Guest Editor: Juraj Hvorecký

Contents

On January 12th, 2015 Professor Petr Koťátko is turning 60. We found it very appropriate to celebrate this occasion by putting forward a volume of papers that not only deals with some of his favorite topics, but also consists of submissions from his close friends and colleagues. The decision was made to choose authors from countries other than his homeland of the Czech Republic, in order to emphasize a strong standing Petr has in the international philosophical community.

We do not feel it is necessary to trace Petr's philosophical journey in any great detail. Yet few comments are in order to elucidate why this collection contains exactly papers it does. Petr's encounters with analytic philosophy can be traced to the very late 1980s, when turbulent political changes and a gradual policy of détente between the two political blocks on each side of the iron curtain created for a first time a limited space for exchanges among scholars. Prior to the Velvet Revolution, Petr spent several months in the UK, visiting top educational institutions. During this stay he got closely acquainted with analytical ways of philosophizing and be-friended many essential figures of this tradition.

When the old regime finally collapsed, Petr started to promote analytic philosophy in Czechoslovakia in a systematic way. He began to teach philosophy of lan-guage and related subjects at various institutions, became the head of the Analytic philosophy unit at the newly re-established Institute of Philosophy of the Czech Academy of Sciences, published extensively and also started to invite foreign scholars to Czechoslovakia. It was especially his organizational activity that made Prague an important center for analytic philosophy in Europe. In the early years of the transition, he has hosted many distinguished philosophers, including J. Katz, P. Strawson and W.V.O. Quine. These early visits had immense impact on the local community, as they were frequented not only by emerging analytic audience, but by then dominant continental philosophers as well. Success of the early days has moved Petr to his most celebrated project—legendary meetings in Karlovy Vary. From 1992 to 1998, Petr headed a team that put together one of the most influential series of analytic conferences on the continent. We need to be reminded that it was still during the time period when now blooming analytic philosophy was still largely restricted to the Anglo-Saxon world. Karlovy Vary Symposia became a household name for everyone working in the areas of philosophy of language and mind. Today, reading the Symposia programs, one cannot but be utterly amazed. The long list of participants not only includes most of the towering figures of the post-war analytic tradition—

Quine, Davidson, Putnam, Føllesdal, Searle—but also contains a sizeable proportion of young scholars, who were just beginning their careers and are now influencing the growth of the analytic tradition in Europe and elsewhere. We know that most of participants still greatly value their chance to witness these unique Symposia. Their recollections emphasize not only fascinating philosophical discussions, but also scenic surroundings and frequent inclusions of artistic performances. A close connection of philosophy and arts comes naturally to Petr. After all, he is also a renowned author of fiction, an avid piano player and a keen observer of visual arts. We should also mention that several volumes of proceedings from Karlovy Vary meetings have been published by important publishing houses and made the meetings and Petr himself known far beyond the circle of conference participants.

It is in the post-Karlovy Vary period that Petr's artistic interests and his philosophy have got closer together. He has moved both his philosophical focus and organizational capabilities into the wide problem of interpretation. In a series of publications and during his Prague Interpretation colloquia, which continue from the year 2000 onwards, he has demonstrated not only his mastery of the subject matter, but also a unique willingness to communicate with leading representatives of the literary theory. These gatherings, smaller in size, have attracted leading figures from aesthetics, philosophy of language and literary theory. Many meetings have again resulted in a publication of proceedings that are widely used as reference books within the field.

Contributions in this volume mirror the above-mentioned two stages of Petr's professional development. The first six papers discuss various issues in the field of the philosophy of language, from the notion of sense through truth conditions to the concepts understanding and intention. The final four papers attempt to solve issues on the borderline between philosophy and literature, with three discussing the nature of fictional characters, a topic Petr has been working on very recently. All contributions were written specifically for this occasion and many provide direct responses to selected issues Petr has addressed in his own work.

We would like to thank all the contributors for making the volume possible, to the editors of Organon F for making it all happen and to Petr for inspiring us to be better philosophers and better people.

