logica – che senso poteva infatti avere parlare dei rimpiazzamenti "disponibili nel linguaggio"? Vogliamo dire: quando si considera la condizione (F) l'intuizione vorrebbe che si debba controllare se *ogni* sistematico rimpiazzamento delle designazioni degli oggetti potrebbe disturbare lo status del rapporto di conseguenza

in questione. Invece si procede tenendo conto del fatto che è possibile controllare soltanto i rimpiazzamenti che possono essere fatti usando le risorse espressive del linguaggio in cui si sta lavorando. Ed è sintomatico di questo cambiamento di prospettiva il fatto che quando, nel §1.2.3, difendeva il carattere *ovvio* del rapporto di conseguenza fra gli infiniti esempi e la corrispondente asserzione universale, diceva che quel rapporto sembra indubitabile dal punto di vista dell'*intuizione quotidiana*. Quando parla dell'insufficienza della condizione (F), nel §2.4.4, fa invece riferimento a un non meglio precisato, ma dato, *linguaggio formalizzato*.

Nella formulazione della condizione (F) per Tarski sono in questione tutti i possibili rimpiazzamenti delle costanti extra-logiche occorrenti negli enunciati coinvolti con le altre costanti presenti nel linguaggio. Quando formulerà la sua propria proposta *semantica*, invece, il riferimento sarà a tutte le possibili *reinterpretazioni* delle costanti extra-logiche in questione. Le nozioni chiave della formulazione semantica, quelle che permettono di superare le difficoltà della condizione (F),[42] saranno le nozioni di 'soddisfazione' e di 'modello'.[43] Parlando di modelli ci si rende indipendenti dalla maggiore o minore ricchezza del linguaggio scelto come oggetto di indagine perché si considerano tutte le possibili interpretazioni dei simboli non-logici del linguaggio. Per far questo bisogna percorrere un *detour* più complicato rispetto a quello previsto dal precedente rimpiazzamento 'simboli costanti/altri simboli costanti', con riferimento alle costanti extra-logiche presenti nel linguaggio della teoria formalizzata in questione (in un certo senso, possiamo quindi definirlo *intra-linguistico*). La nuova procedura prevede che le costanti extra-logiche presenti negli enunciati coinvolti siano rimpiazzate da variabili di tipo corrispondente (che si assume siano disponibili nel linguaggio). Ciò trasforma l'enunciato in una *funzione enunciativa*. A questo punto si considerano tutte le possibili interpretazioni di queste variabili in un modello, cioè tutti i possibili modi di ottenere *enunciati* da quella funzione enunciativa. E

---

[42] E di eliminare il coinvolgimento della modalità presente nell'iniziale condizione (M).

[43] Tralasciando ovviamente la possibilità di avere "nomi per tutti i *possibili* oggetti", è in questo modo possibile ovviare alla possibile povertà di nomi presenti nel linguaggio della teoria. Grazie al concetto di successione si fornisce uno strumento per produrre un nome per ogni oggetto del dominio. Date infatti una successione σ e una variabile individuale x, se a è un oggetto del dominio per il quale la teoria non ha un nome, ecco allora un nome per a: $\sigma[x := a]$.

in questi termini è possibile fornire la cercata definizione di "seguire per logica":

<div align="center">

l'enunciato X segue logicamente dagli enunciati della classe K

se e soltanto se

ogni modello di K è anche modello di X.

</div>

In un certo senso, tuttavia, è come se nel corso della trattazione cambiasse il modo di considerare il linguaggio. Inizialmente, s'è visto, il riferimento è a un linguaggio *formalizzato* dotato di un determinato insieme di costanti extra-logiche, già interpretate relativamente a un dominio assunto e fissato. Questo vale fino alla formulazione della condizione (F). Poi si passa, in qualche modo, a un linguaggio formale dotato di insiemi di variabili di vario tipo in attesa di trovare un'interpretazione in vari modelli. Ma il cambiamento della situazione non è del tutto acquisito, nel senso che il quadro di riferimento resta fondamentalmente sempre quello di un linguaggio formalizzato, dal quale poi ci si allontana per rendere la relazione di seguire logicamente indipendente dalla nostra 'conoscenza del modo esterno', dalla nostra conoscenza degli oggetti di cui si parla negli enunciati in questione. Cioè, per l'appunto, dalla loro interpretazione già assicurata dal fatto di lavorare con un linguaggio formalizzato. Per poter formulare la sua nuova proposta, quella "semantica", Tarski opera la sostituzione 'costanti/variabili', e quindi 'enunciati/funzioni enunciative', ma senza esplicitamente passare al quadro di riferimento tipico di un linguaggio formale.[44] Questo fatto forse spiega la sua convinzione[45] che se tutti i termini extra-logici fossero considerati logici allora si avrebbe un collasso della conseguenza logica su quella materiale. Cioè, se le costanti extra-logiche di un *dato* linguaggio *formalizzato* sono considerate

---

[44] Le differenze di fondo tra questa nozione tarskiana di modello e quella poi diventata usuale, indipendentemente dalla questione della variabilità dei domini, sono ovvie. In quest'ultima, infatti, cade il passaggio attraverso le *funzioni enunciative*, mentre il linguaggio originale degli enunciati è considerato non-interpretato; cioè *formale*, ma in una diversa accezione del termine. Quando passiamo da un modello all'altro, variamo semplicemente l'interpretazione del nostro linguaggio non-interpretato. Ciò comporta che, appropriatamente, dire che un certo enunciato X è vero in un modello $\boldsymbol{M}$ per il Tarski di questo periodo significa dire che la funzione enunciativa X' ottenuta da X rimpiazzando termini con variabili è *soddisfatta* da un'appropriata sequenza di oggetti di $\boldsymbol{M}$.

[45] Altrimenti errata, cfr. al proposito Sher 1991: 45–46.

alla stregua di quelle logiche, nel senso che il loro significato è fissato una volta per tutte, allora *relativamente a quel linguaggio* l'implicazione materiale fra due enunciati A e B è anche logica. Per l'assunto, infatti, o A è falsa o B è vera e visto che il loro significato è fissato questa situazione non può essere cambiata; cioè, non è possibile avere A vera e B falsa.

Riferimenti bibliografici

Anderson, A.R. e Belnap, N. (1967), "The pure calculus of entailment", *Journal of Symbolic Logic*, 27, pp. 19–52.

Bar-Hillel, Y. (1950), "Bolzano's propositional logic", *Archiv für Mathematische Logik und Grundlagenforschung*, 52, pp. 65–98.

Barnes, J. (2007), *Truth etc.: Six Lectures on Ancient Logic*, Clarendon Press, Oxford.

Bay, T. (2001), "On Tarski on models", *Journal of Symbolic Logic*, 66, pp. 1701–1726.

Bolzano, B. (1837), *Wissenschaftlehre*, Seidel, Sulzbach, trad. ing. parziale *Theory of Science*, a cura di J. Berg, D. Reidel, Dordrecht 1973.

Carnap, R. (2000), *Untersuchungen zur allgemeinen Axiomatik*, a cura di T. Bonk e J. Mosterin, Wissenschaftliche Buchgesellschaft, Darmstadt.

Etchemendy, J. (1990), *The Concept of Logical Consequence*, Harvard University Press, Cambridge, MA.

Gabbay, D. e Woods, J. (a cura di) (2008), *Handbook of History of Logic*, volume II: Mediaeval and Renaissance Logic, Elsevier, Amsterdam.

Goldfarb, W. (2005), "On Gödel's way", *Bulletin of Symbolic Logic*, 11, pp. 185–193.

Gomez-Torrente, M. (1996), "Tarski on logical consequence", *Notre Dame Journal of Formal Logic*, 37, pp. 125–151.

Hubert, H. (1976), *Iohannis Buridani Tractatus de Consequentiis*, Peeters, Louvain.

Łukasiewicz, J. (1957), *Aristotle's Syllogistic from the Standpoint of Modern Formal Logic*, Clarendon Press, Oxford.

Mignucci, M. (2002), "Syllogism and deduction in Aristotle's logic", in M. Canto e M. Sperber (a cura di), *Le Style de la Pensée: Recueil de Textes en Hommage à Jacques Brunschwig*, Les Belles Lettres, Paris, pp. 244–266.

Moriconi, E. (2014), "La verità in matematica", *La Matematica nella Società e nella Cultura – Rivista dell'Unione Matematica Italiana*, VII, pp. 417–443.

Moriconi, E. e Mariani, M. (1997), "Etchemendy on logical truth", *Epistemologia*, XX, pp. 267–296.

Moriconi, E. e Napoli, E. (1987), "Dimostrazioni e significato", *Lingua e Stile*, XXII, pp. 153–178.

——— (1988), "Dummett's transcendence", *Philosophia*, 18, pp. 371–383.

Sagüillo, J.M. (1997), "Logical consequence revisited", *Bulletin of Symbolic Logic*, 3, pp. 216–241.

Sher, G. (1991), *The Bounds of Logic*, MIT Press, Cambridge, MA.

Siebel, M. (2002), "Bolzano's concept of consequence", *Monist*, 85, pp. 581–601.

Simons, P. (1987), "Bolzano, Tarski, and the limits of logic", *Philosophia Naturalis*, 24, pp. 378–405.

Tarski, A. (1930), "Fundamentale Begriffe der Metodologie der deduktiven Wissenschaften. I", *Monatshefte für Matematik und Physik*, 37, pp. 361–404, trad. inglese in Tarski 1956: pp. 60–109.

——— (1930a), "Über einige fundamentalen Begriffe der Metamathematik", *Comptes Rendus de Séances de la Société des Sciences et des Lettres de Varsavie*, 23, pp. 22–29, trad. inglese in Tarski 1956: pp. 30–37.

——— (1933), "Einige Betrachtungen über die Begriff der ω-Widerspruchsfreiheit und der ω-Vollständigkeit", *Monatshefte für Matematik und Physik*, 40, pp. 97–112, trad. inglese in Tarski 1956: pp. 279–295.

——— (1936), "Über den Begriff der logischen Folgerung", *Actes du Congrès International de Philosophie Scientifique, VII Logique*, Hermann & Cie, Paris, pp. 1–11, trad. inglese in Tarski 1956: pp. 408–420, traduzione inglese rivista come Tarski 2002.

——— (1956), *Logic, Semantics, Metamathematics*, Clarendon Press, Oxford, II edizione a cura di J. Corcoran, Hackett, Indianapolis 1983.

——— (2002), "On the concept of following logically", *History and Philosophy of Logic*, 23, pp. 155–196.

Part III
Syntactic structure and logical analysis

Symmetry and asymmetry in the definition of syntactic operations: Merge and Labels

Caterina Donati

1. Introduction

A large part of theoretical research on syntax of the past forty years has been devoted to the search for the definition of what is or what are the most basic syntactic operations. While it is clear that syntax is a *combinatorial* component of language, the exact nature and status of this combination, in relation in particular to the *transformational* component of syntax, is still under discussion. This kind of deep general issues is what animated the weekly drives we shared with Ernesto on the roads connecting Bologna with Urbino (which eventually led to a joint seminar on coordination in the remote year 1999). Ernesto is not a syntactician, of course, but he knows how to think about language, ask the good questions and raise the worse objections. I hope to answer some of those in this paper.

The paper is organized as follows.[1] Section 2 reviews the set of assumptions that go under the name of the X-bar theory, underlining its advantages and limits, in particular in relation to coordination. Sections 3 to 5 discuss the concept of Merge in relation to symmetry (section 3), linearization (section 4) and labelling (section 5). Sections 6–8 discuss some possible labelling algorithms and their potential problems. Section 9 and 10 raise the issue of the possibility of labelling conflicts (section 9) and of label-less syntactic objects (section 10). Section 11 goes back to coordination and draws some conclusions.

One proviso before starting. For many years the starting point of the research on the nature of core syntactic operations has been the assump-
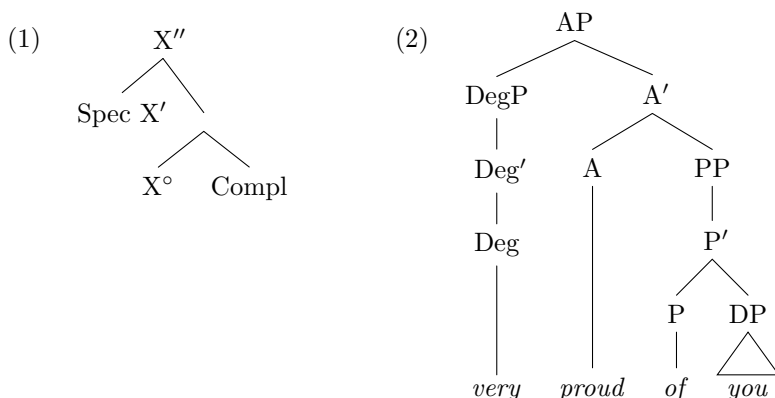
---

[1] This paper is largely based on the second chapter of the monograph written with Carlo Cecchetto (Cecchetto and Donati 2015) and should thus be considered as largely amenable to both authors.

tion of syntax as interfacing with the lexicon, a repository of stored lexical items, defined as bundles of features and handled as such by syntax. Any combinatorial component included in lexical derivation was ignored by syntacticians as belonging to a separate component, namely morphology. More recently, this lexicalist starting point has been severely criticized, and many now push the idea of a "syntax all the way down" (Halle and Maranz 1993), with the same combinatorial rule being responsible both for the formation of phrases and complex structures and for that of words (see Embick and Noyer 2007 and references cited therein). For sake of simplicity, and because we want to focus here on the nature of properly syntactic objects such as phrases, we will largely ignore this recent development and the related controversy (see Cecchetto and Donati 2015 for an overview) and simply assume that syntax interfaces with the lexicon.

2. The X-bar schema

The Government and Binding approach to the lexicon/syntax interface (cf. Chomsky 1981) assumed two ingredients as primitives: the Projection Principle (Chomsky 1986), namely the idea that properties of words (their features) survive or project into the syntactic structure; and the X-bar schema (Chomsky 1970, Jackendoff 1977), namely the idea that this projection is filtered by a rigid, universal and asymmetrical structure, reproduced in (1) and illustrated in (2) with the AP (adjectival phrase) *very proud of you*.

(1)   X″   (2)   AP

         Spec   X′   DegP   A′

                  X°   Compl   Deg′   A   PP

                                                   Deg   P′

                                                                       P   DP

                                                   very   proud   of   you

By (1) each word heads a phrase, projecting its category (i.e. N, V, P, A etc.; Deg (degree), A, P and D in example (2)) and other features (e.g. phi features such as number, gender etc.) up to at most two levels, and each phrase is headed by a word. The schema is a primitive and a universal, and provides the atom of any syntactic structure. The only variable dimension is the order of head and complement, which is supposed to vary cross-linguistically (by the Head Parameter: Stowell 1981, Travis 1984).

This proposal has many descriptive and explanatory advantages: first, the schema posits a single discrete model for any syntactic structure, allowing immediately for the infinite recursion property of natural languages. This advantage has become more and more salient in the field when the X-bar schema was extended to areas that were initially not included under its scope, such as the clause (analysed as a phrase headed by the complementizer, CP: Chomsky 1986), or when the subject came to be analysed as part of the verb phrase, as its specifier (the so-called VP-internal subject hypothesis: McCloskey 1997). Second, it posits a fundamental asymmetry in phrase structure, imposing an element as the head. Third, it allows for important typological predictions, concerning word order variation ant its overall systematic character (see Biberauer et al. 2010 for recent developments and discussion).

The X-bar schema has been however more recently challenged for a number of reasons. One of them has to do with the issue of symmetry and asymmetry: (1) imposes that any phrase is asymmetrically headed, and is as such a good model for complementation and subordination, but what about coordination, or what traditional grammars call parataxis? Clearly, if (1) is the only possible atom of any syntactic structure, coordination is supposed to be an illusion, not involving a symmetrical structure but rather some asymmetry: a coordination as (3) must thus be analysed as a Conjoined Phrase, with the first conjunct in its specifier and the second as the complement of the conjunction head.

(3)   *three dogs and four cats*
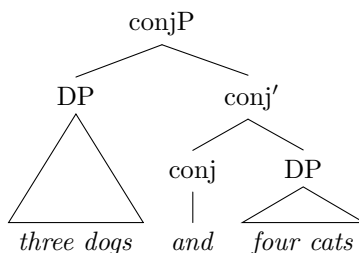
A great amount of interesting research has been produced given these premises singling out the existence of important asymmetries between the two conjuncts, including agreement and binding asymmetries (see a.o. Munn 1993, Kayne 1994, Zhang 2010 for an overview). A simple illustration of such asymmetries is given in (4). It shows that the first conjunct can be an antecedent of a pronoun in the second conjunct, while the second conjunct cannot be the antecedent of a pronoun in the first conjunct.

(4)   a.   Every man$_i$ and his$_i$ dog left.
      b.   *His$_i$ dog and every man$_i$ left.

Still, crucial aspects of coordination seem to escape this forced unification. In particular, it is rather clear that coordination and juxtaposition are simpler than subordination (see Roeper 2011 for a recent formalisation from the point of view of first language acquisition), and that it is not always the case that the two conjuncts behave asymmetrically (see Borsley 2005 for a critical discussion). How to account for this simplicity if coordination is just subordination in disguise? We shall go back to this important issue.

More theory internally, the X-bar theory did not resist the Minimalist stand that was inaugurated in the Nineties by Chomsky (1995) as a tentative to direct research into the most simple and minimal theory of syntactic computation. In particular, the X-bar schema violates what has been called the Inclusiveness Condition (Chomsky 1995: 225), namely the requirement that syntax only deals with lexical entries and their properties, without adding any other object during the computation. The schema in (1) pre-exists any lexical entry, and acts as a filter between the lexicon and the syntactic computation. Bar levels are particularly problematic. What are they? They are not lexical properties in any obvious sense but they display a specific syntax, so under the Inclusiveness Condition they should be discarded.

The rigidity of (1) imposes moreover a number of vacuous projections, since a word cannot enter a syntactic derivation without projecting a phrase structure. This implies a great departure from the minimal assumptions: even elements that really look like words, like, say, "what", or "he", must be analysed as projecting a phrase, which is nevertheless invisible and empty in most cases, as illustrated in (5).

(5)   D″
          |
          D′
          |
          D
          |
          the
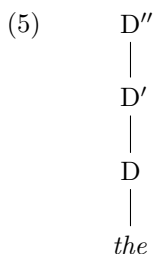
Much of the research into the theory of syntax in the past 15 years can be described as an attempt to get rid of the X-bar schema while keeping some of its results, and in particular the Projection Principle and the asymmetry in phrase structure between the 'leading element' (the head) and the categories with which it combines. We now turn to this research.

3. Merge and symmetry

The Minimalist Program, with its sake of simplicity and quest for reduction of theoretical assumptions, abandoned the X-bar schema as a stipulated primitive and went back to reason about the very notion of combination necessary to account for the syntax of natural languages. The output is the notion of Merge, defined as the simplest operation combining items.

The simplest definition of Merge is that of 'combine', a binary operation putting two elements together. This is a clear *symmetrical* operation that does not yield any prevalence of one element over the other. Merge yields as an output something as simple as $\{\alpha, \beta\}$. Still, it is able to derive the fundamental characteristics of recursion, since it may apply to its own output.

There is overwhelming evidence that a symmetric output is not desirable, at least as far as subordination and complementation are concerned (we will go back to coordination). When two elements are merged, typically one takes priority over the other. This is what underlies the notion of head and projection described under the X-bar schema. A syntactic object is typically hierarchically organized, with one of the two merged

categories determining the distributional and interpretive properties of the object created by Merge. So, what is needed in addition to Merge is something forcing one of the two elements to project, or, in newer terms, to provide a label to the newly formed syntactic object, which will look like $\{\alpha\,\{\alpha,\beta\}\}$. Some asymmetry needs to be built up into Merge or into its output. This is the aspect that we call the labelling problem, on which we will get back extensively. However, there is another aspect of the necessarily asymmetric output of Merge, that we might call linearization. We briefly deal with it in the next section.

4. Merge and linearization

Kayne (1994) proposed an axiom, the Linear Correspondence Axiom (LCA), that determines how purely hierarchical syntactic structures get linearized, that has been very influential in the syntactic theorizing of the last twenty years. In a nutshell, LCA translates asymmetric c-command into linear precedence:[2] if α c-commands β, α precedes β. Crucially, what is needed for linearization is an asymmetric syntactic structure: it is at odds with a purely symmetric operation as Merge.

Many proposals in recent years have tried to reconcile LCA with the symmetric nature of Merge. The basic idea is that some asymmetry needs to be imposed to syntactic structures for linearization to be possible.

The weakest form of this idea posits this asymmetry requirement as a condition on Phonetic Form (PF): symmetric syntactic objects (namely structures that c-command each other) can be generated by syntax. However, they must be destroyed before PF through movement operations, whose output will be an asymmetric structure where one object asymmetrically c-commands the other (cf. Moro's 2000 Dynamic Antisymmetry, but see also Barrie 2011, and 'rolling movement' as in Guimarães 2000, a.o.). This weak antisymmetric approach is attractive in that it is com-

---

[2] C-command is a relation between nodes and can be defined as follows.

A node A *c-commands* node B if and only if:

(i)   A does not dominate B,
(ii)   B does not dominate A, and
(iii) The first branching node that dominates A also dominates B.

patible with a minimal definition of Merge as an operation yielding symmetric syntactic objects, relegating asymmetry to a requirement holding at the PF interface, when sequential strings are needed. However, we see at least two problems. The first is that this approach presupposes that Merge is symmetric but movement introduces an asymmetry. However, under the most consequential minimalist view, movement is just another instance of Merge, one in which α is merged with a category β that is contained within α, so technically movement is Internal Merge. If movement is Internal Merge and Merge is symmetric, movement cannot have a salvific symmetry-breaking role. Furthermore, even if the unification of Merge and movement is not accepted, the weak antisymmetric approach can solve only one part of the symmetry problem, since it does not help with the labelling aspect of it. In principle, labelling determination has nothing to do with linearization: linearization does not tell how we get a label from an operation as simple as Merge.

The stronger version of the Linearization proposal fares better from this point of view: it simply stipulates that syntactic objects must be asymmetric, and forces Merge to only apply to non-symmetric pairs of objects.

All in all, the approach to linearization based on LCA cannot be easily combined with the simplest notion of Merge, unless extra-assumptions are made either about the status of linearization as an interface phenomenon or about its asymmetrical nature. Let us now turn to the other facet of the symmetry problem, the one concerning labelling.

5. Merge and labels

The intuition below the notion of label is that a group of words retains some of the properties of one (and only one) of the words that make up the group. This asymmetry in the output of Merge plays a role in many components of grammar. It plays a role at the interfaces. At the PF interface, if word order is a matter of postcyclic linearization, the notion of head may play a crucial role. But also at the LF interface, semantic selection is sensitive to the notion of label, for example. Even more centrally, labels are usually considered to be needed in the syntactic computation. In the standard view, the one that was implemented in

the Projection Principle we discussed earlier, a syntactic derivation is driven step by step by lexical properties. At the first step, these are directly encoded in the word itself, but the other steps of the derivation appear to be no different. In general, what a label does is introducing a partitioning among words and their features, into two categories: those that survive and remain active in the next step of the derivation, and those which do not.[3] More explicitly, we can define a label as in (6).

(6)   **Labels.** When two objects α and β are merged, a subset of the features of either α or β becomes the label of the syntactic object $\{\alpha, \beta\}$. A label:

   (i) can trigger further computation,
   (ii) is visible from outside the syntactic object $\{\alpha, \beta\}$.

Given the definition in (6), Merge can remain symmetric and minimal. Furthermore, the Inclusiveness Condition is respected since the label is no new object to be inserted in the derivation. On the contrary, a label is a subset of the features that are already present in the derivation. For example, if a verb is merged with a direct object, some feature of the verb (typically its categorial feature) will become the label of the newly formed syntactic object. However, now there is a burden on the computation. In fact, given α and β, two syntactic objects merged together, something in the computation must be able to select one of them as the survivor, the label. This has to be done efficiently, by minimal search. The quest for the proper algorithm(s) able to achieve this label determination for each syntactic object has been at the heart of much debate in recent years. We now turn to this.

6. The Head Algorithm

Chomsky (2008) proposes a labelling algorithm that we will call the "Head Algorithm", given in (7).

---

[3] Collins (2002) sketches a theory of syntax where the notion of label can be dispensed with. However what he tries to eliminate is really the notion of label as an extra object distinct from the two items that are merged, as in X-bar theory. This is not different from what is sketched here. In his system as well, what we call "label" cannot be dispensed with, and he calls it "locus".

(7)   In {H, α}, H a lexical item (LI), H is the label.

This algorithm has a strong intuitive basis: if there is an asymmetry between the two elements that are merged, and one of them is structurally simpler, then it is automatic for the system to distinguish the two elements and select one. By minimal search, the one that is structurally simpler will be the label. This is the lexical item, the word.

  This algorithm is also consistent with a lexicalist point of view, where words are distinct from phrases by being syntactically atomic. However, it is possible to show that the Head Algorithm is too restrictive, since it does not provide the computational system with an automatic device for labelling all core cases of syntactic objects created by Merge. While we might expect labelling to be not always univocal, leaving some work to the interfaces, with (7) alone we would have too much indeterminacy. Let us see some cases of undesirable indeterminacy in details.

  First of all, a system working with one and only one algorithm as (7) would have nothing to say about the very first step of any derivation, when two words get merged, as in (8).

(8)   {saw, John}

This would give us a weird grammar, in which any computation automatically runs at least two parallel derivations given any pair of words, depending on which word provides the label. Chomsky (2005) acknowledges this problem but claims that a multiple spell out system ensures that the 'wrong' derivation will crash early enough. Still, the system would introduce the computational burden of maintaining two parallel derivations up to the next higher phase even in trivial cases like (8), which are not temporarily ambiguous in any reasonable sense.

  A more problematic case systematically arising in a system containing only (7) is illustrated in (9), a configuration where the external argument is merged with VP.

(9)   *the boys saw the cat*
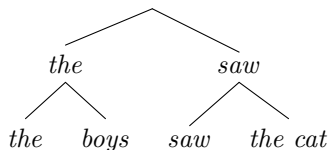
In (9) two objects are merged but none of them is a word: (7) might be taken to mean that the resulting object does not have any label, but this is clearly an unwanted result. Alternatively, a system that has (7) as its only labelling algorithm might be taken to mean that labelling cannot be decided in such cases. This is equally unsatisfactory. In order to treat cases like (9), Chomsky (2013) adopts an idea underlying Moro's (2000) Dynamic Antisymmetry: (9) is a case of undesired symmetry, so, it must be destroyed by some kind of movement. However, we see several problems with adopting this idea in the form proposed by Chomsky (2013). First, the proposal is patently countercyclic. In fact, if what breaks symmetry and creates a label is movement, the symmetric structure lacking a label (say, Spec, VP) will remain in place until the landing site of movement (say, Spec, TP) becomes available. But, in order for the derivation to reach the point where Spec, TP is created, the structure without a label (Spec, VP) should be selected. However, this is impossible under the definition of label given in (6) (a label selects *and is selected*).

Another problem is that the movement that breaks the symmetry at the Spec, VP level creates another point of symmetry at the Spec, TP level. After all, even the configuration created by movement of the external argument is a case in which two objects are merged but none of them is a word. So the Head Algorithm in (7) cannot decide the label.

An alternative that Chomsky considers is the same we have discussed for First Merge (of two words), as in (8): either category in (9) can label the syntactic object, but the interface will filter out the incorrect labelling. But this proposal would be dramatically inconsistent with the very idea of having a labelling algorithm: if the algorithm is effective in only one configuration ({word, syntactic object}) out of the three configurations Merge can generate (the other two being {word, word}; {syntactic object, syntactic object}), it is basically of little use.

7. The Movement Algorithm

This quick review of some representative cases of Merge clearly shows that a system that contains only (7) as a labelling algorithm is unsatisfactory. Chomsky (2008) proposed a second algorithm, adding (10) to (7).

(10) If α is internally merged to β forming α, β, then the label of β becomes the label of $\{\alpha, \beta\}$.

The algorithm in (10) is meant to avoid any problem with labels of internally merged, and thus moved, structures. It ensures that "in all movement operations it is always the target that projects", a very standard assumption in generative grammar of the past 40 years. A configuration immediately captured by the algorithm in (10) is (11).

(11) *which cat did you see*
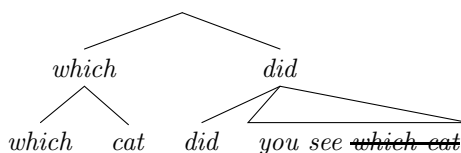
Given X-bar theory, which assumed a rigid structure pre-existing movement and lexical insertion, the condition holding that the target always projects was obtained for free. Movement was assumed to be 'substitution' of an empty node with the displaced element. So, the label for the structure was already given and movement was predicted not to have any effect on such a structure. This is illustrated in (12).

(12)
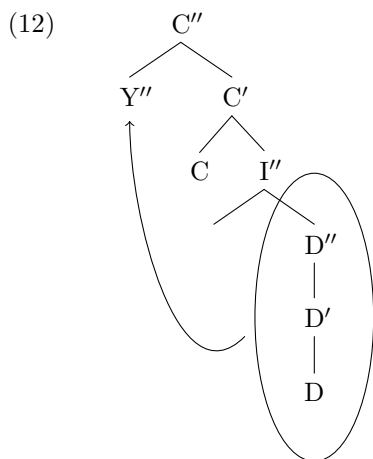
In a minimalist framework, where there is no structure pre-existing syntactic operations and the only operation responsible for movement *and* for structure building is Merge, there is simply no place for the notion of substitution.

The algorithm in (10) actually stipulates that External Merge is different from Internal Merge, an *ad hoc* residual of a movement theory. Also, it does not solve the 'specifier problem' illustrated with (9) and discussed above. When two complex syntactic objects are merged together and none is moved, neither the head algorithm nor the movement algorithm can provide this syntactic object with a label.

8. The Probing Algorithm

Suppose that Merge does not impose any asymmetry. Any application of Merge combining two distinct objects is possible. Suppose also that Merge is free, costless and replicable *ad libitum*. A consequence of this simple view seems to be the impossibility to derive a label, as we discussed in details above.

However Merge is not the only syntactic operation involved in syntactic structure creation. There is another operation, which plays a role as crucial as that of Merge, namely Agree, or, as it is also called, Probing. Probing is an operation by which an unvalued feature, call it a Probe, finds matching feature, a Goal, valuating it. Here is the definition given in Chomsky (2000: 113).

(13)   Call H a *Probe* P, which seeks a *Goal* G within XP; P=H c-commands G [...]. If the P-G relation satisfies the relevant conditions, then uninterpretable features of P, G delete.[4]

---

[4] There is an aspect of the definition in (13) that we will not keep in the rest of the paper, namely the fact that an uninterpretable feature of the Probe deletes as a consequence of the Probing operation. We think that the question of what feature is interpretable is potentially slippery, since it depends on the semantic model that is assumed. In fact, a safer notion, which is commonly used in more recent minimalist work, is the opposition between valued/unvalued features. This opposition in many cases is easier to detect. For example, there is a clear basis to say that the subject DP values the agreement features of the verb, rather than the other way around.

Probing, in contrast to Merge, is intrinsically asymmetric: it is the Probe that needs a Goal and seeks for it, not vice versa. Since probing is clearly asymmetric, it is very natural to derive the asymmetry of syntactic objects from Probing itself. In order to do so, we must assume that Probing itself is the trigger of (most) Merge operations. As for movement (Internal Merge) operations, it is not controversial that Probing plays a role, with the Probe being the (possibly indirect) trigger of movement. Although Probing can take place long-distance, it is typically associated to movement, with the Goal internally merging to the phrase headed by the Probe. But even selection in a case like (14) can be naturally interpreted as a kind of of Probing.

(14) {saw, Kitty}

Although no structural asymmetry is built in the syntactic object created by Merge in (14), there is an asymmetry in the lexical properties of the two words involved (in their features). A classical way to describe this asymmetry is to say that "Kitty" *saturates* "saw", and not vice-versa. We can frame this asymmetric relation between the two members of a merging pair in terms of a Probe-Goal relation: "saw" has an unvalued feature (a selection feature) – a Probe – which gets valued by some feature(s) of "Kitty", the Goal.

The case of External Merge of the external argument to the VP, illustrated in (9) above, does not seem fundamentally different. The category V has to assign a theta-role, therefore it can be seen as the Probe of the merging operation.

Capitalizing on this extensive interpretation of the concept of Probing, Cecchetto and Donati (2010, 2015) propose the following algorithm.

(15) **Probing Algorithm**: The label of a syntactic object $\{\alpha, \beta\}$ is the feature(s) which act(s) as a Probe of the merging operation creating $\{\alpha, \beta\}$.

The proposal in (15) is very natural, therefore it is no surprise that very similar approaches have been proposed by several authors. For example, Adger (2003: 91) reduces selection to a Probe-Goal relation and defines the head as the element which selects in any merging operation. The Probing Algorithm is also reminiscent of Pesetsky and Torrego's (2006)

Vehicle on Merge Requirement. Boeckx 2008: chap. 3 contains a detailed discussion of labelling, reaching similar conclusions.

Before getting to discuss one by one the cases that were problematic given Chomsky's (2008) labeling algorithms, let us make explicit an important consequence of what we have said so far. Remember that we assume that words exist and that they are special entities seen by syntax as atomic elements. We propose to formalize 'wordhood' in syntax by assuming that words come with a special property, which forces them to Merge with other material. If this is so, then words are intrinsic probes *qua* words and they always activate the Probing Algorithm. This means that a word can always label the syntactic object it contributes to form. With this in mind, let us now review all the possible cases we already discussed in the preceding sections.

Consider the case in (16). Here "saw" unambiguously labels the structure because (i) "saw" is a word, hence a Probe by definition, and (ii) "saw" selects (i.e. probes) "the cats".

(16)   *saw the cats*
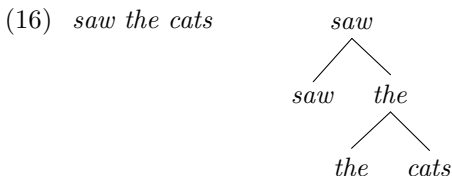
Remember that the configuration in (14) could not be satisfactorily treated by Chomsky (2008), who had to allow for parallel derivations until the first Spell-Out step because no asymmetry gave a structural label to the phrase. In (14) there is no labelling indeterminacy under the probing approach, though: while both merged elements are words (hence intrinsic probes), only "saw" is a double Probe, being also a selector. If a double Probe 'wins' over a single Probe, "saw" labels the structure, as is desirable.

The case of External Merge of the external argument to the VP, illustrated in (9) above is also not problematic: the verbal subconstituent assigns a theta role to its specifier. It is thus a Probe, and by (15) it provides the label.

9. Labelling conflicts

The system predicts some conflicts to be possible, in particular when a word (hence an intrinsic probe) merges with a probing syntactic object. A configuration instantiating this case is illustrated in (17).

(17)   *what you see*
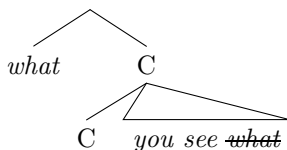
In this configuration, both the internally merged word ("what"), which is an intrinsic probe, and the probe of the movement operation (C) should provide the label given the Probing algorithm (15). If there is no clear winner, there are three logical possibilities, listed in (18).

(18)   When two probes are merged:

      a.   neither provides the label,
      b.   both provide the label,
      c.   either one or the other provides the label.

It seems that at least for the case of (17) it is the possibility (18c) that correctly describes what happens, namely that a labelling ambiguity arises: either the word labels the structure being an intrinsic Probe, or the syntactic object does, being the Probe. We refer to Cecchetto and Donati (2010), who argue that labelling ambiguity is what accounts for the dual nature of sentences like *what you see*, which can either be interpreted as embedded questions, as in (19a), or as free relatives, as in (19b).

(19)   a.   I wonder what you see here,
       b.   Picasso drew what you see here.

In (19a) the clause gets a C label and thus satisfies the selection of *wonder*, which takes only clauses as its internal argument (see (20a)); in (19b) the label comes from *what*, which is a D, and the string thus satisfies the selection requirements of a DP-selecting verb such as *draw* (20b).

(20) a.   I wonder whether you like cats / *cats,
      b.   I draw cats / *whether you like cats.

10. Label-less objects

As for possibility (18a), namely that the resulting object remains with-
out a label, this is clearly a possibility that we cannot exclude from the
model entirely: the fact that Merge typically results from a Probing op-
eration does not imply that it has to. Chomsky (2008) proposes that
Merge, either external or internal (movement), is a costless operation
that must apply freely. The approach to labelling that was presented
here can naturally explain why Merge is typically (although not obli-
gatorily) associated to Probing. In fact, the Probing Algorithm in (15)
severely constrains the application of free (= unprobed) Merge, because
according to (15), each time Merge is *not* probed, its output will have
no label. But, given (6), an object without a label has a very restricted
distribution: it cannot be selected and no further computation can take
place inside it. Given these restrictions, do label-less objects actually ex-
ist? One obvious candidate is clauses, that are very special objects, in
that they can be root structures. Given (6), if labels are needed for a
derivation to proceed (labels can trigger further computation) and feed
external Merge (through selection), when a structure is neither embed-
ded nor triggers further computation it needs no label. In Cecchetto and
Donati 2015 we explore the possibility that root clauses are label-less,
obtaining interesting results concerning classical syntactic problems such
as successive cyclic movement and islands.

11. Coordination and conclusion

As for the possibility given in (18b), namely the option that two cate-
gories can provide the label at the same time, this is perhaps what could
best account for the properties of coordination we mentioned above.[5]

---

[5] See Citko 2008 for an interesting proposal in this direction. She adopts an analy-
sis in terms of what she calls "Project Both" to capture properties of comparative
correlatives and of extended projections in the sense of Grimshaw (1991, 2005).

Typically coordination does not have a specific label: coordinated phrases have the same distribution of the two conjuncts taken separately.

(21) I see three dogs / four cats / three dogs and four cats.

In (21) the complex conjoined constituent has the same distribution of its two separate conjuncts, receives the same theta-role and satisfies the selection of the same predicate. It might be tempting to describe this state of affairs by assuming that conjoined structures are indeed the output of simple symmetric merge, with no labelling. This would not work however if we take the definition of label as in (6): no labelling would make conjoined structures syntactically inert and invisible. This does not seem to be the case. Conjoined structures are indeed active: they can be moved ((22a)–(22b)), they trigger (plural) agreement (which means that the complex phrase they form has formal features on its own: (22c)), they can be referred to by an anaphora (22d).