In the closing, and on behalf of all the contributors to this volume, his co-workers, friends and collaborators, we want to wholeheartedly congratulate Petr on this special occasion and wish him all the best in both his professional and personal lives. You have been an inspiration, a great colleague and a valuable friend to all of us and we hope to continue these relations for many more years.

Juraj Hvorecký

Intending to be misinterpreted[1]

For the occasion of Petr Koťátko's 60th anniversary

PETER PAGIN

Department of Philosophy. Stockholm University
SE – 10691 Stockholm. Sweden
peter.pagin@philosophy.su.se

ABSTRACT: In his paper 'Two Notions of Utterance Meaning', Petr Koťátko criticises Davidson's conception of the relation between meaning and intention. He ascribes the following view (D) to Davidson: "If S makes an utterance in order to perform a certain speech act, he intends and expects that act to be assigned to the utterance in A's interpretation". Koťátko's objection to (D) is that a speaker can intend to be misinterpreted. The present paper discusses this objection. It is argued that Koťátko's main example of such an intention fails. It is also argued that although there can be cases that would be adequately described as examples of intending to be misinterpreted, they are not of the kind needed for an objection against (D).

KEYWORDS: Intention – meaning – misinterpretation – utterance.

1. Koťátko on Davidson

In his celebrated *Aristotelian Society* paper 'Two notions of utterance meaning' (Koťátko 1998), Petr Koťátko designs an argument, both intricate and intriguing, against a certain view of *utterance meaning*. He ascribes the view he attacks to Donald Davidson, in particular as expressed in Davidson (1986) and in Davidson (1994). I find his representation of David-

---

son's views fair and shall not have any quarrel with him on that score. In this section I shall present Koťátko's discussion of Davidson's view. In the following section, I shall examine his main line of criticism, which depends on the idea that a speaker can *intend to be misinterpreted*.

Koťátko summarizes Davidson's view as follows:

> [...] the principle of determination of meanings of particular utterances: the utterance has the propositional content *p* and illocutionary force *f* if and only if it was so meant by the speaker and understood by the audience. (Koťátko 1998, 225-226)

Later on (1998, 230), Koťátko specifies two Davidsonian conditions about the speaker's intentions and the audience's interpretation. Here *S* is the speaker and *A* the audience:

Condition (*a*)   If *S* makes an utterance in order to perform a certain speech act, he intends and expects that act to be assigned to the utterance in *A*'s interpretation.

Condition (*b*)   *A* interprets the utterance in the way which he believes to have been intended by *S*.

Condition (*b*) is perfectly in order. There is a question, however, with respect to Condition (*a*). To what extent can a speaker *intend* the audience to interpret her this way or that? By normal standards, you can intend to do only what is under your own control to achieve. You cannot intend to win a (fair) lottery. You can intend to win a race only if you are certain that you will win if you try. By the same token, you can intend *A* to interpret you in a particular way (assign a speech act to your utterance), only if you are certain that there is an utterance you can make that will cause *A* to interpret you that way. In case you are not certain, a weaker alternative appears more appropriate:

Condition (*a'*)   If *S* intends to perform a certain speech act *I* by means of an utterance *u*, then *S* expects that *A* will assign *I* to *u*.

Because you can expect something even if not being certain that it will happen, Condition (*a'*) can be true even if Condition (*a*) is false. However, since Koťátko's objection will turn out to be an objection equally much against Condition (*a'*), I shall set this issue aside.

Kot'átko discusses two versions of intending and interpreting the use of a proper name, in his example 'Woody Allen'. In both versions there are two possible referents to choose between, X and Y. The first version assumes that there is a standard, or real, bearer of the name, Y, while the second version just has two alternatives of equal standing.

In his treatment of the first version, Kot'átko rightly assumes that speaker and hearer, in trying to conform to the two Davidsonian conditions, will take into account their beliefs about each other, including beliefs about the other's beliefs. He shows that beliefs that ascribe higher-order false beliefs will have effects on actual intending and interpreting, under the Davidsonian assumptions.