(22) a.   Three dogs and four cats are seen by me (passivization: A movement).
      b.   How many dogs and how many cats do you see? (wh-question: A-bar movement).
      c.   One cat and one dog are strolling around.
      d.   I see three dogs and four cats and I want to chase them.

On the other hand, it is a well-known fact about coordination that you can only conjoin phrases belonging to the same syntactic category. A possible way to capture this is by imposing that both conjuncts provide the label, which thus must be categorically non distinct.

If this is true, and there is no choice to be made in conjoined structure for determining its label, it is maybe possible to account for why coordination appears to be easier and is acquired earlier in acquisition. But this is a speculation that requires much further research. I hope that it will nevertheless be sufficient to capture the imagination of Ernesto and feed many more endless discussions with him in the future.

References

Adger, D. (2003), *Core Syntax: A Minimalist Approach*, Oxford University Press, Oxford.

Barrie, M. (2011), *Dynamic Antisymmetry and the Syntax of Noun Incorporation*, Springer, Dordrecht.

Biberauer, T., Holmberg, A. and Roberts, I. (2010), "A syntactic universal and its consequences", *Linguistic Inquiry*, 45, pp. 169–225.

Boeckx, C. (2008), *Bare Syntax*, Oxford University Press, Oxford.

Borsley, R. (2005), "Against conjP", *Lingua*, 115, pp. 461–482.

Cecchetto, C. and Donati, C. (2010), "On labeling: Principle C and head movement", *Syntax*, 13, pp. 241–278.

——— (2015), *(Re)Labeling*, MIT Press, Cambridge, MA.

Chomsky, N. (1970), "Remarks on nominalization", in A.J. Roderick and P.S. Rosenbaum (eds.), *Readings in English Transformational Grammar*, Ginn, Waltham MA, pp. 184–221.

——— (1981), *Lectures on Government and Binding*, Foris, Dordrecht.

——— (1986), *Barriers*, MIT Press, Cambridge, MA.

——— (1995), *The Minimalist Program*, MIT Press, Cambridge, MA.

——— (2000), "Beyond explanatory adequacy", in A. Belletti (ed.), *Structures and Beyond: The Cartography of Syntactic Structure*, volume 3, Oxford University Press, Oxford, pp. 104–131.

——— (2005), "Three factors in language design", *Linguistic Inquiry*, 36, pp. 1–22.

——— (2008), "On phases", in R. Freidin and M.L. Zubizarreta (eds.), *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, MIT Press, Cambridge, MA, pp. 133–166.

——— (2013), "Problems of projection", *Lingua*, 130, pp. 33–49.

Citko, B. (2008), "Missing labels: Head movement as project both", in C.B. Chang and H.J. Havnie (eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistic*, Cascadilla Proceedings Project, Somerville MA, pp. 121–128.

Collins, C. (2002), "Eliminating labels", in S.D. Epstein and D. Seely (eds.), *Derivation and Explanation in the Minimalist Program*, Blackwell, Oxford, pp. 42–64.

Embick, D. and Noyer, R. (2007), "Distributed morphology and the syntax/morphology interface", in G. Ramchand and C. Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, Oxford University Press, Oxford, pp. 289–324.

Grimshaw, J. (1991), *Extended Projection*, PhD Thesis, Brandeis University, Waltham MA.

——— (2005), "Extended projection", in *Words and Structure*, CSLI Publications, Stanford, pp. 1–74.

Guimarães, M. (2000), "In defense of vacuous projections in bare phrase structure", *University of Maryland Working Papers in Linguistics*, 9, pp. 90–115.

Halle, M. and Maranz, A. (1993), "Distributed morphology and the pieces of inflection", in K. Hale and S.J. Keyser (eds.), *The View from Building Twenty: Essays*

in Linguistics in Honor of Sylvain Bromberger, volume 20, MIT Press, Cambridge, MA, pp. 111–176.

Jackendoff, R. (1977), *X-Bar Syntax: A Study of Phrase Structure*, MIT Press, Cambridge, MA.

Kayne, R. (1994), *The Antisymmetry of Syntax*, MIT Press, Cambridge, MA.

McCloskey, J. (1997), "Subjecthood and subject positions", in L. Haegeman (ed.), *Elements of Grammar*, Kluwer, Dordrecht, pp. 197–236.

Moro, A. (2000), *Dynamic Antisymmetry*, MIT Press, Cambridge, MA.

Munn, A. (1993), *Topics in the Syntax and Semantics of Coordinate Structures*, PhD Thesis, University of Maryland.

Pesetsky, D. and Torrego, E. (2006), "Probes, goals and syntactic categories", in Y. Otsu (ed.), *Proceedings of the Seventh Tokyo Conference on Psycholinguistics*, Hituzi Syobo, Tokyo, pp. 25–60.

Roeper, T. (2011), "The acquisition of recursion: How formalism articulates the child path", *Biolinguistics*, 5, pp. 57–86.

Stowell, T. (1981), *Origins of Phrase Structure*, PhD Thesis, MIT.

Travis, L. (1984), *Parameters and the Effects of Word Order Variation*, PhD Thesis, MIT.

Zhang, N. (2010), *Coordination in Syntax*, Cambridge Studies in Linguistics Series 123, Cambridge University Press, Cambridge.

Some reflections on negation in the vernacular*

Wolfgang Künne

1. Suppose you are patient enough to answer the following (more or less silly) questions: (1) "Is snow blue?", (2) "Are all philosophers wise?", (3) "Do some Greeks love the German Minister of Finance?", and (4) "Is your cousin Anne pretty, and is her friend Betty witty?", and each time your answer is "No". Suppose furthermore that your patience is not yet exhausted: in each case you also want to tell the interrogator what your monotonous monosyllabic answer logically comes to, not more and not less. So in order to unpack your answers you use negations of the declarative sentences that correspond to the yes/no interrogative you have heard. In English there is no *systematic* way of producing such answers to questions (1) to (4). (If at this point you feel like exclaiming that you do know a uniform strategy that delivers for each and every declarative English sentence its negation, please wait and see. Ernesto Napoli is ready to disillusion you, and I shall try to come to his aid.) So how can the feat be accomplished? You can enlarge upon your two-letter answer in case (1) by sliding "not" between copula and general term, in case (2) by prefixing "not" to "all philosophers" and in case (3) by transforming "some Greeks" into "no Greek" (and adapting the grammatical

---

* Many of these reflections were triggered by reading and re-reading Ernesto Napoli's insightful and crisp paper "Negation" (Napoli 2006). Much to my surprise it passed unnoticed in Horn and Wansing 2015. As you may have guessed, by "vernacular" I mean languages like English, Italian and German, languages Frege called *Volkssprachen*. I dislike the well-liked appellation "natural languages", for *Volkssprachen* are as much the product of human labour as are so-called artificial languages.

number of the verb). In the more demanding case (4) you can unpack the "No" by saying, "Anne is not pretty, *or* Betty is not witty".

It is tempting to describe each of the non-laconic answers as formulating *the* negation of the declarative sentence that corresponds to the yes/no interrogative, but would that description be correct? Napoli seems to think so. In his (2006: 233) he claims that "*the* negation of a sentence is a sentence that contradicts it", and he spells this out as follows: "A sentence is *the* negation of another *if* and only if their conjunction is contradictory, i.e. is a logical falsity" (246).[1] I have italicized what worries me in this statement: the presumption of uniqueness and the "if" half of the biconditional. Not only one sentence is such that its conjunction with "Anne is pretty, and Betty is witty" is a logical falsehood. Apart from the disjunction that Napoli favours (245) and that I used above, there are, for example, its reversal "Betty is not witty, or Anne is not pretty" and a conjunction of two conditionals, "If Anne is pretty then Betty is not witty, and if Betty is witty then Anne is not pretty". And even in non-molecular cases like "Snow is blue" there is more than one sentence whose conjunction with "Snow is blue" is logically false. Apart from "Snow is not blue" there is, for example, "Snow is not blue, and if it is raining then it is raining", etc. Of course, the uniqueness problem also arises in formal languages. Not only "$\neg p$" stands in contradictory opposition to "p", but also, for example, "$\neg\neg\neg p$" and "$\neg p \wedge (q \to q)$".

Peter Geach noticed the problem – and swept it under the carpet: "We may speak of *the* contradictory of a proposition, since no proposition has two (non-equivalent) contradictories".[2] Referring to Geach's remark Horn and Wansing (2015: §1.6) maintain that "co-contradictories" are only "syntactically distinct" because they have "the same truth conditions". This is a startling contention. Let us rehearse once again (this time in a formal language) some contradictories of "$p \wedge q$". Admittedly,

---

[1] All unaccompanied bracketed page numbers in this paper refer to Napoli 2006.

[2] Geach 1970: 71. Readers should bear in mind that Geach makes a point of always using the term "proposition" not like his contemporaries but like the mediaevals, that is, he applies it to declarative *sentences*. Horn and Wansing (2015) use "proposition", "statement" and "sentence", sometimes two of them in one breath, as if they were just stylistic variants of "declarative sentence". Some of the philosophers on whose work they draw, e.g. Frege and Strawson, used these terms (or their German counterparts) to express three entirely different concepts, and some of the points Napoli makes in his paper cannot be made without heeding these distinctions.

"$\neg p \vee \neg q$" and "$\neg q \vee \neg p$" differ only syntactically. But I would have thought that the difference between "$\neg p \vee \neg q$", "$(p \rightarrow \neg q) \wedge (q \rightarrow \neg p)$" and "$(\neg p \vee \neg q) \wedge \neg(r \wedge \neg r)$" is not only syntactical but a difference in meaning. Sameness of truth conditions, as determined by truth tables, does not guarantee sameness of meaning.

If we face the fact that there are always many (logically equivalent) contradictories of a given sentence S, we must also concede that there is no such thing as *the negation* of S either, unless we can offer a criterion for selecting one of the co-contradictories as the negation of S. I find the following criterion reasonable: among the many sentences that contradict S, the negation of S is the one that is conceptually most parsimonious, that is, that invokes fewer concepts than the others and that invokes one concept less often than the others. Thus "$\neg p$" invokes fewer concepts than "$\neg p \wedge (q \rightarrow q)$", and it invokes one concept less often than "$\neg\neg\neg p$". Even if this works very often, it does not always work: the two disjunctive contradictories of "$p \wedge q$" are conceptually balanced, and there is no reason to call one of them rather than the other the negation of the conjunction. The situation exemplified by this pair is bound to arise if a language contains sentences that have the same meaning. (Thus if you read in the book you are to review, "*The Principles of Mathematics* were written by two authors", you can articulate your protest by writing "... not by two ..." as well as by writing "... not by 2 ...".) So in the end I also need a bracketed qualifier: no sentence has two (non-synonymous) contradictories. And like Geach I shall talk as if this justifies talking of *the* negation of a sentence. Whenever I say that a sentence S is the negation of another sentence, the reader is asked to abstract from those features of S that distinguish it from other sentences with the same meaning.[3]

2. Logicians have invented an unbreakable unary connective[4] that can be prefixed to *any* declarative sentence to deliver its negation, and in their

---

[3] One could hire Wilfrid Sellars' "dot quotes" for this purpose (without subscribing to his metalinguistic reductionism). They are explained in such a way that the following identity statement is true: The sentence •Anne is not pretty, or Betty is not witty• = the sentence •Betty is not witty, or Anne is not pretty• = the sentence •*Anne non è carina, o Betty non è buffa*• = the sentence •*Betty non è buffa, o Anne non è carina*•. Cp. Sellars 1963: 228f.

[4] "One-place sentence-forming operator on sentences" would be a more appropriate title, but since it is not very handy, I follow Napoli and the majority of philosophers

informal expository prose they have forced the long-winded expression
"It is not true (the case) that" into the same service. They are to be
congratulated for the invention, and they are to be mildly criticized for
the act of force.

Let us see how two of the greatest logicians performed this act of force.
In his late essay "*Die Verneinung*" Gottlob Frege formulates the double
negation of a geographical truth as follows:

(DN) It is not true that the Schneekoppe is not higher than the Brocken.[5]

The negation sign that is "inserted (*eingefügt*)" in the embedded sen-
tence is atomic, but Frege wisely refrains from expressing the dou-
ble negation of that sentence by (the German counterpart of) "The
Schneekoppe is not not higher than the Brocken", for that is an un-
grammatical stutter.[6] He does not put "not" in front of the internally
negated sentence either, for the resulting chain would not be (clearly)
well-formed either. I have put this point more hesitantly than Napoli
who calls the result of prefixing "not" to "Mary is happy" simply "a-
grammatical" (239). In German an utterance of, say, "*Nicht Maria ist
glücklich, sondern Joseph*", with optional stress on the first name, makes
perfectly good sense, and maybe the same holds of "Not Mary is happy,
but Joseph is". Be that as it may, a *clearly* ungrammatical sentence re-
sults, for example, if one prefixes "not" to a "There is" sentence.[7] As an
emergency solution for the annoying grammatical problem Frege prefixes

---

and logicians in using the misnomer. One would have thought that only two or more
things can be connected.

[5] Frege 1919: 148. For a line-by-line commentary on this essay see my 2010: 543–588.
I wonder what non-German readers can make of Frege's example. The Schneekoppe
(now Sněžka) is the highest mountain of the Czech Republic, the Brocken is the
mountain in Northern Germany on which the witches in Goethe's Faust have a lot of
fun.

[6] Napoli 2006: 247f. In Horn and Wansing (2015: §1.8.1) the cartoon character Homer
Jay Simpson is quoted as saying, "I am not not licking toads." Let us assume that this
is not yet deemed to be a well-formed English sentence. (A very optimistic assumption,
for in a BBC online poll in 2003 Mr Simpson was voted the greatest American of all
times, well ahead of Thomas Jefferson, Abraham Lincoln and Martin Luther King.)

[7] Italian readers of this paper need not to be told that there are languages in which
such a move does not produce garbage: the negation of "C'è birra in frigo" is "Non
c'è birra in frigo".

a string of five words to the internally negated sentence, clearly intending this 'prologue' to perform the very same (truth-value reversing) job as the inserted "not".

In his famous monograph "The Concept of Truth in Formalized Languages" Alfred Tarski says about his own use of the same 'prologue' for the same purpose:

> For stylistic reasons we sometimes use, instead of the word "not", the expression "it is not true that". In doing so we treat the whole expression as a single word, without ascribing to its parts, especially to the word "true" it contains, any independent meaning.[8]

It is clear why Tarski finds the note of warning in the second sentence very important for his project. Surely, Frege would have endorsed Tarski's remark. In his last paper "*Gedankengefüge*" he takes as his object language a *regimented* version of German. In a similarly regimented version of English for which I shall use the nickname "Loglish", the English sentence (DN) would be replaced by

(DN)^L Not (not (the Schneekoppe is higher than the Brocken)).[9]

The negation sign in (DN)^L is just a typographical variant of the small vertical stroke in Frege's *Begriffsschrift*, Peirce's and Hilbert's overline, Peano's and Russell's tilde, Łukasiewicz's and Prior's letter N, Quine's em-dash, and Heyting's and Gentzen's hook that the editors asked us to employ in our contributions to this book. Negations that begin with an atomic negation operator on sentences are foreign to ordinary English,[10] and, as Geach (1956: 75) rightly adds,

> propositional negation was as foreign to ordinary Greek as to ordinary English, and [Aristotle] never attained to a distinct conception of it. The Stoics did reach such a conception, but in doing so they violated accepted Greek usage; their use of an initial οὐχι must have read just as oddly as sentences like "Not: the sun is shining" do in English.

---

[8] Tarski 1935: §2, note 17. (Throughout this paper translations are always mine.)

[9] Frege 1923: 41 *et passim*. For a commentary on this essay see my 2010: 589–684.

[10] Following some of Frege's footsteps Napoli shows that sentences like "Not everybody is happy" only seem to falsify this contention (239–241). He does not comment on the fact that in his mother tongue "Non piove" is an immaculate sentence.

Tarski *stipulates* that the 'prologue' is to be understood as if it were as unstructured as the word "not" or as the logicians' negation signs. One could visually mark this intention by hyphenating the five words, "it-is-not-true-that". Surely, sentences beginning with this expression also "read oddly". Since it presents itself as an indissoluble unit, its word-shaped components have, as Tarski puts it, no "independent meaning (*selbständige Bedeutung*)":[11] comprehension of the whole expression is not founded upon comprehension of the words it ostensibly consists of. (When it comes to understanding the word "attic", understanding "at" and "tic" is no help.) To be sure, the typographical presence of "not" and "true" in the (implicitly or explicitly) hyphenated 'prologue' may serve as a reminder of its explanation as a truth-value reverser, but this mnemonic role does not make these components contributors to the meaning of the whole. So there is a vast difference between "it is not true that" in English and its hyphenated rewrite in Loglish, since understanding the English expression is based upon understanding "not" and "true". Furthermore, the sentence with the hyphenated 'prologue' no more contains a genuine occurrence of a that-clause than the sentence

(*)   The woman who married Socrates was Greek

contains a genuine occurrence of the sentence "Socrates was Greek" (Geach 1955: 229 and 1965: 110). In (*) the name "Socrates" goes with the words that precede it to form a complex subject term, and in the hyphenated 'prologue' the word "that" goes with the preceding string to form a unary connective. But in the English expression the word "that" goes with the succeeding sentence to form a clause. This becomes conspicuous as soon as we insert a parenthesis: "It is not true – as we all know – that snow is blue" (cp. Künne 2010: 559). You cannot insert a parenthesis into the hyphenated 'prologue', anymore than you can insert a parenthesis into the word "not" or into the hook – they all are seamless wholes.

---

[11] The phrase is Husserl's (and it is worth mentioning that Tarski's translator Leopold Blaustein had studied with Husserl), but in Husserl's mouth the phrase expresses a concept that is used to explain what is special about the meanings of syncategorematic expressions. Cf. Husserl 1913: 4th Investigation, §§4-9. That concept is of no help here, so I propose a different explanation of the phrase.

The following comparison shows, I think, that replacing the English 'prologue' in a sentence by its hyphenated counterpart changes the meaning of the sentence. Consider

(S1) It is not true that Anne left because Betty came.
(S2) It-is-not-true-that Anne left because Betty came.

In (S1) an answer to the question why Anne left is rejected as untrue. The rejected explanation is singled out by a that-clause built from a molecular sentence, so "that" goes with the molecular sentence. By contrast, (S2) in which "that" is tied to the preceding string is a denial of Anne's departure, and it explains why Anne did *not* leave. So the substitution does not at all preserve the sense of (S1).[12]

3. Consider this set of sentences (in what follows repeatedly referred to as *the Negation Quintet*):

(0) $\neg$ (Snow is blue).
(1)   Snow is not blue.
(2)   It is not true that snow is blue.
(3)   That snow is blue is not true.
(4)   The proposition that snow is blue is not true.

(0) is a substitution-instance of the logicians' schema "$\neg p$" in a variant of the Loglish Frege uses in $(DN)^L$. By contrast, (1) to (4) belong to the vernacular. Neither the hook in (0) nor the inserted "not" in (1) is in the same boat as the string of words that precedes the sentence "snow is blue" in (2). Although the victims of our elementary logic courses are hardly ever cautioned at this point, using (2) to pronounce (0) is misleading, for in the vernacular the proper parsing of (2) is before the word "that", not after it.