For instance, suppose that S believes that W, the referent of the name 'Woody Allen', is Y, i.e. that $W = Y$ ($B_S p$, in Kot'átko's abbreviation), but believes that A believes that $W = X$ ($B_S B_A q$). Since S intends to refer to what she expects A will interpret her as referring to, she will refer to X, not to Y. Similarly, if S believes that A believes that $W = Y$ ($B_S B_A p$), but also believes that A believes that she herself, S believes that $W = X$ ($B_S B_A B_S q$), the result will be the same. For, since she believes that $B_A B_S q$, S expects A to interpret her as referring to X, since she expects that A will interpret S in accordance with what A expects S to intend, which will be to intend to refer to X. And so on. Beliefs in higher-order mistakes have consequences for actual intentions and interpretations.

Note that, as Kot'átko points out, this does not depend on the *truth* of these higher-order beliefs. S and A may have false higher-order belief about each other, but still adapt intention and interpretation to the effect that communicative success results. Kot'átko sums up the result as follows, correctly as far as I can judge:

> This shows that, according to the Davidsonian account of utterance meaning, an utterance including a proper name can be an assertion about some person X, even if the name uttered is not the name of X in the community to which S and A belong, S does not believe that A regards X as a bearer of that name (neither in 'official' nor in any other sense), A does not believe that S regards X as a bearer of that name, S does not believe that A believes that S regards X as a bearer of that name, etc. And the utterance has this meaning even if S's relevant beliefs do not correctly represent S's relevant beliefs and A's relevant beliefs do not correctly represent S's relevant beliefs. (Kot'átko 1998, 230-231)

Is this result disastrous for the Davidsonian account? Should we rather judge communication to *fail* if S and A base their intentions and interpretations on false mutual beliefs, even if they end up with the same referent? Kot'átko does not claim so, and I think that would be an unjustified conclusion. Note that although S and A may have false beliefs about each other, these beliefs only concern attitudes about who is the *standard* referent of the name 'Woody Allen'. In case the referent *is* shared, S does *not* have a false belief about how A will interpret S's utterance, and A does *not* have a false belief about what S intends to refer to. On Davidson's model, those beliefs are true. And if S has a rational belief about what A believes about S's intention, that belief will be true as well, for otherwise A will have a false belief about S's intention to refer, and the referent will not be shared. And so on. In exceptional cases, as in Kot'átko's example, the true higher-order beliefs about intentions and interpretations may be based on false higher-order beliefs about standard reference, but in general, taking into account what one's interlocutor does believe about standard reference seems like a good idea.

This is actually pretty close to Kot'átko's own conclusion, for he observes that we do get a different result if we leave standard reference out of the picture, and just focus on mutual beliefs about intentions and interpretations. This is Kot'átko's second version of the example. In the second version, the question of the standard or real referent of the name plays no role. Instead of the propositions $W = X$ and $W = Y$, there are four alternatives (1998, 232):

(p_1) If S utters 'Woody Allen' (on a given occasion), he intends to refer to Y.

(p_2) S's utterance of 'Woody Allen' (on a given occasion) would be interpreted by A as referring to Y.

(q_1) If S utters 'Woody Allen' (on a given occasion), he intends to refer to X.

(q_2) S's utterance of 'Woody Allen' (on a given occasion) would be interpreted by A as referring to X.

Kot'átko now observes that there is no longer any room for false higher-order beliefs as in the first version, for those false beliefs all concerned standard reference. In this case, any false mutual belief would, if S and A are rational, lead to not sharing reference, just as the beliefs about intentions and interpretations in the first version. Kot'átko notes that the following belief is not consistent with Davidson's conditions:

(1) $B_S p_2 \ \& \ B_S B_A q_1$

If (1) is true, S believes that he will be interpreted as referring to Y but *also* believes that A believes that S will intend to refer to X. But if S's second-conjunct belief is true, and A thus expects S to refer to X, then by *Condition (b)*, A will interpret S as intending to refer to X, and so S's first-conjunct belief is false. So, holding both beliefs conjointly is not coherent.

This looks like a welcome result for Davidson, but Koťátko has a real objection: it is possible to *intend to be misinterpreted*. This would be a direct violation of *Condition (a)*. The next section will be concerned with this idea.