What is the relation between (2) and (3)? Napoli (245) takes it to be very close indeed, for he says about (a sentence like) (2) that it "can be rewritten as" (3). I entirely agree. The "it" in (2) is semantically vacuous:

---

[12] Even if one thinks that (S1), as it stands, can mean what (S2) unambiguously means one has to admit that (S2) differs semantically from (S1). If we insert in (S1) a comma between "left" and "because", the resulting sentence implies that Anne did not leave, but the resulting sentence is a *different* sentence.

it is an expletive that contributes nothing to the content of (2).[13] If you look at the translation of (2) into Italian, you see that the Italians, though not exactly famous for being taciturn, spare themselves the luxury of an expletive: "*Non è vero che la neve è azzurra*". Bernard Bolzano (whose father came from the Lago di Como) agrees with the 'Italian' verdict on the first word in (2):

> Sometimes the word "it" seems to be quite superfluous [*überflüssig*], as in the expression: "It is true that etc." For this is completely equivalent [*durchaus gleichgeltend*] to "The proposition that etc. is true". Similarly, the remark "It is fine weather today [*Es ist heute schönes Wetter*]" means the same as [*eben so viel heißt als*] "The weather is fine today etc." (Bolzano 1837: vol. II, 216, quotation marks added.)

As for the key term "superfluous", it is worth recalling that some grammarians call expletives "pleonastic pronouns". If part x of a sentence is superfluous in the Bolzanian sense then x can be deleted *salva propositione*. In this sense the adverb in "Socrates was really wise" is superfluous (Bolzano 1837: I, 123). Sometimes, after removal of the superfluous expression an inversion is required, as in the case of (2). Unfortunately, in the displayed passage Bolzano actually identifies the sense of (2) with that of (4) rather than with that of (3). But one may very well wonder whether it is his considered view that the sense of the apposited noun phrase "the proposition" is part of the sense of (2). (I shall return to this issue.)

Here is what Arthur Prior once said about the frame that surrounds "snow is blue" in (3), or rather about the frame "That _____ is not the case":

> When, instead of saying simply "Snow is not blue", we say "That snow is blue is not the case", we construct from the sentence "Snow is blue" what looks like a name ("That snow is blue") and then complete the sentence with what looks like a verb ("is not the case"). We are not, however, *very* strongly tempted to treat this name, in this sentence at any rate, as genuinely denoting an object, and its "verb" as genuinely describing an activity of this object — it is sufficiently obvious that the whole complex "That _____ is not the case" simply has the force of the adverb "not", appropriately placed. (Prior 1963: 193. I took the liberty of replacing his example by mine.)

---

[13] Mulligan (2010) prompted the correction of Künne 2003: 351 in Künne 2010a: 601 ff.

If Prior were right the ad-sentential frame "That _____ is not true" and its "the case" variant would enclose a sentence in a similar way as the ad-verbial frame "*ne* _____ *pas*" in French encloses the verb in "*Elle ne sourit pas*": the parts are not such that they contribute their own meanings to that of the scattered whole.[14] As regards the weak temptation that Prior wants us to resist, his description of those who yield to it is a caricature, or did you ever come across anyone who thought of sentences like (3) or their "the case" variants as saying of an object (a proposition, or a state of affairs, presumably) that it performs an activity called "not being true" ("not being the case")? We can let that pass as a bit of propaganda. Now compare the following strings of words:

(1*)¿ Ali is not a terrorist, and it is no longer taken to be true by the inquisitors.

(3*) That Ali is a terrorist is not true, and it is no longer taken to be true by the inquisitors.

Since the anaphoric pronoun in (1*) is looking in vain for an antecedent, (1*) is grammatical garbage, hence the depreciating suffix. By contrast, (3*) is a well-formed sentence. The first conjunct of (1*) contains a negation in the style of (1) while the first conjunct of (3*) contains a negation in the style of (3), but that difference should not make much of a difference if Prior is right. According to him, those conjuncts differ only in that the first negation sign in (1*) is inserted into a sentence whereas in (3*) it surrounds the same sentence. Allegedly, both negation signs are unary sentence-forming operators on sentences. But in (3*) the anaphoric "it" contributes to the content of the second conjunct what "that Ali is a terrorist" contributes to the content of the first conjunct. So once again we see that the word "that" goes together with the succeeding sentence to form a *clause* – rather than with "is not true (the case)" to form a unary connective, as Prior claims.

4. How are the members of the Negation Quintet related to the affirmative sentence (A)?

(A)   Snow is blue.

---

[14] Unsurprisingly, Prior says elsewhere (1967: 458, 461; 1971: 11, 19) about (2), or rather about its "the case" variant, what we now heard him say about the variant of (3).

Obviously, (0) is the negation of (A) formulated in Loglish, and (1) is the negation of (A) formulated in English. Neither (2) nor (3) nor (4) is the negation of (A). Why not? The negation of a sentence cannot thematically diverge from the negated sentence. (A) is about snow, not about a proposition. (4) is unmistakably about a proposition (and only indirectly about snow), and the same holds, though less obviously so, for (2) and (3). Sentence (2) is not the negation of (A) but of "It is true that snow is blue". The latter sentence is not only different from (A), it also expresses a different proposition. In saying this I reveal my disbelief in the redundancy account of truth.[15] As Napoli points out, (3) is not the negation of (A) but of "That snow is blue is true" (245). The latter sentence is not only different from (A), it also expresses a different proposition. Since it expresses the same proposition as (2), I reaffirm here my rejection of the redundancy account. Although Napoli explicitly only maintains that (3) is not the negation of (A), I am confident that he is ready to endorse my other non-identity claims as well. I am less confident, though, that he is happy about the *reason* I have offered for those claims. I shall return to this suspicion at the very end of this paper. At this point it may suffice if I explain my use of the term "proposition" with the help of an inference pattern: From any meaningful instance of the schema

(P)   N.N. advances / attacks / defends / embraces / endorses / evaluates / rejects the {axiom, belief, claim, dogma, hypothesis, law, principle, tenet, theorem, thesis, thought, . . . } that things are thus-and-so

you can infer the corresponding instance of

(C)   N.N. (VERB)s the proposition that things are thus-and-so.

What the general terms between the curly brackets in (P), as used in such a context, apply to are propositions. The term "proposition" in (C) is just the most general of all the general terms that can meaningfully replace the bracketed nouns in (P).

   So far I have been (almost) completely silent on sentence (4) in the Negation Quintet. It is not the negation of (A) but of "The proposition

---

[15] This is not the place to argue for the rejection of this account. Cp. Künne 2003 on Truth-Theoretical Nihilism and Künne 2014.

that snow is blue is true".[16] How is (4) related to (3)? In the same way,
I take it, as "The number 8 is not prime" is related to "8 is not prime".
In both pairs, the subject of the verbose sentence is prefixed by an ap-
positive noun-phrase that is lacking in its sparing counterpart, but in
spite of this difference the two sentences are cognitively equivalent: if
you understand both you cannot rationally accept one without immedi-
ately being ready to accept the other. But in both pairs, the sentences
do not have the same meaning, for they are not conceptually balanced:
there is a concept (*number* in the one case, *proposition* in the other) such
that only one of the two sentences contains an expression that expresses
nothing but that concept. For the same reason (4) does not have the
same meaning as any other member of the Negation Quintet.

In each of the sentences (2) to (4) lack of truth is ascribed to one and
the same proposition. Lack of truth is what is predicated by the (Fregean)
predicate "x is not true". In calling this open sentence a predicate I may
seem to contradict Napoli's contention that "the negation sign and the
predicate do not constitute a new predicate" (248). But this appearance
is deceptive, for the notion of a predicate Napoli employs in his paper
is vastly different from the Fregean account I favour.[17] Napoli's notion
allows him to assert, for example, that the phrase "the property of being
a horse" is a "proper name that acts as predicate in the sentence 'Varenne
has the property of being a horse'" (243). One may reasonably wonder
whether a notion that permits us to say this about a *singular term* really
is a notion of a *predicate.* The famous case of "Trieste is no Vienna" in
Frege 1892: 200 is entirely different, for in that sentence "Vienna" does
not play the same role as in "Vienna is a charming city", whereas in
Napoli's example the singular term plays exactly the same role as in "The
property of being a horse is multiply instantiated": it serves to identify
an (abstract) object. According to the Fregean account, Napoli's sample
sentence contains the predicates "x has y", "x has the property of being
a horse" and "Varenne has y", and a singular term is part of two of them.
It is in *this* sense of "predicate" that sentences (2), (3) and (4) contain,

---

[16] Bolzano carefully distinguished (1) and its ilk from (4) & Co. (1837: II, 16, 44 ff,
63, 269, 419).

[17] As I am not familiar with the linguistic theory on which Napoli relies at some
points in his paper, I cannot do more than register this difference.

apart from a predicate that applies to snow, a predicate that applies to propositions. Or so I contend.

What is the semantical relation between (2)–(4) on the one hand and (0) and (1) on the other? The difference between (0) and (1) is irrelevant to this issue, so let us focus on the following three biconditionals:

(2) if, and only if, (1);
(3) if, and only if, (1);
(4) if, and only if, (1).

Each of these biconditionals is true if lack of truth (in the case of a truth candidate) coincides with falsity. But if bivalence does not hold then the left-hand side of each of these biconditionals may be true while the right-hand side isn't (Künne 2010: 548f). The point to be made here could also be made if we were to assume that snow is as non-existent as luminiferous aether, but let us use an actual-world example. Suppose we follow Frege and Strawson and maintain that "Kant's wife was blonde" is neither true nor false because Immanuel never got married. Then "Kant's wife was not blonde" also falls into the truth-value gap. But "It is not true that Kant's wife was blonde" and "(The proposition) that Kant's wife was blonde is not true" are as true as can be.

5. My contention that not only in (4) but also in (3) and even in (2) lack of truth is ascribed to a proposition has to face certain substitution worries. As has often been pointed out in recent debates about the 'logical form' of ascriptions of propositional acts or states (*vulgo* attitudes), instances of "the proposition that *p*" (*propositional terms*, for short) and corresponding instances of the unadorned schema "that *p*" (*naked that-clauses*) are not always interchangeable *salva congruitate*, that is, without loss of grammaticality.[18] (Just try to squeeze the noun phrase "the proposition" into "Sophia hopes that Marcello will return".) This observation is often taken to show that naked that-clauses, unlike propositional terms, do not single out propositions.

Now this kind of substitution failure also occurs in sentences like

(2)   It is not true that snow is blue.

---

[18] You will find many references to the pertinent literature in Mulligan 2010 and Künne 2014.

Replacement of the clause by the corresponding propositional term results in grammatical garbage:

(¿)   It is not true the proposition that snow is blue.

Of course, if we insert a comma after "true", grammaticality is restored:

(+)   It is not true, the proposition that snow is blue.

But in (+) we have done (a tiny bit) more than just replace the naked that-clause by its adorned counterpart, and we have transformed the pleonastic or expletive pronoun in (2) into a cataphoric or anticipatory pronoun. (Compare "He was wise, the man who drank the hemlock".) The change of the role of the pronoun becomes more conspicuous when we look at the German translation of (+):

> Sie ist nicht wahr, die Proposition, dass Schnee blau ist.

The "*Es*" variant of this is grammatically inacceptable: because of its gender, the phrase "*die Proposition*" calls for an anticipatory "*sie*". (There is no pronoun in the Italian rendering of (+), "*Non è vera, la proposizione che la neve è azzurra*", but note that the adjective is not "*vero*" but "*vera*", as required by the gender of "*la proposizione*".) Now, does the ungrammaticality of (¿) show that the clause in (2) does not single out the proposition that snow is white? I do not think so. Since the first word in (2) is an expletive, (2) expresses the same proposition as

(3)   That snow is blue is not true,

and in (3) the naked that-clause *is* replaceable *salva congruitate* by the propositional term. So the fact that in (2) the clause is not replaceable without further ado by the term has got nothing to do with the content of (2): it is due to a semantically insignificant grammatical constraint. Consider a similar case. In "He kissed Annabella" the name cannot be replaced by "she" without violence to grammar, but this substitution does preserve well-formedness in "Annabella was kissed by him", in a sentence that expresses in the same context the same proposition as its active counterpart. So the non-exchangeability in the first sentence is due to a whimsy of grammar.

In certain embeddings sentences like (4) suffer from a substitution failure in the other direction: sometimes the propositional term cannot be replaced *salva congruitate* by a naked that-clause. Thus

> If the proposition that snow is blue is not true then snow is not blue

is grammatically kosher, while the string

> If that snow is blue is not true then snow is not blue

is not acceptable. At least it seems to be unacceptable to some keen-eared native speakers (Koster 1978, referred to, and confirmed by, Mulligan 2010: 574). Again, I do not think that this shows that the naked that-clause does not serve the same purpose as its adorned counterpart, namely to single out a proposition. It only shows that in certain positions that-clauses need a crutch. A comparison might help. Consider the sentence frame "The Hamburg composer ... was much admired in Vienna". We can insert the name "Johannes Brahms" but not the co-designative definite description "the friend of Clara Schumann".[19] The latter needs some add-on before it can enter the slot, "*who was* the friend of Clara Schumann". For all I know, no adherent of the (highly debatable) Russellian view of definite descriptions ever used this failure of *exchangeability without further ado* as evidence for the claim that "J.B." and "the friend of C.S." do not serve the same purpose, namely to pick out a certain man. Even Russellians can agree that this non-exchangeability is just a quirk of grammar. Now I think that the (alleged) non-replaceability of a propositional term by a naked that-clause in the frame "If ... is not true" deserves the same verdict. (I am confident that grammarians are good at explaining such restrictions on substitution.)[20]

6. If we mean by "negation" not the output of an operation but the operation itself then (Napoli argues) negation is not an operation on the linguistic act of assertion, and similarly, if we mean by "disjunction"

---

[19] As the adjective "co-designative" betrays, I endorse the Frege-Strawson view of definite descriptions.

[20] The reflections on ascriptions of propositional 'attitudes' in Künne 2014 may serve as a supplement to section 5.

or "conjunction" not the output of an operation but the operation itself then neither disjunction nor conjunction is an operation on two linguistic acts of assertion (234 f). In all three cases, I wholeheartedly agree – for a Fregean reason that Napoli does not give. The output of those operations can be the antecedent of a conditional, and even if the conditional is put forward as true its antecedent is not. But an account of those operations should apply across the field. Therefore the act-of-assertion view should be rejected.

In the course of his discussion of the operations disjunction and conjunction Napoli makes two questionable moves.

> Obviously any rational speaker who had asserted p or had asserted q would be ready to assert p *or* q. Yet what he would be ready to do is not what he has done. (235)

What is here declared to be obvious is not even true. The Preacher (Solomon?) who had said with assertoric force, "All is vanity", did not manifest lack of rationality by not being ready to say with assertoric force, "All is vanity, or the entropy of an isolated system that is not in equilibrium tends to increase over time". After all, the concepts of thermodynamics were not yet available to anyone when the book Koheleth was written.[21] A repair along the following lines suggests itself: a rational speaker who assertively uttered one of two declarative sentences that he fully understands would be ready to utter their disjunction assertively. Even this is not beyond doubt, but let me turn to the other move I declared to be questionable.

> For sure if anyone asserted p *and* q then he asserted p and asserted q. (235)

For sure? Frege thought otherwise, and that may be quite a good reason for not being *sure*. Frege thought that asserting that (p and q) does not include asserting that p, anymore than asking whether (p and q) includes asking whether p (Frege 1923: 37ff). He would vary Napoli's earlier remark and say: "Obviously any rational speaker who had asserted that ($p \land q$) would be ready to assert that p. Yet what he would be ready to do is not what he has done." The symbolism of the *Begriffsschrift* permits no embedding of the "judgement stroke" ("assertion stroke" would have

---

[21] For a criticism of an analogous mistake in Jerry Fodor's theory of *belief* see Künne 1995: 376f.

been a more appropriate name), so a string like "$\vdash p \wedge \vdash q$" is ill-formed. But let us return to the operation that is our main topic.

The operation of negation, Napoli says, is an operation not on acts of assertion but rather on "what is asserted" when you perform such an act (234). I have added the "when" clause because often this operation is not on what is asserted but on *what is said without commitment as to its truth-value.* Recall the Fregean point about conditionals. (Acts of assertion are to such acts of non-committal saying as mental acts of judgement are to mental acts of merely entertaining a thought.[22]) Now what is it that is asserted in an act of assertion? As you may have expected, my answer is: a proposition – in the sense that I tried to elucidate in section 4 *via* the inference pattern "(P), therefore (C)". What is Napoli's answer? Look at the schematic letters and the molecular schemata in the last-displayed quotations. At first sight, it is unclear what the two letters are placeholders for: grammar allows substitution of quotational names of sentences as well substitution of that-clauses. In the schemata, however, grammar demands that we replace the letters by sentences and then flank the result by quotation marks. This suggests that Napoli takes *sentences* to be the items on which the operation negation operates, and in his summary he explicitly states that "negation is not an operation on linguistic acts but rather an operation on the objects of linguistic acts, namely sentences" (233). I feel a bit uneasy about "objects". Admittedly, "to assert" is a transitive verb, but can anything be called an object of assertion – in the sense in which a painting may be an object of perception and of admiration? The noun phrase in "John rode his bike" is the direct object of the verb, but was the bike the object of John's riding? Well, the locution "the object of an assertion" may be less than felicitous, but we know what Napoli means by it: what is asserted. But are sentences (of whatever language) really what is asserted in certain utterances? Is the sentence "All is vanity", or rather its Hebrew counterpart, what the Preacher asserted? This sentence is what he uttered with assertoric force, to be sure, but what he asserted is that all is vanity. The sentence was the vehicle of his linguistic act, not its content, not what is specified when we answer the question, "What did the Preacher assert in the first verse of the book Koheleth?".

---

[22] Cp. the paper mentioned in the last footnote.

At the beginning of section 4, I made a rather melancholy remark to the effect that Napoli, though presumably ready to accept various non-identity claims I made, may not be ready to agree with the reasons I gave for those contentions. Those reasons invoked propositions (*Sätze an sich* or *Gedanken*, as my favourite philosophers call them), and alas! quite a few of my friends regard propositions as creatures of darkness, as *entia non grata*. Of course, the final page of this paper is not a place for addressing the serious worries behind such Quinean affronts. But I shall close by replying to a local anti-propositionalist argument Napoli gives.

All through this paper I have followed Napoli in taking sentences to be what negation operates on. I think this has many advantages, but I do not think that Frege made a mistake when (in his last two papers) he took propositions to be what is negated in a negation. By contrast, Napoli makes an attempt at refuting this view:

> Negation [...] does not apply to propositions. No doubt, natural language sentences are meaningful, still their meaning is quite irrelevant for negation. The negation of a sentence is a sentence that contradicts it. Two sentences are contradictory iff their conjunction is a logical falsity, or their disjunction is a logical truth. However, logical truth and logical falsity, unlike truth and falsity, are a matter of form, not of meaning. (233)

To be sure, under the assumption that the sentence "Snow is blue" is univocal, the question whether "Snow is blue, and snow is not blue" is a logical falsehood is not a matter of the meaning *of that sentence*. But even under the (apparently correct) univocity assumption the answer to our question very much depends on the meaning of "or" and "not". Furthermore, "Ernesto owns a bank, and Ernesto does not own a bank" is not a contradiction if the first conjunct is understood as affirming that Ernesto is an owner of a sitting accommodation, while the second conjunct is understood as denying that he is an owner of a financial institution (cp. Strawson 1957). So the announcement that the meaning of a sentence is irrelevant to the question whether it is logically false seems to be premature.[23]

References

Bolzano, B. (1837), *Wissenschaftslehre*, 4 vols., Sulzbach, Seidel (quoted by volume number and page number), translated as *Theory of Sciences*, George, R. and Rusnock, P. (eds.), 4 vols., Oxford University Press, Oxford 2014 (original pagination in the margin).