2. Koťátko's counterexample

Koťátko's example, the *Martial Example*, is the following:

> Let us imagine that Paul says to John: 'Martial wrote witty epigrams'. He hopes that John, due to embarrassing gaps in his education, will interpret him as asserting that Martial wrote witty epitaphs and that this will come to light in John's reaction: that would provide a welcome opportunity to give John a lesson in literary terminology. I think the most natural thing to say here is that Paul wants to be misinterpreted, which means: there is a discrepancy between what he wants to assert and what he wants to be taken as asserting. Then the condition (a) is not fulfilled. And Paul can very well succeed in both respects: even if he is interpreted as he wanted, we shall, I think, say that he asserted that Martial wrote witty epigrams (this is also what he is going to later explain to John). This is certainly something which the Davidsonian notion of utterance meaning does not allow us to say. Now imagine that John knows what 'epigram' conventionally means and that he also sees through the trick intended by Paul: then he will obviously interpret the utterance in the standard way, even if he knows that this is not the way John wanted him to interpret it (and it will be quite natural if he manifests to John that he understood him correctly, i.e. not as John wanted). In that case the condition (b) is not fulfilled. (Koťátko 1998, 234-235)

The questions that immediately present themselves are of course these: Is the Martial Example an example of intending to be misinterpreted? And: Is

the Martial Example an example of interpreting contrary to how you take the speaker's intention?

Let's focus on the first question. The initial point about what you can intend is clearly relevant: according to the example, Paul *hopes* that John will interpret 'epigram' as *epitaph*. It is questionable whether you can intend something to happen that you can merely hope to happen. But maybe the example can be reconstructed so as to avoid this problem. I'll set it aside.

My objection against Kot'átko's alleged counterexample is that the object of intention is misdescribed. In order that John's reaction will show "embarrassing gaps in his education", what is relevant is not *which* meaning Paul intends to express by means of 'epigram', but that he intends to express *the standard meaning* (in their speech community), whatever it is. In order that John reveal any gap in education by his interpretation, it is necessary that he does interpret 'epigram', as uttered by Paul, as expressing *its standard meaning*, whatever it is. The mere fact that Paul would intend *epigram* and John interpret the word as meaning as *epitaph* does not by itself reveal any gap in education. Without reference to standard meaning, it is a mere case of failed communication.

Taking this into account, how should the speaker's intention and the audience's interpretation be described in the Martial Example? One alternative is to distinguish between a *primary* and *secondary* intention and interpretation. We can then describe it as follows.

(Paul)   i)   Paul primarily intends to express the standard meaning of 'epigram'.
          ii) Paul believes that the standard meaning of 'epigram' is *epigram*.
         iii) Hence, Paul secondarily intends 'epigram' to mean *epigram*.

We can make an analogous derivation for John:

(John)   i)   John primarily interprets 'epigram' as having its standard meaning.
          ii) John believes that the standard meaning of 'epigram' is *epitaph*.
         iii) Hence, John secondarily interprets 'epigram' to mean *epitaph*.

If (Paul) and (John) are adequate descriptions of Paul's and John's linguistic attitudes in the Martial Example, then it turns out that it does not after all constitute a counterexample to Davidson's conditions. For Paul primarily intends *epigram* to mean just what he expects Paul to primarily interpret it as meaning, viz. its standard meaning, in accordance with *Condition (a)*. And John primarily interprets 'epigram' to mean exactly what John primarily intends it to mean, in accordance with *Condition (b)* (at least, this is in accordance with *Condition (b)* if John also believes Paul to intend the standard meaning, but we may assume that).

On this analysis of the Martial Example, we still have a desired secondary misinterpretation. Paul believes (John-ii), and can therefore, on the assumption of (John-i), derive (John-iii), which Paul desires to be made true, and perhaps also intends it to be made true. Does this violate Davidson's conditions?