Frege, G. (1892), "Über Begriff und Gegenstand", *Vierteljahrsschrift für Wissenschaftliche Philosophie*, 16, pp. 192–205, reprinted in Frege 1967: 167–178, translated as "On concept and object" in Frege 1984: 182–194.

——— (1919), "Die Verneinung", *Beiträge zur Philosophie des Deutschen Idealismus*, 1, pp. 143–157, reprinted in Frege 1967: 362–378, translated as "Negation" in Frege 1984: 373–389.

——— (1923), "Gedankengefüge", *Beiträge zur Philosophie des Deutschen Idealismus*, 3, pp. 36–51, reprinted in Frege 1967: 378–394, translated as "Compound thoughts" in Frege 1984: 390–406.

——— (1967), *Kleine Schriften*, Georg Olms Verlag, Hildesheim.

——— (1984) *Collected Papers on Mathematics, Logic, and Philosophy*, McGuinness, B. (ed.), Basil Blackwell, Oxford.

Geach, P. (1955), "Class and concept", *Philosophical Review*, 64, pp. 561–570, reprinted in Geach 1972: 226–235.

——— (1956), "The law of excluded middle", *Aristotelian Society Supplementary Volumes*, 30, pp. 59–90, reprinted in Geach 1972: 74–87.

——— (1965), "Logical procedures and the identity of expressions", *Ratio*, 7, pp. 199–205, reprinted in in Geach 1972: 108–115.

——— (1970), "Contradictories and contraries", *Analysis*, 29, pp. 187–190, reprinted in Geach 1972: 70–74.

——— (1972), *Logic Matters*, University of California Press, Berkeley.

Horn, L.R. and Wansing, H. (2015), "Negation", in E.D. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/sum2015/entries/negation/.

Husserl, E. (1913), *Logische Untersuchungen*, Niemeyer, Tübingen, 2nd edition.

Koster, J. (1978), "Why subject sentences don't exist", in S.J. Keyser (ed.), *Recent Transformational Studies in European Languages*, MIT Press, Cambridge, MA, pp. 53–64.

Künne, W. (1995), "Some varieties of thinking: Reflections on Meinong and Fodor", *Grazer Philosophische Studien*, 50, pp. 365–395.

——— (2003), *Conceptions of Truth*, Clarendon Press, Oxford.

——— (2010), *Die Philosophische Logik Gottlob Freges*, Klostermann, Frankfurt.

——— (2010a), "Replies to Boghossian and Mulligan", *Dialectica*, 64, pp. 585–615.

——— (2014), "Frege on that-clauses", in P. Weiss (ed.), *Dummett on Analytical Philosophy*, Macmillan Palgrave, New York, pp. 135–173.

Mulligan, K. (2010), "The truth-predicate vs. the truth connective. On taking connectives seriously", *Dialectica*, 64, pp. 565–584.

Napoli, E. (2006), "Negation", *Grazer Philosophische Studien*, 72, pp. 233–252.

Prior, A.N. (1963), "Is the concept of referential opacity really necessary?", *Acta Philosophica Fennica*, 16, pp. 189–199.

——— (1967), "Negation", in *The Encyclopedia of Philosophy*, volume 5, Macmillan, London, pp. 458–463.

——— (1971), *Objects of Thoughts*, Clarendon Press, Oxford.

Sellars, W. (1963), "Abstract entities", *Review of Metaphysics*, 16, pp. 627–671, reprinted in his *Philosophical Perspectives*, Thomas, Springfield 1968, pp. 229–269.

Strawson, P.F. (1957), "Propositions, concepts, and logical truth", *Philosophical Quarterly*, 7, pp. 15–25, reprinted in his *Logico-Linguistic Papers*, Meuthen, London 1970, pp. 116–130.

Tarski, A. (1935), "Der Wahrheitsbegriff in den Formalisierten Sprachen", *Studia Philosophica*, 1, pp. 261–405, reprinted in his *Collected Papers*, Birkhäuser, Basel 1986, pp. 51–198.

La rappresentazione sintattica del credere

Pierdaniele Giaretta

Che cosa si crede? Ovvero, di che genere sono le cose che si credono? Ecco una domanda della quale Ernesto Napoli contesterebbe la legittimità nella misura in cui essa presuppone o suggerisce che il credere sia una relazione tra qualcuno e qualcosa. Senza entrare nel merito della contestazione, si può evitare di affermare che il credere sia una relazione e supporre soltanto che quando si dice che x crede che p ciò implichi che tra x e p sussista una relazione R tale che x ha la relazione R con p se e solo se x crede che p. L'ipotesi dell'esistenza di una tale relazione non è una risposta alla domanda riguardante cosa sia veramente il credere, o come si debbano correttamente intendere le attribuzioni di credenza. Certamente si può identificare il credere con una tale relazione, ma si può anche non fare questa identificazione e, ad esempio, concepire il credere come una relazione tra più entità (la *multiple relation theory of judgment* di Russell) o, al contrario, come una classe di fenomeni descritti mediante predicati complessi all'interno dei quali le posizioni di argomento hanno caratteristiche anomale (Quine).

Tuttavia, l'esistenza stessa di una relazione che sussiste tra x e p se e solo se x crede che p potrebbe essere contestata. Ad esempio, l'astrazione potrebbe non essere considerata una via percorribile per affermarla e una ragione che si potrebbe addurre, ed è stata addotta, è che non ci sarebbe nulla per cui "p" può stare che possa essere inteso come un argomento, o almeno come un argomento nello stesso senso in cui lo è Ernesto quando si dice che Ernesto è un filosofo o che tutti ammirano Ernesto. Assumendo solo come un'ipotesi l'esistenza di una tale relazione, possiamo prescindere da queste e altre obiezioni e, per semplicità di

esposizione, ci permettiamo anche di parlare di relazione di credenza e di nozioni da essa derivabili.

Specificando il primo termine di quella che per semplicità chiamiamo "relazione di credenza" si ottengono delle proprietà o, se si preferisce, dei predicati unari rispetto ai quali è stata sollevata una questione di consistenza che è apparsa rilevante per le prospettive di un certo tipo di orientamento della ricerca cognitiva, quella che ricorre all'elaborazione di modelli computazionali di classi di fenomeni mentali. Si tratta della rappresentabilità sintattica del credere, riguardo alla quale questo contributo ricorda innanzitutto i risultati rilevanti di Montague (1963) e Thomason (1980) e poi, dopo aver brevemente illustrato l'analisi di van Fraassen (2011) del risultato di Thomason, fornisce e commenta un risultato ulteriore, che è un rafforzamento di quello di Thomason poiché viene evitato il ricorso a una condizione che può essere considerata problematica.[1]

1. Tipicamente gli oggetti di credenza possono essere intesi come enunciati o come proposizioni. Riteniamo che si possa sostenere che si credono enunciati senza considerare il credere un fatto sintattico e si possa sostenere che si credono proposizioni senza escludere che del credere sia possibile una rappresentazione sintattica. La questione è stata affrontata in modo analogo, o parzialmente analogo, a come è stata affrontata la questione della esprimibilità della verità nel linguaggio di una teoria del primo ordine che comprenda l'aritmetica e quindi abbia la possibilità di esprimere proprietà e relazioni sintattiche. Presupponendo che la verità si applichi a enunciati, Tarski (1935) dimostrò che nessuna formula di una tale teoria può esprimere la verità. Tuttavia, ai fini della dimostrazione del teorema, nulla impedisce di pensare la verità come un predicato che si applica primariamente a proposizioni. La questione della natura delle cose alle quali si applica la verità diventa rilevante a livello di analisi filosofica dell'inesprimibilità della verità.

---

[1] Nella letteratura si parla di predicati sintattici nel senso di predicati che si applicano a entità linguistiche senza coinvolgimento del loro significato. La nozione è vaga, ma sufficiente per comprendere che quando si dice che una proprietà o relazione è rappresentabile sintatticamente si intende che per essa esiste un corrispondente predicato sintattico, tale che anche nel caso che la proprietà o relazione rappresentata si applichi a entità non linguistiche tale predicato si applica a corrispondenti entità linguistiche.

Montague (1963) ha posto un analogo problema di esprimibilità riguardo ai predicati modali, precisamente riguardo a predicati per i quali sono formulate certe condizioni minimali che corrispondono a quelle spesso assunte per alcuni operatori modali. Le condizioni da lui formulate sono soddisfatte sia da un concetto intuitivo di necessità che da un concetto di conoscenza riferita a un soggetto ideale che, senza limitazioni, sia in grado di dimostrare enunciati che seguono logicamente da enunciati che egli riconosce come veri. Montague assume che per una teoria T e una formula α di T, con una unica variabile libera, valgano le seguenti condizioni, dove con 'F' si deve intendere l'espressione nella teoria del gödeliano di F:[2]

$\vdash_T \alpha(\text{`}F\text{'}) \rightarrow F$;
$\vdash_T \alpha(\text{`}\alpha(\text{`}F\text{'}) \rightarrow F\text{'})$;
$\vdash_T \alpha(\text{`}F\text{'})$, dove F è un assioma logico;
Se $\vdash_T \alpha(\text{`}F \rightarrow G\text{'})$ e $\vdash_T \alpha(\text{`}F\text{'})$, allora $\vdash_T \alpha(\text{`}G\text{'})$.

T comprende, come sottoteoria, $Q^{(\beta)}$, dove Q è una teoria dell'aritmetica di Robinson, β è una formula con una unica variabile libera e $Q^{(\beta)}$ è la relativizzazione di Q a β.

Si può assumere che $Q^{(\beta)}$ abbia una formula χ come suo unico assioma e, sulla base di questa assunzione, Montague dimostra che T è inconsistente. Egli basa la sua prova sul cosiddetto lemma di diagonalizzazione, del quale dice che può essere considerato come un principio di autoriferimento. In virtù di tale lemma Montague afferma, in particolare, l'esistenza di una formula F tale che

$$\vdash_T F \leftrightarrow \alpha(\text{`}\chi \rightarrow \neg F\text{'}),$$

cioè tale da essere equivalente alla formula nella quale l'espressione formale del gödeliano della sua negazione condizionata all'unico assioma aritmetico di $Q^{(\beta)}$ sostituisce la variabile libera di α.

Nello stesso modo Thomason (1980) applica la diagonalizzazione a una teoria T, che sia analogamente estensione di $Q^{(\beta)}$, nella quale sono assunte per α le seguenti condizioni:

---

[2] Gödeliano di una formula è un numero associato alla formula mediante un metodo, ora chiamato gödelizzazione, che ha permesso a Gödel di dimostrare i suoi famosi teoremi di limitazione.

$\vdash_T \ \alpha(\text{`}F\text{'}) \to \alpha(\text{`}\alpha(\text{`}F\text{'})\text{'});$
$\vdash_T \ \alpha(\text{`}\alpha(\text{`}F\text{'}) \to F\text{'});$
$\vdash_T \ \alpha(\text{`}F\text{'}),$ dove F è un assioma logico;
$\vdash_T \ \alpha(\text{`}F \to G\text{'}) \to (\alpha(\text{`}F\text{'}) \to \alpha(\text{`}G\text{'})).$

Sulla base di tali condizioni Thomason dimostra

$\vdash_T \ \alpha(\text{`}\chi\text{'}) \to \alpha(\text{`}G\text{'})$

per ogni formula G. Se ad α si dà il significato di "essere creduto da s", dove s è un soggetto razionale per il quale è naturale assumere

$\vdash_T \ \alpha(\text{`}\chi\text{'})$
$\vdash_T \ \neg\alpha(\text{`}0 = 1\text{'})$

la teoria T risulta contraddittoria.

Thomason, che aveva modificato le condizioni su α allo scopo di poter presentare α come un predicato psicologico che potesse rappresentare la competenza cognitiva di un soggetto idealizzato, alla fine commenta: "This seems enough to show a coherent theory of idealized belief as a syntactic predicate to be problematic" (1980: 393).

Recentemente van Fraassen (2011) ha preso in considerazione quello che egli chiama "il paradosso di Thomason per la credenza" cercando di fornire una motivazione adeguata per la seconda condizione di Thomason, cioè $\vdash_T \ \alpha(\text{`}\alpha(\text{`}F\text{'}) \to F\text{'})$. Van Fraassen, che fa riferimento anche a Thomason (2011), connette la condizione di Thomason con il paradosso di Moore e ne propone una giustificazione non strettamente logica sulla base di una semantica supervalutazionista. L'analisi di van Fraassen è sottile e complessa e apparentemente porta a un rafforzamento del paradosso di Thomason poiché finisce per giustificare le condizioni aggiuntive dalle quali esso segue in virtù di alcune "naturali assunzioni". Van Fraassen non indica alcuna soluzione o via di uscita e considera il paradosso una realtà che dobbiamo accettare:

> I *cannot coherently conceive of myself* as successfully represented in the mode that Thomason criticizes. That goes for you too. Even after the modifications I introduced, Thomason's striking point remains, mutatis mutandis: if such a theory correctly represented you, and it was taught to you, and you believed it, and (constrained to be pragmatically coherent) you described the world in sentences you believed to be true, your description would be logically inconsistent. Can we live with this: that all sufficiently rich metaphysical theories that

we could write *about ourselves* will land in an indefensible and unsustainable sort of absurdity? I say yes [...]. (2011: 31)

Verosimilmente van Fraassen arriva a questa conclusione anche perché ritiene che la dimostrazione di Thomason non coinvolga l'autoriferimento. Egli dice: "This argument is remarkably minimal. Nothing involves self-reference" (18). In realtà questo è un aspetto che dovrebbe essere considerato più accuratamente.

2. Non è irragionevole pensare che l'autoriferimento sia coinvolto nella prova di Thomason tanto quanto lo è in quella di Montague, poiché in entrambe, allo stesso modo, si ricorre al lemma di diagonalizzazione e questo, come Montague dice esplicitamente, può essere considerato come un principio di autoriferimento. Forse van Fraassen non ritiene che la diagonalizzazione debba necessariamente essere intesa come autoriferimento. *Stricto sensu* gli enunciati dei quali si afferma l'esistenza in virtù della diagonalizzazione non sono autoreferenziali poiché, in base a una gödelizzazione standard, il loro numero di Gödel deve essere maggiore di qualunque numero rappresentato da un numerale che in esse occorra e, inoltre, si potrebbe contestare che la gödelizzazione sia una forma di riferimento. Tuttavia la formula F che si afferma esistere per diagonalizzazione è dimostrabilmente equivalente a una nella quale un termine rappresenta indirettamente una formula della quale F è parte. Nelle dimostrazioni di Montague e di Thomason le formule che per diagonalizzazione sono asserite equivalenti a una qualche F sono esemplificazioni di formule α sottoposte a vincoli parzialmente diversi. In entrambi i casi la dimostrazione di inconsistenza della teoria T può essere vista come una dimostrazione della non esistenza di un'interpretazione di α che (i) soddisfa i vincoli posti su α e (ii) è tale che il valore di verità di F è lo stesso della formula nella quale un termine rappresenta indirettamente una formula della quale F è parte.

Tuttavia la difficoltà di interpretare α come un predicato di credenza emerge già prima che, in seguito all'introduzione delle condizioni di Thomason, ne risulti impossibile una interpretazione che soddisfi anche gli assiomi aritmetici, in particolare il singolo assioma che si può formulare per l'aritmetica di Robinson. Un modo per fare emergere tale difficoltà consiste nel riprodurre la dimostrazione del teorema di Löb (1955), scegliendo, come caso particolare, degli enunciati nei quali un predicato

di credenza compaia nel ruolo della formula α di Thomason. Rispetto a tale predicato ci si limita a fare assunzioni che corrispondono a proprietà dimostrabili della provabilità formale che, come è noto, può essere espressa nel linguaggio dell'aritmetica del primo ordine e, in quanto tale, è oggetto specifico del teorema di Löb. In particolare, si assume di poter applicare il lemma di diagonalizzazione, dal quale la dimostrazione di Löb prende le mosse. Lo scopo finale è esaminare la possibilità di dare una rappresentazione aritmetica della cognizione, specificamente di un insieme di credenze che un soggetto razionale può avere o acquisire.

Consideriamo un linguaggio basato su un alfabeto che, oltre ai simboli aritmetici, contenga il predicato "s crede", introdotto con l'intenzione che s sia un soggetto in grado di avere credenze e acquisirne altre mediante deduzione da enunciati da lui creduti, e la formula "s crede t" sia vera se e solo se il termine t rappresenta il gödeliano di un enunciato p tale che s crede che p. Tuttavia "s" non è un termine del linguaggio. In altre parole, "s crede" deve essere considerato come una lettera predicativa della quale non si esclude a priori la possibilità di una definizione mediante nozioni del linguaggio aritmetico o in esso rappresentabili.

Immaginiamo che Ts includa l'aritmetica di Robinson e sia una teoria delle credenze che si possono attribuire a s nel linguaggio di Ts, che Ts sia nota a s e che a s siano attribuibili anche credenze che s può avere in virtù della sua conoscenza di Ts.

Assumiamo in particolare che s creda gli assiomi di Ts, anche quelli che contengono occorrenze di "s crede":

$C_1 \quad \vdash_{Ts} s$ crede $([A])$

dove A è un assioma e $[A]$ è un termine del linguaggio aritmetico che lo rappresenta, ad esempio l'espressione formale diretta, detta numerale, del suo gödeliano. Assumiamo che s sia in grado di fare le deduzioni che usando Ts si possono fare a partire da assiomi o ipotesi formulate nel linguaggio di Ts e pertanto creda gli enunciati che in questo modo si possono provare.

Tranne che per ragioni di enfasi, ometteremo nel seguito l'uso del simbolo metateorico "$\vdash_{Ts}$"; inoltre, ometteremo per semplicità di lettura le parentesi tonde in espressioni di forma "s crede $(\dots)$".

Applicando il lemma di diagonalizzazione alla formula con unica variabile libera x

$(s \text{ crede } x) \rightarrow F$

dove F è una formula chiusa, scelta a piacere, si ottiene che esiste un enunciato E tale che

D $E \leftrightarrow (s \text{ crede } [E] \rightarrow F)$

dove "$[E]$" è un termine che per diagonalizzazione rappresenta E stesso.

Possiamo assumere che s consideri il predicato "s crede" come un predicato aritmetico e quindi creda D in quanto conseguenza valida di assiomi che egli crede veri:

(1) $s \text{ crede } [E \leftrightarrow (s \text{ crede } [E] \rightarrow F)]$.

s è in grado di ragionare in modo deduttivamente valido e crede le conseguenze deduttive che deriva da ciò che crede (chiusura deduttiva delle credenze di s). Pertanto, credendo l'equivalenza D come affermato in (1), crede un membro dell'equivalenza, se crede l'altro e viceversa:

(2) $s \text{ crede } [E] \leftrightarrow s \text{ crede } [s \text{ crede } [E] \rightarrow F]$.