It is unclear what Davidson would or should have said, based on his published writings. The distinction between primary and secondary meaning intention does not occur there, and is somewhat alien to his way of thinking. The distinction in Davidson (1986) between *prior theory* and *passing theory* is different, which concerns replacing one theory by another, in interpretation, regardless of whether either is standard: the prior theory might well have been an idiosyncratic scheme of interpretation for a particular speaker, even if it be replaced by a scheme that is perhaps even more idiosyncratic.

Davidson would clearly have recognized the possibility of devious calculations such as in the Martial Example, but it is not obvious how he would have related these to his views about intentions and interpretations. There is some reason think that he would have insisted that agreement in primary intention and interpretation is what matters. One reason is that in virtually all normal linguistic communication, the primary intention is all that matters, because speakers normally directly intend to express a certain meaning, and similarly for hearers. A second reason is that a Davidsonian might insist that any desire for a misalignment between secondary attitudes of speaker and audience *presuppose* an alignment between their primary attitudes. A third reason is that secondary attitudes depend on certain background beliefs about other factors than the speaker's intention and the hearer's interpretation, in this case standard meaning, and such background beliefs did not play a significant role in Davidson's account. The conditions on intention and interpretation should then concern primary meaning attitudes.

A consequence of this choice is that, on Davidson's view, as set forth in the first Kot'átko quote, and given the primary intention in (Paul) and the primary interpretation in (John), the utterance meaning of 'epigram' is its standard meaning, *epigram*. This would follow even on a variant of the Martial Example where both Paul and John are misinformed about the standard meaning (maybe Paul thinks it means *epigraph*). Is this reasonable? Let's compare it to the outcome of an alternative analysis of the Martial Example. On this analysis, the beliefs about the standard meaning only serve to motivate and explain the primary intention:

(Paul′)   i) Paul desires that his primary meaning intention with respect to 'epigram' coincides with its standard meaning.
           ii) Paul believes that the standard meaning of 'epigram' is *epigram*.
           iii) Hence, Paul primarily intends 'epigram' to mean *epigram*.

(John′)   i) John desires that his primary interpretation of 'epigram' coincides with its standard meaning.
           ii) John believes that the standard meaning of 'epigram' is *epitaph*.
           iii) Hence, John primarily interprets 'epigram' to mean *epitaph*.

On this alternative analysis, if both Paul and John are mistaken about the standard meaning, to the effect that both believe that it means *epitaph*, the utterance meaning, on according to Davidsonian principles, would be *epitaph*, not *epigram*. The desire to agree in intention and interpretation with standard meaning would cause them to *revise* their primary intentions and interpretations, should they learn the truth about the standard meaning, but before this has happened, they each means and interpret what they *believe* is the standard meaning, even if the standard is different from what they believe.[2]

---

[2]   This discussion runs parallel to the discussion of social externalism following the work of Tyler Burge (starting with Burge 1979). The original analysis corresponds to Burge's own externalism, where a speaker who defers to the community actually means what the community means, even if he is mistaken about it, and therefore mistaken about what he himself means. The second analysis corresponds to an alternative way of understanding what deference to the linguistic community amounts to: a readiness to

The alternative analysis, (Paul′) and (John′), corresponds better, I think, to how ordinary speakers relate to their speech community. If this would also be the correct analysis of the Martial Example, it clearly would provide counterexamples to Davidson's principle: Paul now intends/desires John's primary interpretation to be one that does not agree with his own primary meaning intention. The question is whether it does justice to the example.

According to the example, Paul is *primarily* interested in exposing John's lack of knowledge of the *standard* meaning of 'epigram'. He also believes, correctly, that he himself knows the standard meaning, and that therefore, if he intends to express the standard meaning, John's misunderstanding Paul will be equivalent to having a false belief about the standard meaning. But in the variant scenario, where Paul himself is *wrong* about the standard meaning, John's misunderstanding of Paul will be irrelevant to the truth value of John's belief about the standard meaning. Therefore, I think the alternative analysis, (Paul′) and (John′), does not really capture the scenario in the Martial Example.

Since we therefore should keep the first analysis, and the example is taken to concern Davidson's conditions with respect to primary meaning intention, Koťátko's Martial Example is not really a counterexample. Could there be others?