Il condizionale

(3) $s \text{ crede } [E] \rightarrow (s \text{ crede } [s \text{ crede } [E]] \rightarrow s \text{ crede } [F])$

può essere dimostrato nel modo seguente, dove "PC" sta per "regola di prova condizionale" e "CDs" sta per "chiusura deduttiva delle credenze di s":

(a)	$s \text{ crede } [E]$	ip.
(b)	$s \text{ crede } [s \text{ crede } [E] \rightarrow F]$	(a) e (2)
(c)	$s \text{ crede } [s \text{ crede } [E]]$	ip.
(d)	$s \text{ crede } [F]$	(b) e (c) per CDs
(e)	$s \text{ crede } [s \text{ crede } [E]] \rightarrow s \text{ crede } [F]$	(c)–(d) per PC
(f)	$s \text{ crede } [E] \rightarrow (s \text{ crede } [s \text{ crede } [E]] \rightarrow s \text{ crede } [F])$	(a)–(e) per PC.

Analogamente a Thomason e a van Fraassen, per ogni formula G assumiamo come assioma che se s crede G, s creda di credere G:

C$_2$ $\vdash_{Ts} s \text{ crede } [G] \rightarrow s \text{ crede } [s \text{ crede } [G]]$

e quindi in particolare:

(4) $s \text{ crede } [E] \rightarrow s \text{ crede } [s \text{ crede } [E]]$.

Da (3) e (4) segue:

(5) s crede $[E] \rightarrow s$ crede $[F]$.

Infatti:

(a)	s crede $[E]$	ip.
(b)	s crede $[s$ crede $[E]]$	(a) e (4)
(c)	s crede $[s$ crede $[E]] \rightarrow s$ crede $[F]$	(a) e (3)
(d)	s crede $[F]$	(b) e (c)
(e)	s crede $[E] \rightarrow s$ crede $[F]$	(a)–(d) per PC.

Assumendo che s creda anche (4), oltre che gli assiomi logici e aritmetici, abbiamo che s crede (5) per chiusura deduttiva delle credenze di s:

(6) s crede $[s$ crede $[E] \rightarrow s$ crede $[F]]$.

Assumiamo ora l'ipotesi:

A s crede $[s$ crede $[F] \rightarrow F]$.

L'antecedente "s crede $[F]$" del condizionale creduto da s per l'ipotesi A è il conseguente dell'implicazione creduta da s in base a (6); si può applicare la legge della transitività dell'implicazione, e quindi, per chiusura deduttiva delle credenze di s, ottenere:

   s crede $[s$ crede $[E] \rightarrow F]$

che per (2) è equivalente a:

   s crede $[E]$

da cui per (5)

   s crede $[F]$.

Dunque, imitando il ragionamento che porta alla dimostrazione del teorema di Löb, in modo simile a quanto fatto da Raymond Smullyan,[3] abbiamo dimostrato

(I) s crede $[s$ crede $[F] \rightarrow F] \rightarrow s$ crede $[F]$.

---

[3] Raymond Smullyan è un logico di grande valore e anche un brillante autore di libri che presentano in modo intuitivo, e spesso divertente, risultati fondamentali della logica. Il libro al quale si fa riferimento qui è: *Forever Undecided. A Puzzle Guide to Gödel* (1987).

Viceversa,

(a) s crede $[F]$ ip.
(b) $F \rightarrow (s$ crede $[F] \rightarrow F)$ assioma logico
(c) s crede $[F \rightarrow (s$ crede $[F] \rightarrow F)]$ per C$_1$
(d) s crede $[s$ crede $[F] \rightarrow F]$ (a) e (c) per CDs.

Dunque:

(II) s crede $[F] \rightarrow s$ crede $[s$ crede $[F] \rightarrow F]$.

Da (I) e (II) segue

S s crede $[s$ crede $[F] \rightarrow F] \leftrightarrow s$ crede $[F]$.

S è una equivalenza dimostrabile in Ts. Un membro dell'equivalenza è dimostrabile in Ts se solo se lo è l'altro. Quindi abbiamo

S′ $\vdash_{Ts}$ s crede $[s$ crede $[F] \rightarrow F]$ se e solo se $\vdash_{Ts}$ s crede $[F]$.

A questo punto un confronto con il teorema di Löb può essere utile, anche per una analisi e valutazione del risultato ottenuto per Ts. Il teorema di Löb afferma:

L $\vdash_T$ $Dim([F]) \rightarrow F$ se e solo se $\vdash_T$ F

dove "Dim" sta per una formula che rappresenta la dimostrabilità formale in T e soddisfa le seguenti condizioni:

  se $\vdash_T$ G, allora $\vdash_T$ $Dim([G])$;
  $\vdash_T$ $Dim([G \rightarrow H]) \rightarrow (Dim([G]) \rightarrow Dim([H]))$;
  $\vdash_T$ $Dim([G]) \rightarrow Dim([Dim([G]])$.

Ammettendo

  se $\vdash_T$ $Dim([G])$, allora $\vdash_T$ G

– che si dimostra se T è ω-consistente – segue

L′ $\vdash_T$ $Dim([Dim([F]) \rightarrow F])$ se e solo se $\vdash_T$ $Dim([F])$.

L'equivalenza S′ è chiaramente analoga all'equivalenza L′.

3. S′ e L′ sono analoghi ma non ugualmente giustificati. Mentre L e L′, quest'ultima sotto l'ipotesi della ω-consistenza di T, hanno la solidità delle dimostrazioni dell'aritmetica, altrettanto non si può dire di S e S′.

Consideriamo l'affermazione più debole S′. Se la teoria Ts delle credenze di s è consistente, s è un soggetto razionale e F è l'enunciato assurdo "$0 = 1$", è lecito aspettarsi che Ts non dimostri che s crede che $0 = 1$, cioè:

$$\text{non} \vdash_{Ts} \; s \text{ crede } [0 = 1].$$

Supponiamo che effettivamente s non creda che $0 = 1$:

$$\neg(s \text{ crede } [0 = 1]).$$

Supponiamo anche che s non lo creda perché sa che "$0 = 1$" è falso. È possibile e plausibile che s riconosca di non credere che $0 = 1$ e, quindi, creda di non credere che $0 = 1$:

$$s \text{ crede } [\neg(s \text{ crede } [0 = 1])].$$

In virtù delle sue capacità logico-deduttive s può stabilire, e quindi credere, che se crede che $0 = 1$ allora $0 = 1$:

$$s \text{ crede } [s \text{ crede } [0 = 1] \rightarrow 0 = 1]$$

ma per S′

$$\text{non} \vdash_{Ts} \; s \text{ crede } [s \text{ crede } [0 = 1] \rightarrow 0 = 1].$$

Sembra inevitabile concludere che la nostra teoria Ts delle credenze di s è inadeguata perché esclude una possibilità che la competenza aritmetica e la razionalità logico-deduttiva di s richiedono di ammettere.

È facile vedere formalmente che Ts non può essere resa adeguata, o maggiormente adeguata, aggiungendo come assioma l'informazione che s non crede che $0 = 1$. Infatti sia ETs l'estensione di Ts così ottenuta e perciò tale che

$$\vdash_{ETs} \; \neg(s \text{ crede } [0 = 1]).$$

S′ vale anche per ETs, come pure C_1 e la chiusura deduttiva delle credenze di s. Quindi ETs si può dimostrare inconsistente nel modo seguente:

(a) $\neg(s \text{ crede } [0 = 1])$ assioma
(b) $s \text{ crede } [\neg(s \text{ crede } [0 = 1])]$ per C_1
(c) $s \text{ crede } [0 = 1] \to 0 = 1$ (a) per tautologia
(d) $s \text{ crede } [s \text{ crede } [0 = 1] \to 0 = 1]$ (b) e (c) per CDs
(e) $s \text{ crede } [0 = 1]$ (d) per S$'$.

Ts è resa inconsistente dall'aggiunta di un assioma che nega che un certo enunciato è creduto, data l'assunzione che gli assiomi siano creduti, cioè sostanzialmente nel modo nel quale Thomason (1980) indica che la sua teoria T può essere resa inconsistente. Tuttavia la derivazione dell'inconsistenza non ricorre alla problematica seconda condizione di Thomason che van Fraassen cerca di giustificare.

Si può evitare l'atteggiamento di rassegnazione di van Fraassen e fornire un'analisi del perché si ottiene l'inconsistenza? Mettendo da parte le riserve che si possono avere a considerare la diagonalizzazione un modo di ottenere l'autoriferimento e identificando la formula E con quella indicata da $[E]$, si può individuare la fonte della contraddizione proprio in D, cioè:

D $E \leftrightarrow (s \text{ crede } [E] \to F)$.

D è problematica quando il predicato "s crede" è intuitivamente inteso come tale da applicarsi solo a enunciati che dovrebbero essere comprensibili indipendentemente dalla sua applicazione a essi. Sembra infatti naturale sostenere che l'equivalenza D può essere compresa se e solo se l'enunciato "s crede $[E] \to F$" ha già un suo significato e, in una concezione composizionale stretta del significato, tale enunciato ha un significato se e solo se le sue parti "s crede $[E]$" e F hanno già un significato. D'altra parte "s crede $[E]$" ha un significato se e solo se l'oggetto di credenza E ha un significato, quindi se e solo se "s crede $[E] \to F$" ha un significato, quindi se e solo se "s crede $[E]$" ha un significato. Così, *apparentemente*, non è possibile afferrare il significato di "s crede $[E]$" senza presupporre come già dato nella comprensione di E il significato di "s crede $[E]$", contro il principio generale che un enunciato p deve essere comprensibile indipendentemente dalla comprensione di "x crede che p".

Tuttavia si potrebbe obiettare che l'analisi proposta del principio D riguarda il significato intuitivo che esso può avere, cioè un aspetto che non sembra avere un ruolo essenziale nella derivazione formale dell'inconsistenza di ETs. Non potrebbe essere che siano fatti sintattici non

rappresentabili formalmente la ragione dell'inconsistenza? Ad esempio, il fatto che s non crede che $0 = 1$ potrebbe essere di natura sintattica ma non esprimibile in modo compatibile con altri fatti sintattici riguardanti le credenze di s. Si tratta di una possibilità che non può essere esclusa a priori e che dovrebbe essere ulteriormente approfondita.

Una ragione che induce ad approfondire questa possibilità è la generalità con la quale il principio D risulta problematico nella prospettiva semantica strettamente composizionale sopra delineata. È possibile tuttavia adottare anche un punto di vista semantico diverso, noto nella letteratura come quello sotteso alle tavole di verità forti di Kleene, rispetto al quale non tutte le istanze di D risultano problematiche. Da questo punto di vista si può pensare che affinché un enunciato complesso abbia un significato sia sufficiente che una sua sottocomponente enunciativa abbia un valore di verità tale da determinare il valore di verità dell'enunciato intero quale che sia il valore che si può attribuire ad altre sottocomponenti. Nel caso di E, se F è un enunciato vero, la verità di "s crede $[E] \rightarrow F$", e quindi di E, può essere ritenuta del tutto determinata, poiché il condizionale "s crede $[E] \rightarrow F$" è reso vero dalla verità del conseguente indipendentemente da quale sia l'eventuale valore di verità del suo antecedente. In questo caso l'equivalenza D risulta vera. Se invece F è falso, ad esempio è "$0 = 1$", il condizionale "s crede $[E] \rightarrow F$", quindi E, dovrebbe essere considerato indeterminato, e quindi anche l'equivalenza D risulterebbe indeterminata.

In conclusione ci sono ancora vari aspetti da chiarire riguardo alla non rappresentabilità sintattica del credere che spesso si tende a dare per scontata.[4]

Riferimenti bibliografici

Löb, M. (1955), "Solution of a problem of Leon Henkin", *Journal of Symbolic Logic*, 20, pp. 115–118.

Montague, R. (1963), "Syntactical treatment of modality, with corollaries on reflection principles and finite axiomatizability", *Acta Philsophica Fennica*, 16, pp. 153–167, ristampato nel suo *Formal Philosophy*, Yale University Press, New Haven e Londra 1974, pp. 286–302.

Smullyan, R. (1987), *Forever Undecided. A Puzzle Guide to Gödel*, A. Knopf, New York, ristampato da Oxford University Press, Oxford 1988.

Tarski, A. (1935), "Der Wahrheitsbegriff in den Formalisierten Sprachen", *Studia Philosophica*, 1, pp. 261–405.

Thomason, R.H. (1980), "A note on syntactical treatments of modality", *Synthese*, 44, pp. 391–395.

——— (2011), "Some limitations to the psychological orientation in semantic theory", *Journal of Philosohical Logic*, 40, pp. 1–14.

van Fraassen, B. (2011), "Thomason's paradox for belief, and two consequence relations", *Journal of Philosophical Logic*, 40, pp. 15–32.

Gödel, Priest e la nozione di *prova informale*

Massimiliano Carrara ed Enrico Martino

1. Introduzione

Nel suo "The Logic of Paradox" (1979), Priest ha sviluppato un argomento, fondato sul concetto di *prova informale*, il cui scopo è mostrare che il primo teorema di incompletezza di Gödel suggerirebbe la presenza di dialeteie (enunciati sia veri che falsi) nel modello *standard* dell'aritmetica (su questo stesso tema si veda anche Priest 1994. Per una breve introduzione all'argomento si veda Murzi e Carrara 2015).

L'argomento è stato criticato da Chihara (1984), Shapiro (2002), Berto (2009) e altri. Anche Ernesto Napoli vi ha dedicato un lavoro, intitolato "Priest's Paradox" e pubblicato su *Logique et Analyse* nel 1985. L'argomento è stato riproposto da Priest nella prima edizione di *In Contradiction* (1987) ed espanso nella seconda edizione (2006).

Gran parte della critica all'argomento di Priest è diretta contro la stessa nozione di prova informale, in particolare contro la tesi che tutto ciò che è informalmente dimostrabile è vero. Sicuramente, se la nozione di prova informale è intesa abbracciare tutte le prove effettivamente eseguibili dai matematici sul campo, come Priest sembra suggerire, la tesi è insostenibile. Qui, però, adotteremo una nozione ideale di prova informale, compatibile con i principi del dialeteismo, per cui la tesi è, in linea di principio, accettabile. Obiettivo del saggio è sostenere che, anche per una tale nozione idealizzata di prova, le assunzioni di Priest sono difficilmente sostenibili. Cercheremo di rendere più perspicuo ciò che è implicito nelle assunzioni di Priest e di presentare una loro conseguenza che dovrebbe essere respinta anche da un dialeteista.

2. Sulla nozione di prova informale

Priest descrive la nozione di prova informale nel modo seguente:

> La prova, come è intesa dai matematici (non dai logici), è quel processo di
> argomentazione deduttiva con cui stabilisco che certe affermazioni matematiche
> sono vere. In altre parole, supponiamo di avere un'affermazione matematica,
> per esempio un'affermazione sulla teoria dei numeri, di cui si voglia stabilire
> verità o falsità. Cerco una prova o una refutazione, cioè una prova della sua
> negazione [...] chiamerò le argomentazioni deduttive informali effettuate a
> partire da enunciati fondamentali *prove informali.* (Priest 2006: 40)

Il presunto paradosso dovrebbe essere suggerito dall'analogia della *provabilità informale* dell'enunciato indecidibile G di Gödel con il paradosso
del mentitore:

> Come è chiaro a chiunque abbia familiarità con il teorema di Gödel, al suo cuore
> si trova un paradosso. Informalmente l'enunciato 'indecidibile' dice "Questo
> enunciato non è dimostrabile". Supponiamo che sia dimostrabile; allora, dal
> momento che tutto ciò che è dimostrabile è vero, non è dimostrabile. Perciò
> non è dimostrabile. Ma lo abbiamo appena dimostrato. Quindi è dopo tutto
> (anche) dimostrabile. (Priest 2006: 237)

Questo argomento è ben noto ed è stato ampiamente discusso in letteratura. Seguendo Dummett, chiamiamo tale argomento "la prova semplice" (Dummett 1978: 186–201). Vale la pena notare che la *prova semplice* vale anche dialeteicamente. Alcuni hanno sostenuto che una tale prova utilizza implicitamente la consistenza dell'aritmetica di Peano (PA) che, per il secondo teorema d'incompletezza, è formalmente indimostrabile. Si osservi, tuttavia, che la prova in questione non assume la consistenza di PA, ma il fatto che ciò che è dimostrabile è vero, da cui, dialeteicamente, non segue la consistenza. Né la prova sfrutta la *reductio ad absurdum*; sfrutta invece il *tertium non datur*, che vale dialeteicamente. La prova è così fatta:

G è dimostrabile o non lo è. Ma, se è dimostrabile, allora è vero e quindi, come dice, è (anche) indimostrabile. Quindi, in ogni caso, non è dimostrabile e quindi vero.

Siamo d'accordo con Priest che G è informalmente dimostrabile, anche da un punto di vista dialeteico. L'aspetto paradossale, però, del teorema di

Gödel è solo apparente. Infatti, ciò che l'interpretazione metalinguistica di G dice è che l'enunciato *aritmetico* G non è dimostrabile.

Questa proposizione, a differenza degli autentici paradossi semantici, non comporta alcun autoriferimento, perché l'interpretazione aritmetica di G è espressa in termini di operazioni aritmetiche primitive e non in termini di dimostrabilità. L'apparente paradosso nasce dalla confusione tra due differenti nozioni di *prova*. G non è dimostrabile nel senso che non è derivabile dagli assiomi dell'aritmetica di Peano (PA); è dimostrabile nel senso che si può riconoscere informalmente la sua verità nel modello *standard* dell'aritmetica.

3. Prova informale nella prova semplice

Esaminiamo il tentativo di Priest di riprodurre la *prova semplice* sostituendo la PA-dimostrabilità con la dimostrabilità informale.

Si introduca, innanzitutto, nel linguaggio L dell'aritmetica del primo ordine, un predicato $\beta_N(x)$ sui numeri naturali, significante "x è (il codice di) un enunciato informalmente dimostrabile". Priest afferma che il predicato soddisfa i seguenti principi:

(1) $\quad \vdash_N \beta_N(\langle\alpha\rangle) \to \alpha$,
(2) $\quad$ Se $\vdash_N \alpha$ allora $\vdash_N \beta_N(\langle\alpha\rangle)$,

dove $\langle\alpha\rangle$ è il codice di α e dove $\vdash_N$ esprime la *provabilità informale*. Scrive Priest:

> Per (1), è analitico che tutto ciò che è dimostrabile informalmente è vero. [...] E poiché ciò è analitico è esso stesso informalmente dimostrabile. Per (2), se qualcosa è informalmente dimostrato allora questo stesso fatto costituisce una prova che α è dimostrabile. (Priest 2006: 238)

Poi, con il metodo usuale dell'autoriferimento, Priest costruisce un enunciato γ della forma $\neg\beta_N(\langle\gamma\rangle)$ e utilizza la dimostrazione semplice per mostrare che sia γ che $\neg\gamma$ sono dimostrabili. Ne conclude che l'aritmetica è incoerente.

Priest afferma esplicitamente di non aver assunto che β_N possa essere costruito a partire dall'usuale vocabolario aritmetico $(', +, \times)$. Osservia-

mo, in proposito, che se β_N non è aritmetico, ma è introdotto intensionalmente come il predicato di dimostrabilità informale degli enunciati aritmetici, (1) e (2) devono essere limitati a un α aritmetico. E poiché γ non è aritmetico non ne segue alcuna contraddizione.

D'altra parte, se β_N è introdotto intensionalmente come il predicato di dimostrabilità informale degli enunciati del linguaggio esteso $L \cup \{\beta_N\}$, otteniamo un autentico paradosso dell'autoriferimento. In questo secondo caso a essere inconsistente è la teoria estesa. Così, in entrambi i casi, l'inconsistenza dell'aritmetica non segue.