3. Intending a misalignment of indexicals

Could there be counterexamples to Davidson's principles that do not depend, as the Martial Example does, on a discrepancy between the audience's interpretation and standard meaning? Two types of possible discrepancies come to mind. One concerns a possible misalignment in the interpretation of indexicals, and the other in disambiguation (of various kinds). These are not really different in principle, from the current perspective, and I shall focus on the indexical case.

In the use of indexicals, the speaker could aim at misleading the audience by trying to make him assign the wrong value. Consider the case where *S* writes an email to *A*, saying

---

change once intentions and interpretations in the light of new information, but not an externalism about what one in fact means.

(2)   I am now in Buenos Aires.

What does 'now' in this message refer to? Does it refer to the time of the context of S's writing, or the time of the context of A's reading?[3] Suppose that S is sending the email late in the evening, counting on it's not getting read until the morning. He knows that A does not expect him to arrive in Buenos Aires until the morning, but in fact S has arrived already the evening before. He wants no questions to be asked about it, but figures it will be too dangerous to plainly lie. Therefore, he hopes that A will interpret S as referring, by 'now', to the time of reading, i.e. the morning. His plan is further, in case he is explicitly asked about it, to say that by 'now', he referred to the time of writing, i.e. the evening. Is it adequate to describe S as meaning *the evening*, but intending to be interpreted as meaning *the morning*, i.e. as intending to be misinterpreted?[4]

As with respect to the Martial Example, I think this would be an inadequate description of the example. When we use indexicals and demonstratives, we take them to refer to what is cognitively salient in the context. With demonstratives, typically but not always, an entity is made salient by means of a demonstration. The pure, or automatic, indexicals, including 'I', 'here', and 'now', have as default referents the speaker, the location, and the time, of the context of utterance, but in exceptional cases, the default interpretations are overridden.[5] In case of 'now', the default can be overridden e.g. in the historical present tense of narratives, and, as in the current example, in messages with an expected time delay between the speaker's production of the utterance and the hearer's perception. This latter case is not,

---

[3]   These alternatives in the interpretation of temporal indexicals, and especially 'now', are discussed in the literature on the so-called 'Answering Machine Paradox', typically in connection with Kaplanian semantics. For an overview, see Cohen and Michaelson (2013).

[4]   There is a fairly wide discussion, starting with Kaplan himself, of what indexicals *really* refer to in tricky cases, in particular if it is the speaker's intention that decides, or something else. For instance, see Gauker (2008) and Åkerman (2009). This discussion is, however, irrelevant to present concerns. The present question concerns the possibility for the speaker to combine two intentions, regardless of what the "real", or "correct", reference might be.

[5]   For 'I' there are descriptive uses (cf. Nunberg 1993, 20-21), and in some languages, including German, the first person singular pronoun also has impersonal uses (cf. Zobel 2010), where English predominantly uses the second person.

however, a counterexample to the principle that the time referred to is the salient time. It is only that in these exceptional cases, a time other than the time of utterance is salient.

The term 'context' is ambiguous between the real situation of utterance and what is *cognitively relevant* in the situation of utterance. In the latter sense, it can be seen as a collection of values given to semantically relevant parameters. Kaplan (1989a) used the term in the latter sense, and made it explicit in Kaplan (1989b, 591-593). David Lewis used it in the former, a location "in physical space-time and in logical space" (Lewis 1980, 85). The need for the distinction arises because a real situation, in the sense of the objective features of the environment of a speaker or interpreter at a time, do not uniquely fix all relevant parameters.[6] It does not automatically determine what is cognitively salient to a speaker or interpreter, even though certain perceptible features can have a strong influence (a very big dog surrounded by small dogs is likely to be the referent of 'that dog' unless preceding discourse leads in another direction).

In particular, what is salient to the speaker may not be what is salient to the audience, which may be a source of misunderstanding in the use of an indexical or demonstrative. Therefore, in the collection-of-parameters sense of 'context', we strictly speaking need to distinguish between the *speaker's context* and the *audience's context*, even if speaker and audience are in the same objective situation, i.e. in roughly the same place at the same time. The need is even greater if they are not, as in the example.