Priest è ben consapevole di ciò; pensiamo che, in questa parte dell'argomentazione, la sua affermazione che l'aritmetica è inconsistente non sia altro che un modo impreciso di dire che la teoria estesa, unitamente ai principi (1) e (2) per il linguaggio esteso, è inconsistente.

Successivamente, Priest suggerisce che β_N sia aritmeticamente esprimibile. Questa è la parte più interessante della sua argomentazione. Il suggerimento nasce dal fatto, sottolineato da Dummett, che, quando una prova viene presentata, siamo in grado – almeno in linea di principio – di riconoscerla come tale. Priest cita il seguente passo di Dummett:

> Gli intuizionisti sono inclini a scrivere che, mentre non possiamo delimitare in anticipo il regno di tutte le possibili prove intuizionisticamente valide, tuttavia, per particolari prove date e particolari principi di prova enunciati, possiamo accertare che siano intuizionisticamente corrette. (Dummett 1959: 347)

In effetti, c'è un senso in cui, secondo Dummett, anche la nozione di prova informale è decidibile: quando ci venga esibita una *presunta* prova di un enunciato matematico, siamo in grado, in linea di principio, di decidere se è o no un'autentica prova. In tal senso, ove si assuma la possibilità di una codificazione numerica del linguaggio delle prove informali, l'insieme (dei codici) delle prove informali è decidibile. Priest ne deduce che, a norma della tesi di Church, tale insieme è ricorsivo. Ne segue che l'insieme (dei codici) degli enunciati informalmente dimostrabili è ricorsivamente enumerabile e quindi che il predicato di dimostrabilità informale è aritmetico.

Sulla base di queste considerazioni, Priest introduce un predicato β *aritmetico* (senza pedice) coestensivo con β_N. Così, il corrispondente enunciato di Gödel γ è aritmetico e, in virtù della prova semplice, risulta essere una dialeteia. Dunque, Priest conclude, l'aritmetica è contraddittoria. Analizziamo nel dettaglio l'argomento di Priest.

4. Analisi critica dell'argomento di Priest

In primo luogo, per quel che riguarda le prove informali, è dubbio che l'uso della tesi di Church sia appropriato. La tesi garantisce che qualsiasi insieme di numeri decidibile mediante una qualunque procedura *meccanica* sia ricorsivo. Ora, siamo d'accordo con Dummett che, in linea di principio, l'insieme delle prove ingenue deve essere decidibile nel senso chiarito sopra. Certo, si potrebbe obiettare che la questione se una prova putativa sia corretta o meno è, in larga misura, soggettiva. Ad esempio, ciò che è accettato come corretto da un classico può essere rifiutato da un dialeteista o da un intuizionista. Tuttavia, come detto sopra, considereremo una nozione *ideale* di prova informale relativa a una concezione ben determinata di prova e correttezza deduttiva.

Allo scopo, introduciamo un matematico ideale – chiamiamolo Ernesto-detto-Graham – libero da vincoli empirici spazio-temporali, il cui ragionamento è sempre privo di errori. Possiamo considerarlo simile al *soggetto creativo* di Brouwer; tuttavia, a differenza di quest'ultimo, non si suppone che sia un intuizionista, ma piuttosto un dialeteista. Per "prova informale", intendiamo una prova, informale appunto, che sia eseguibile da un tale personaggio di fantasia. Per questa nozione di prova la tesi che tutto ciò che è dimostrabile è vero è certamente accettabile. Assumiamo quindi che, quando gli sia presentata una sequenza finita di parole, Ernesto-detto-Graham sia in grado di decidere se conta come prova o meno.

Quel che è problematico è la questione se una tale capacità sia di natura *meccanica*. Certo non lo è per Dummett che, da buon intuizionista, è esplicitamente favorevole a una concezione anti-meccanicistica del ragionamento intuitivo. Nel passo sopra citato osserva esplicitamente che "non possiamo delimitare in anticipo il regno di tutte le possibili prove intuizionisticamente valide". In particolare, se la prova informale è assiomatica, così come sostiene Priest, si può utilizzare come assioma qualsiasi enunciato *evidente* e sembra essere molto dubbio che esista una procedura meccanica in grado di stabilire, per qualsiasi enunciato, se è evidente o meno. Priest insiste che una concezione meccanicistica della mente umana sia appropriata per spiegare come la nozione di *prova informale* è appresa. Al contrario, riteniamo che la nozione di prova informale

sia acquisita per mezzo di un *atto di astrazione*, dopo aver visto un'adeguata varietà di particolari prove. Così come acquisiamo il concetto di *cavallo* dopo aver visto un congruo numero di cavalli; sicché, quando ci viene presentato un animale, siamo in grado, in linea di principio, di decidere se si tratta di un cavallo o meno. Ma non vi è alcuna evidenza che tale capacità sia di natura meccanica. E se, per ragioni filosofiche, crediamo che lo sia, tuttavia non siamo in grado di esibire effettivamente una macchina che abbia tale capacità. Questo punto, come si vedrà, è piuttosto rilevante per la nostra discussione.

Assumiamo, per amore di discussione, una concezione meccanicistica della mente umana, secondo la quale l'insieme delle prove informali sia ricorsivo. Sia β, come sopra, la controparte aritmetica del predicato β_N di dimostrabilità informale. Come osserva Priest, β è solo *estensionalmente* equivalente a β_N. Una tale concezione meccanicistica della mente umana può, al più, suggerire l'esistenza di qualche predicato aritmetico estensionalmente equivalente a β_N, ma non per questo è in grado di fornire l'espressione aritmetica di un tale predicato. In altre parole, il solo fatto che Ernesto-detto-Graham è convinto, per ragioni filosofiche, che deve esistere una controparte β aritmetica di β_N non gli consente di trovare effettivamente un tale β e di dimostrarne l'equivalenza estensionale a β_N. In mancanza di ciò, (1) e (2) – evidenti per β_N – non valgono per β. Infatti, supponiamo che a Ernesto-detto-Graham sia presentata l'espressione aritmetica di un β che, di fatto, sia equivalente a β_N, ma tale che non vi sia alcuna prova informale di tale equivalenza. In questa situazione Ernesto-detto-Graham non può provare che, per tutti gli α, $\beta(\langle\alpha\rangle) \rightarrow \alpha$, perché non sa che, per ogni α, $\beta(\langle\alpha\rangle)$ è vera se e solo se α è informalmente dimostrabile. In particolare, Ernesto-detto-Graham non è in grado di dimostrare che l'enunciato gödeliano γ, la cui forma è $\neg\beta(\langle\gamma\rangle)$, è vero se e solo se è indimostrabile.[1] Tutto ciò che Ernestodetto-Graham può congetturare, in virtù del suo credo filosofico, è che, per alcuni predicati ricorsivamente enumerabili il corrispondente enunciato gödeliano sia vero e indimostrabile, ma, non conoscendo un tale enunciato, non può dimostrarlo e ottenere, quindi, una contraddizione.

Comunque, al fine di raggiungere la conclusione di Priest, facciamo

---

[1] Un'obiezione simile è stata formulate da Chihara (1984), ma per quel che ne sappiamo non è mai stata presa in considerazione da Priest.

l'assunzione implausibile che, per un certo β ricorsivamente enumerabile, Ernesto-detto-Graham sia in grado di dimostrare che β è equivalente a β_N e che quindi γ è una dialeteia.

Si assuma, per il momento, così come fa Priest, che la nozione di prova informale sia assiomatica e sia Σ l'insieme degli assiomi (enunciati aritmetici immediatamente evidenti). L'evidenza degli assiomi può essere intesa nel senso che, quando gli sia presentato un qualunque enunciato di Σ, Ernesto-detto-Graham ha immediata evidenza della sua verità. Ciò non implica, tuttavia, che Ernesto-detto-Graham possieda la *simultanea evidenza* di tutti gli (infiniti) enunciati di Σ. Ma assumiamo, a favore di Priest, che Ernesto-detto-Graham abbia simultanea evidenza di tali enunciati e abbia accesso ai loro codici, come accade nel caso familiare degli assiomi di Peano.

Sia $\Pi(m,n)$ il predicato aritmetico:

> m è una prova informale di n;

sia $P(n) = \exists x \Pi(x,n)$ il predicato aritmetico di dimostrabilità informale e γ il relativo enunciato di Gödel. Domanda: può Ernesto-detto-Graham dimostrare che γ è una dialeteia?

Si osservi che ciò che è derivabile da Σ deve valere in qualsiasi modello di Σ. Ne segue che ciò che è informalmente deducibile da Σ è precisamente ciò che Ernesto-detto-Graham può riconoscere vero in un modello arbitrario di Σ, a lui sconosciuto, sotto la sola assunzione che si tratti di un modello di Σ. Sia $\mathcal{M}$ un tale modello. Come è noto, qualunque insieme di enunciati, vero nel modello *standard*, ha modelli non-*standard*. Così Ernesto-detto-Graham non sa se $\mathcal{M}$ è *standard*. Ma P rappresenta il predicato di dimostrabilità solo nel modello *standard*: se $\mathcal{M}$ è non *standard*, l'enunciato esistenziale $\exists x \Pi(x,n)$ può essere esemplificato da un numero non *standard*, che non è un codice di una prova. Così Ernesto-detto-Graham sa che, se $\gamma = \neg \exists x \Pi(x, \langle\gamma\rangle)$ è vero, allora è indimostrabile; ma non può sapere che, se è falso, allora è dimostrabile. Così Ernesto-detto-Graham non può concludere che γ è una dialeteia.

Forse si potrebbe obiettare che la prova informale riguarda il modello *standard*: ciò che è informalmente dimostrabile è ciò che si può riconoscere come vero nel modello *standard*. E in effetti Ernesto-detto-Graham riconosce che γ è vero nel modello *standard*. Ma ciò che il nostro argomento dimostra, mediante la considerazione di un modello arbitrario, è che la

verità di γ nel modello *standard* non può essere derivata da Σ. E poiché, d'altra parte, la verità di γ è riconosciuta da Ernesto-detto-Graham, ne concludiamo che la nozione di prova informale non è assiomatizzabile. Questa conclusione è in accordo con la caratteristica peculiare del modello *standard*, ovvero il fatto che ogni numero ha un numero finito di predecessori, dove la *finitezza* è qui da intendersi come *assoluta*, cioè non relativa a un modello della teoria degli insiemi. Tale nozione assoluta, come è noto, non può essere catturata da nessun sistema assiomatico. Tuttavia è afferrata dalla intuizione umana e può essere sfruttata in una prova informale sul modello *standard*. Infatti, è sfruttata nell'aritmetizzazione del predicato di provabilità formale, relativo all'aritmetica di Peano, e quindi nella dimostrazione informale della verità, nel modello *standard*, del relativo enunciato gödeliano. Nel caso dell'aritmetica di Peano, si osservi che il requisito della *simultanea* evidenza degli assiomi è chiaramente soddisfatto per la presenza di un numero *finito* di assiomi più lo schema di induzione, gli infiniti esempi di questo essendo evidenti proprio in virtù della nozione informale di finitezza assoluta.

L'ipotesi che la nozione di prova informale sia assiomatizzabile, tuttavia, non sembra indispensabile per gli scopi di Priest. Anche se le prove informali sono espresse nel linguaggio naturale e possono sfruttare qualunque intuizione circa il modello *standard*, in particolare la finitezza assoluta, si può supporre che, con riferimento a una codificazione adeguata del linguaggio informale, i codici di tutte le prove informali degli enunciati aritmetici formino un insieme ricorsivo accessibile a Ernesto-detto-Graham. La possibilità di inventare prove informali, va notato, sembra così ampia e incontrollabile che è difficile immaginare come il povero Ernesto-detto-Graham, anche se altamente idealizzato, possa dimostrare che un dato insieme aritmetico codifichi tutte le possibili prove informali. Ma supponiamo, di nuovo per amore di discussione, che possa farlo. Supponiamo cioè che Ernesto-detto-Graham conosca l'espressione aritmetica del predicato $\Pi(m, n)$ e possa informalmente dimostrare che, per ogni m, n, $\Pi(m, n)$ è vera se e solo se m è una prova di n. In questo caso Ernesto-detto-Graham può riconoscere che il predicato aritmetico $P(x) = \exists y \Pi(y, x)$ esprime la dimostrabilità informale e può informalmente dimostrare che, secondo i *desiderata* di Priest, l'enunciato γ di Gödel è una dialeteia.

Vogliamo mostrare, tuttavia, come, una volta accettate tutte le assunzioni che portano a riconoscere con Priest l'esistenza di dialeteie nel modello *standard*, si possa andare oltre e dimostrare, con principi dialeteicamente validi, che il modello *standard* dell'aritmetica è banale (cioè che ogni enunciato è vero), un risultato che anche un dialeteista dovrebbe rifiutare.

Sia $P(m, n)$ la relazione aritmetica:

n è (informalmente) dimostrabile dall'assunzione m.

Supponiamo di nuovo che Ernesto-detto-Graham abbia accesso al predicato aritmetico P e abbia dimostrato la sua interpretabilità in termini di prove. Dato un enunciato arbitrario A, sia H l'enunciato della forma $P(\langle H \rangle, \langle A \rangle)$. Vogliamo dimostrare che A è vero. Supponiamo che H sia vero. Allora A è dimostrabile da H; e, dal momento che le prove sono corrette (cioè conservano la verità), A è vero. Così abbiamo dimostrato A da H. Pertanto $P(\langle H \rangle, \langle A \rangle)$ è vero, cioè H è vero. Sicché, essendo A dimostrabile da H, A è vero.

Analizziamo questo argomento presentando le regole formali per la sua derivazione in deduzione naturale.

Una sequenza della forma $\Gamma \vdash A$ stia a significare:

A è deducibile informalmente dalle assunzioni Γ.

Useremo le seguenti regole di deduzione:

$\Gamma, A \vdash A$ (rifl.)

$$\frac{\Gamma \vdash A \qquad \Delta, A \vdash B}{\Gamma, \Delta \vdash B} \text{ (taglio)} \qquad \frac{\Gamma, A, A \vdash B}{\Gamma, A \vdash B} \text{ (contrazione)}$$

$$\frac{\Gamma, A \vdash B}{\Gamma \vdash P(\langle A \rangle, \langle B \rangle)} \text{ (P el.)} \qquad \frac{\Gamma \vdash P(\langle A \rangle, \langle B \rangle)}{\Gamma, A \vdash B} \text{ (P intr.)}$$

Le regole per P sono analitiche, le altre sono banali.[2] Sia H, come sopra, un enunciato della forma $P(\langle H \rangle, \langle A \rangle)$. Di seguito abbiamo la prova formale di A:

$$\cfrac{\cfrac{\cfrac{\cfrac{H \vdash P(\langle H \rangle, \langle A \rangle) \ \text{(rifl.)}}{H, H \vdash A} \ (P \ \text{el.})}{H \vdash A} \ \text{(contr.)}}{\vdash P(\langle H \rangle, \langle A \rangle)} \ (P \ \text{intr.}) \qquad H \vdash A}{A} \ \text{(taglio)}$$

È importante osservare che la nostra prova della banalità del modello *standard* non utilizza le regole classiche (*reductio ad absurdum* e ECQ) che portano da una qualunque contraddizione alla banalità. Per questo pensiamo che dovrebbe essere accettata anche da un dialeteista. Ma, ovviamente, Priest non vuole suggerire che il modello *standard* dell'aritmetica sia banale.

Infine, si consideri il seguente scenario. Supponiamo ancora una volta che le prove informali siano ricorsivamente enumerabili per mezzo di un algoritmo accessibile (nel senso sopra specificato) a Ernesto-detto-Graham, in modo che egli possa dimostrare che l'enunciato di Gödel G è una dialeteia. Lasciamo che Ernesto-detto-Graham corra attraverso le prove alla ricerca di una prova dell'enunciato di Gödel G. Poiché G è vero e dice di essere indimostrabile, il povero Ernesto-detto-Graham non riuscirà mai a trovare una prova di G. Eppure Ernesto-detto-Graham è riuscito a dimostrare G. Potrebbe Priest accettare come un 'fatto della vita' (i famosi "facts of the life" di cui spesso parla), che non si possa trovare qualcosa che si è trovato? Un tale scenario ci pare inintelligibile.

5. Conclusioni

Concludiamo osservando che la discussione di Priest su prova informale e teorema di Gödel è inadeguata a supportare il suo suggerimento che

---

[2] Alcune teorie logiche, piuttosto sofisticate, rifiutano la contrazione. Noi pensiamo che essa debba essere banalmente accettata in una prova informale: nessun matematico al lavoro (classico o di altro genere) è vessato dal bizzarro dubbio che ciò che segue da A e A possa non seguire da A!

qualche enunciato aritmetico sia una dialeteia nel modello *standard*. Certamente le prove informali sono, almeno in linea di principio, corrette e preziose per scoprire la verità di alcuni enunciati del modello *standard*, in particolare degli enunciati gödeliani. Riteniamo, tuttavia, che le nostre riflessioni supportino la conclusione che la natura stessa della nozione di prova informale esclude la possibilità di costruire una macchina in grado di generare tutte le *possibili* prove informali.

Riferimenti bibliografici

Berto, F. (2009), "The Gödel paradox and Wittgenstein's reasons", *Philosophia Mathematica*, 17, pp. 208–219.

Chihara, C.S. (1984), "Priest, the liar, and Gödel", *Journal of Philosophical Logic*, 13, pp. 117–124.

Dummett, M. (1959), "Wittgenstein's philosophy of mathematics", *Philosophical Review*, 68, pp. 324–348.

——— (1978), *Truth and Other Enigmas*, Harvard University Press, Cambridge, MA.

Murzi, J. e Carrara, M. (2015), "Paradox and logical revision. A short introduction", *Topoi*, 34, pp. 109–119.

Napoli, E. (1985), "Priest's paradox", *Logique et Analyse*, 28, pp. 403–407.

Priest, G. (1979), "The logic of paradox", *Journal of Philosophical Logic*, 8, pp. 219–241.

——— (1994), "Is arithmetic consistent?", *Mind*, 103, pp. 337–349.

——— (2006), *In Contradiction: A Study of the Transconsistent*, Oxford University Press, Oxford.

Shapiro, S. (2002), "Incompleteness and inconsistency", *Mind*, 111, pp. 817–832.

From Ernesto to eternity, and beyond*

Giuseppe Spolaore

> The ideas of time and of eternity really have a common
> source, for we can in our thoughts add certain lengths of
> duration to one another, doing this as often as we please.
>
> G.W. Leibniz, *New Essays*, II.XIV.10

0. There are a few things that Ernesto Napoli, and nearly nobody else, takes to be wrong. Sometimes, these are things that break the strictures of his general no-loose-talk policy. One example is resorting to scare quotes. (If you are going along with the wrong word, signalling that you are going along with the wrong word does not make it any less wrong.) Another is using "the fact that" (unless you believe that the things philosophers call "facts" exist, which only makes your situation worse). One more example is using "necessary statements" to mean *necessarily true* statements. (To Ernesto's horror, Kripke himself made this very mistake at times, in *Naming and Necessity* and elsewhere.) Of course, often Ernesto takes certain things to be wrong for more substantial reasons. (I take responsibility for this distinction between loose talk and substantial mistakes; notoriously, Ernesto is prepared to go a long way to avoid drawing it.)

1. Some years ago, Ernesto told me that it is wrong to equate "necessarily" with "at all possible worlds". For when we say, for example, that "$1 + 1 = 2$" is necessarily true, we do not mean that "$1 + 1 = 2$" is true at all possible worlds. What we mean is, rather, that "$1 + 1 = 2$" is true *independently* of the world, that is, in a world-insensitive way.

2. Ernesto, I believe, would be happy to extend this remark from the modal to the temporal case. Or at least, this is a conclusion I am prepared to draw from his claim that "[t]ruth is not constant in time but insensitive to time" (2010: 298, the translation is mine; this passage is possibly reminiscent of Frege 1969). It is wrong, again, to equate the time-insensitivity of truth with truth at all times, for essentially the same reason that it is wrong to equate world-insensitivity to truth at all worlds.

3. "Eternally" is often taken to be a temporal homologous of "necessarily". Interestingly enough, it appears that there are two notions of eternity in the philosophical market, and one is tempted to let them correspond to time-insensitivity and constancy though time, respectively. On the one hand, we have a timeless ('Boethian' or atemporal) notion of eternity. To be timelessly eternal (or simply timeless, or atemporal, or the like) is to be wholly unaffected by time – to be *out of time*, as it were.[1] On the other hand, what Leibniz has in mind in the opening quote of this note is definitely a temporally loaded, non-timeless kind of eternity. (Timeless eternity is thought to be timeless precisely because it does *not* resemble time in many essential respects; for instance it involves no temporal succession.) There is a general (albeit not a universal, see Kneale 1968) consensus that spatiotemporal entities, including the cosmos, can only be said to be eternal in the latter, temporal sense.

4. World-insensitivity is an elusive notion. Thus, it comes as no surprise that many philosophers tend to ignore it, and to hold that constancy across possible worlds is all that there is to necessity. Analogously, one might argue that constancy through time is all that there is to eternity. Ernesto (so much may be deemed certain at this point) would disagree. He would likely insist that eternity, atemporally understood as time-insensitivity, is a legitimate notion, and definitely not to be confused with

---

[1] In Boethius the notion of timeless eternity is especially designed to apply to God ("Aeternitas est interminabilis vitae tota simul et perfecta possessio", *De Consolatione Philosophiae*, V, 6). It is common, however, to conceive atemporal eternity less restrictedly, so as to cover also lifeless entities such as Platonic *abstracta*. I assume that, in this non-theological understanding, timeless eternity need not entail the possession of temporal duration. For overviews of the (notoriously slippery) theological debate about eternity, see Kretzmann and Stump 1998 and Helm 2014.

constancy through time.[2] The first thing I do in this short note is to go beyond Ernesto, and observe that *temporal* eternity is not to be confused with constancy through time either (§§ 5–11). In other words (henceforth "eternally" is always understood in its temporal reading unless otherwise specified):

(E)   "Eternally" just does not mean the same as "at all times".

In the second place, I show that when we claim that mathematical or logical truths are – literally and unrestrictedly – eternal, what we (should) mean is that they are time-insensitive, that is, timelessly eternal (§ 13). To this aim, I make use of Ernesto's intuition that "$1 + 1 = 2$" is a world-insensitive truth.

5. To begin with, let me put some emphasis on the word "just" in (E). The point is not that there is a nuanced, super-subtle difference between "eternally" and "at all times". Not at all. The difference in meaning is robust, thick, and obvious.

Put in this way, (E) might sound incredible. After all, the opposite view seems deeply entrenched and regularly taken for granted; witness the following (more or less randomly chosen) quotes:

> A proposition is eternally true iff it is true at all times. (Zalta 1987: 225)

> [A] parallel proof with the phrase 'at all times' in place of 'necessarily' [. . .] would prove eternal rather than necessary existence. (Williamson 2002: 234; the 'proof' at stake is Williamson's well-known argument to the conclusion that everything exists necessarily; see 2002: 233–234)

> Let us define, for any verb 'to v' another verb 'to eternally v' thus: 'x *eternally* v's' is 'x *has* v'ed or x v's or x *will* v'. The important thing about a sentence whose verb is an eternal one is, as the name implies, that if it is true at any time it is true at all times. (Priest 1986: 165)

It is not my intention to claim, however, that the above, authoritative sources are all on the wrong track. For instance, Zalta's quote need not amount to a claim of *synonymity* between "eternally" and "at all

---

[2] To be sure, here I mean that Ernesto would take the notion of timeless eternity as legitimate to the extent that it can be made to coincide with the notion of time-insensitivity. I suspend judgement as to how Ernesto would assess other notions of timeless eternity, including those that are at play in the theological debate (see above, fn. 1).

times". Zalta only commits himself to take "x is eternally true" and "x is true at all times" as (necessarily) coextensional, possibly under some assumptions that both he and his readers are prepared to grant.[3] One gets the impression that the synonymity of "eternally" and "at all time" is taken for granted, but what is really taken for granted is that they are broadly interchangeable under some assumptions. What assumptions? Well, that is a question I address toward the end of this note. For the time being, we may start with a different, albeit related, question.

6. In what follows, by "cosmos" I mean nothing less than the totality of all spatiotemporal stuff. Moreover, I take the cosmos to include not just the things that exist at the present time, but also those that exist at any previous or later time.[4] Now, the question is:

(1)   Do times exist only insofar as the cosmos exists?

I believe it true – and true by necessity – that the timeline does not extend beyond the limits of the cosmos, and so that the right answer to (1) is yes. But even if you do not agree with me on this point, I assume that you are at least prepared to concede that (1) is an *open question*, in Moore's (1903) sense. We need not suppose that the answer is nowhere to be found – maybe future physicians, or metaphysicians, will be skilled enough to stand upon our shoulders and see. But we do not expect it to be found *just* by consulting a dictionary, or by exploiting our semantic competence, or by using logic, and so on. In other words, the question whether some time exists without the cosmos existing is definitely not a semantic one.

7. Since what bothers us in the end is a semantic issue – whether "eternally" and "at any time" are synonyms – we are free to answer question (1) in any way we please. Suppose, then, that the answer is yes:

(2)   Times only exist insofar as the cosmos exist.

---

[3] As for Priest's quote, one might also argue that Priest is offering a *definition* of "to eternally v", and that the meaning of defined notions need not coincide with that of their informal, non-defined counterparts.

[4] I am leaving certain ontological qualms aside for simplicity here. I assume that presentists are as able to make sense of what follows as anybody else.

Now consider:

(3)   The cosmos exists at all times;
(4)   The cosmos is eternal.[5]

Obviously, (2) is equivalent to (3). But is it equivalent to (4)? Arguably, it is not. For one can reject (4) for reasons that are independent of (2), for instance, because one holds that the cosmos has a beginning, or an end, in time.

8. If "eternally" does not mean *at any time*, what does it mean? Or, if you prefer, what gloss might we provide, were we to explain the meaning of "eternally" to someone who has never met the word? A first, natural candidate is "endlessly", in its literal sense of "with no end" (or, depending on the context, "with neither a beginning nor an end"). Please note that the difference in meaning between "endlessly", so understood, and "at all times" is significant. By a reasoning that strictly parallels the one just made about "eternally", we can conclude, say, that "The cosmos exists at all times" is equivalent to "The cosmos is endless (toward both temporal directions)" only at the price of some additional assumption.[6]

Still, "endlessly" does not mean the same as "eternally", although the difference is thinner in this case. A sequence of events counts as endless iff it lacks (depending on the context) either a beginning – an initial element – or an end – a final element, or both. Therefore, an empty sequence of events qualifies as endless. But it is not temporally eternal. What if we require that a sequence is eternal iff it is endless *and* nonempty? Things get better, but we can still figure out a counter-example. Consider denumerable, Zeno sequence of events

(Z) $\langle \ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots \rangle$

such that (i) x_0 lasts for 10 minutes, (ii) with $n < 0$, the duration of x_n is half that of x_{n+1}, and (iii) with $n > 0$, the duration of x_n is half of that of x_{n-1}. (Z) is (literally) endless, but it lasts for half an hour. And nothing eternal lasts for less than one hour.

---

[5] Here and henceforth, I use "eternal" as strictly synonymous with "eternally existing/obtaining".

[6] Not to go too far, an assumption that does the trick is "The cosmos is endless" (!).

To be sure, one might object that (Z) is not genuinely endless, for we can find something like an end (a beginning) for it – that is, some more or less arbitrarily chosen time that follows (precedes) all the elements of (Z). But for all that has been said so far, it is perfectly possible that there exists no such time at all. To dramatize a little bit, we can even imagine that (Z) takes place in a half-hour long cosmos. Even so, it is reasonable to deny that (Z) is eternal. Consider a minimally larger sequence (Z*), obtained by enriching (Z) with a beginning and an end, that is, events e, e' such that $e < x_n < e'$ for all n. A limited sequence like (Z*) is uncontroversially non-eternal, and it is very peculiar to suppose that an eternal thing may be contained within a non-eternal one.

9. The point just made can be extended and made more precise with reference to the semantics of so-called tense logics. Let $\mathcal{L}$ be a simple temporal propositional language endowed with two primitive tense operators H ("at any past time") and G ("at any future time").[7] A *basic temporal frame* for $\mathcal{L}$ is a pair $\mathcal{F} = (T, <)$, where T is a nonempty set of *times* and $<$ is a strict total ordering on T. Informally, $<$ is understood to be a relation of *temporal precedence* between times. A *basic temporal model* is a pair $\mathcal{M} = (\mathcal{F}, I)$, where $\mathcal{F}$ is a basic temporal frame and I is an interpretation function from atoms $p, p_1, \ldots$ of $\mathcal{L}$ to subsets of T. The semantics of $\mathcal{L}$ is defined in the usual Tarskian way. The only interesting semantic clauses for our purposes are those relative to H and G (A varies over formulas of $\mathcal{L}$, t over times in T):

> HA (GA) is true in $\mathcal{M}$ at t iff A is true in $\mathcal{M}$ at any t' such that $t' < t$ ($t < t'$).

In $\mathcal{L}$ we can paraphrase "At any time, A" as HGA. This sentence, far from being an adequate paraphrase of "Eternally, A", is not even equivalent to "Endlessly, A". For when we say that A is endlessly true (in both directions in time), we do not say *just* that A is true at any time. We also say, intuitively, that for any time at which A is true there is *another* time (both toward the future and the past) at which A is true. If that is correct, an adequate rendition of "Endlessly, A" in $\mathcal{L}$ must also involve an existential quantification over times, that is, the duals P ("at some past time") and F ("at some future time") of H, G, respectively:

(5) $HG(FA \land PA)$.

---

[7] I omit quotes and corner-quotes in the mention of formal expressions for readability.

We can familiarly summarize this sentence as:

(5′)   There is always another time at which A.

Note that HGA ("At any time, A") is frame-equivalent to (5) ("Endlessly, A") only if $<$ is both left- and right-serial, that is, has neither a first nor a last element (formally: for any t there exist t', t'' such that $t' < t < t''$). The following (schemata of) theorems of $\mathcal{L}$ correspond to left- and right-seriality, respectively:

(LS) $HA \rightarrow PA$
(RS) $GA \rightarrow FA$.

Left- and right-seriality of $<$ or, equivalently, theorems (LS) and (RS) are often thought to ensure that the modelled timeline (or universe) extends indefinitely toward both the past and the future. For reasons we have discussed in the last section, however, that is false. Not to go too far, the above Zeno sequence (Z) can be easily turned into a right- and left-serial temporal frame $\mathcal{F}'$. The 30-minute timeline corresponding to $\mathcal{F}'$ can hardly be said to extend indefinitely in time, though.

10. To make sure that a temporal frame models a timeline (or a universe) of indefinitely long temporal extension, something more is needed, for instance, that a metric is defined on T (as in Koymans 1990). A metric temporal frame $\mathcal{F}_m = (T, <, d)$ can be gotten from a basic temporal frame by adding a *duration function* d from $T \times T$ to a range $\Delta \subseteq \mathbb{R}_0^+$ (metric temporal models $\mathcal{M}_m$ are defined in the obvious way). We let δ vary over numbers in Δ and t over times in T. Intuitively, for any times t, t', $d(\{t, t'\})$ is the amount of time, expressed in time units, that elapses between t and t' (if $t = t'$ then $d(\{t, t'\}) = 0$). Among other things, it is standard to require that there are no 'gaps' in the temporal structure: if $d(\{t, t'\}) > \delta$, then there exists t'' such that $d(\{t, t''\}) = \delta$. A metric frame $\mathcal{F}_m$ models a timeline (or a universe) of indefinitely long temporal duration toward the future iff, for any $\delta \neq 0$ and any t, there exist both a time that follows t and a time that precedes t by δ time units. It is only within frames that satisfy this condition that HGA and $HG(FA \wedge PA)$ are both equivalent to "Eternally, A". This is, of course, but another sign that the meaning of "eternally" cannot be fully captured in purely temporal, non-metric terms.

If we allow ourselves a few metric operators, then it is straightforward to provide an adequate formal counterpart of "Eternally, A". Let $\mathcal{L}_m$ be a language obtained by enriching $\mathcal{L}$ with operators P_δ ("δ time units in the past") and F_δ ("δ time units in the future"). Here are the relevant semantic clauses:

> $P_\delta A$ ($F_\delta A$) is true in $\mathcal{M}_m$ at t iff there exists $t' \leq t$ ($t' \geq t$) such that $d(\{t, t'\}) = \delta$, and A is true in $\mathcal{M}_m$ at t'.

Now consider:

(6) $HG(F_\delta A)$.

This statement is true iff, for any time t, there exists a time t' that is δ time units later than t, and A is true at t'. Apparently, with $\delta \neq 0$, the truth of (6) ensures that A remains true no matter how many time units (days, hours...) we wait. If we add the mirror condition $HG(P_\delta A)$ we obtain a perfectly sensible formal counterpart of "eternally", that is (with $\delta \neq 0$):

(7) $HG(F_\delta A) \wedge HG(P_\delta A)$, or simply $HG(F_\delta A \wedge P_\delta A)$.

Since we are free to choose any δ we please, we can familiarly summarize (7) as:

(7′)   There is always another day (hour, minute,...) at which A.

11. Where do all these considerations lead us? If anything, they show that the notion of temporal eternity is *metric* in nature. The idea of eternity is strictly tied to that of temporal duration. The whole of eternity may be understood as the result of an endless, stepwise process of addition. What is added are time units, that is, amounts of time that are thought to remain constant at each step.[8] When we say that A is eternally true, we are not saying that its truth is constant in time. Rather, we are saying that its truth remains constant no matter *how much* we move forward, or backward, in time.

---

[8] See Leibniz's quote at the beginning of this note. Cf. the corresponding passage in Locke's *Essays* (II.XIV): "[We come by the Idea of Eternity by] being able to repeat [the] Idea of any length of time... and add them one to another, without ever coming to the end of such addition".

The remark that temporal eternity is a metric notion is indeed a rather obvious one, and by no means original. If sometimes we tend to overlook it, it is most likely because, as said above, we are more or less unconsciously drawn to make some nontrivial assumption. Presumably, we use "endlessly" and at "at all times" as synonyms because we presuppose that the timeline is endless. And we use "eternally" and "endlessly" as synonyms because we presuppose that times are arranged in a non-Zeno, steady-paced sequence.

12. The conclusion that "eternal" does not mean the same as "with no end" is partly based on the claim that nothing eternal lasts less than one hour (see above, §8). There is a possible objection here, though. Suppose that the following sequence of events takes place: ⟨Romeo falls for Juliet, Romeo swears eternal love to Juliet, Romeo (sadly) dies⟩. We are prone to concede that, in the end, Romeo has kept his word. But then Romeo's arbitrarily short-lived love must count as eternal.

Whatever the initial appeal of this objection may be, it is easy to see how it can be addressed. If I am right, Romeo's oath is tantamount to the promise of a love that always lasts for *another* day (hour, minute...). But if such a promise is sensible at all, then the range of days (hours, minutes...) that are quantified upon must be restricted to those that lie within Romeo's future lifetime – otherwise, Romeo could only keep his word if he lived forever. (Here I am making appeal to the familiar phenomenon of contextual quantifier restriction.) Therefore, there is no contradiction in maintaining that Romeo has kept his word and, at the same time, denying that his love counts as eternal in a full, unrestricted sense.[9]

---

[9] Thanks to Andrea Bianchi for having brought this case to my attention. The same explanation may be extended to cover also promises of endless or everlasting love. I also believe that lovers' vows of this kind tend to be subtly ambiguous, and often have an additional, unrestricted but hyperbolic reading. (After all, love speech is full of exaggeration – and clichés.) The hyperbolic reading is sometimes the only possible one, as witnessed by the following excerpt from W.H. Auden's poem, *As I Walked One Evening*:

I'll love you till the ocean
Is folded and hung up to dry
And the seven stars go squawking
Like geese about the sky.

13. At this point, you might wonder how these remarks concerning temporal eternity relate to our starting point, Ernesto's claim (§2) that mathematical truths are world-insensitive. Well, I believe there is more to their connection than meets the eye. Consider:

(8)   Eternally, $1 + 1 = 2$.

Virtually all the philosophers who subscribe to "$1 + 1 = 2$" would also subscribe to (8), in at least one reading of "eternally" (if you have some nominalistic qualms with "1+1=2", just choose any substitute for (8) you deem appropriate). I assume that they mean (8) to be true in a full, literal and unrestricted sense. However, if (8) is taken to involve a temporal reading of "eternally", then (8) entails that "$1 + 1 = 2$" is true for an infinite amount of time. Therefore, (8), so understood, entails that an infinite amount of time *exists*: either the cosmos or the timeline is infinite in duration. I believe that most philosophers would be uncomfortable with drawing such factual conclusions from (8) (or an appropriate substitute). The reason is that, of course, they take truths like (8) to be supremely insensitive to the duration of our universe, the relation between our universe and time, and similar mundane conditions. Therefore, if they assert (8) as an obvious claim, not in need of substantial defence, then they should reasonably understand "eternally" therein in its timeless sense rather than in the temporal one. The same goes if they insist that (8) is a necessary truth, for it is only natural to assume that the cosmos, if eternal at all, is just contingently so.

References

Frege, G. (1969), "Logik", in H. Hermes, F. Kambartel and F. Kaulbach (eds.), *Nachgelassene Schriften*, Felix Meiner, Hamburg, pp. 1–8, translated as "Logic" in his *Posthumous Writings*, Basil Blackwell, Oxford 1979, pp. 1–8.

Helm, P. (2014), "Eternity", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, spring 2014 edition, `http://plato.stanford.edu/archives/spr2014/entries/eternity/`.

Kneale, M. (1968), "Eternity and sempiternity", in *Proceedings of the Aristotelian Society*, 69, pp. 223–238.

Koymans, R. (1990), "Specifying real-time properties with metric temporal logic", *Real-Time Systems*, 2, pp. 255–299.

Kretzmann, N. and Stump, E. (1998), "Eternity", in T. Crane (ed.), *Routledge Encyclopedia of Philosophy*, `https://www.rep.routledge.com/articles/eternity/v-1/`.

Moore, G.E. (1903), *Principia Ethica*, Oxford University Press, Oxford.

Napoli, E. (2010), "Realtà, verità e linguaggio", in M. Carrara and V. Morato (eds.), *Verità. Annuario e Bollettino della Società Italiana di Filosofia Analitica*, Mimesis, Milano, pp. 295–302.

Priest, G. (1986), "Tense and truth-conditions", *Analysis*, 46, pp. 162–166.

Williamson, T. (2002), "Necessary existents", in A. O'Hear (ed.), *Logic, Thought and Language*, Cambridge University Press, Cambridge, pp. 233–251.

Zalta, E.N. (1987), "On the structural similarities between worlds and times", *Philosophical Studies*, 51, pp. 213–239.