The speaker refers by 'now' to the time that is salient to the speaker at the time of utterance, t_S. This statement be taken as an axiom concerning the relation between salience and indexical reference. Lacking an independent precise definition of *salience*, it may also be taken as part of what characterizes our notion of salience. This notion at least also involves the idea of *having in mind*; what is salient is what most strongly attracts attention in a collection of candidates.[7]

---

[6]   Lewis preferred the former sense because contexts have "countless features", not given by a fixed list.

[7]   In vision research, visual salience is regarded as the result of a combination of visual-phenomenal and visual-"semantic", i.e. conceptually categorized, information, with the property of attracting selective attention. What does get attention does not only depend on properties of the stimulus, but also on the cognitive state. See e.g. Parkhurst, Law and Niebur (2002).

The audience interprets 'now' as referring to the time that is salient to the audience at the time of perception, t_A. To this we should add that both speaker and audience *presuppose* that they have the same time in mind, i.e. that $t_S = t_A$. If the presupposition fails, there is misunderstanding.

Applying Davidson's conditions to indexical time reference yields the following:

Condition $(a)_T$ If S by 'now' refers to time t, he expects that t is salient to A.

Condition $(b)_T$ If A interprets S's utterance of 'now' as referring to time t, he believes that t is salient to S.

The idea that S in the Buenos Aires Example intends to be misinterpreted amounts to a violation of *Condition* $(a)_T$. But this idea does not cohere with the assumptions about salience. In uttering (writing) (2) with the intention that A takes 'now' to refer to the time of reading, S must have the time of reading in mind. That time must then be salient to S. But with the time of reading salient in uttering the indexical, that is also what S intends the indexical to refer to. The idea of intending to be misinterpreted leads to a conflict in assumptions about the salient time.

It may be objected that S may have just a prior plan about interpretation. That is, S's intention about A's interpretation can be formed *before* the utterance, while the utterance itself only has the time of utterance as the intended referent. This would be unusual, but not impossible. However, if normal utterances, without prior plans, are made with intentions or desires about interpretation, during the time of the utterance itself, then this would hold for S's utterance in the Buenos Aires Example as well. If so, the prior intention is irrelevant, for it is the simultaneous intention that matters, and then the conflict would arise anyway.

This is a simple argument, and not unassailable, but a rejection needs to appeal to yet unmotivated complications, such as having multiple alternative times salient in parallel, for separate but simultaneous mental acts, or having a sequence of acts associated with a simple one-word utterance. Without very good independent reasons for such complications, they are not serious candidates. What goes for indexicals carries over, I think, mutatis mutandis, to lexical and structural disambiguation, and anaphora resolution.

Still, this argument leaves one possibility open: that S has a prior plan about A's interpretation, but at the time of writing means *nothing at all*,

since he has neither the time of writing nor the time of reading in mind. He just writes down the sentence according to his prior plan. This seems to me a description of the example that is both coherent and plausible. However, although the utterance is made with an intention do mislead the audience, it does not violate *Condition* $(a)_T$, or *Condition* (a), since the antecedents are false: S does make an utterance, but not in order to perform any speech act. He just wants A to interpret it as a speech act anyway.

Can S nevertheless be reasonably described as intending to be misinterpreted? Strictly speaking not, as long as being *mis*interpreted amounts to being interpreted in a way that diverges from what is meant, and by assumption nothing was meant. S can certainly be correctly described as having intended to be interpreted in a way that will *not agree* with anything that S will mean (on the occasion). On the other hand, to the extent that S does not intend to mean anything at all, he can also be described as intending to be interpreted in a way that will *agree* with everything that S will mean (on the occasion).

The final possibility to consider is whether the description might apply solely to S's prior plan. For perhaps S mistakenly *believes* that he can write down sentence (2), while both meaning the evening and intending to be interpreted as meaning the morning. Doesn't S then, in forming this plan, intend to be misinterpreted? I find this hard to deny. If this description is correct, there can after all be states of mind to which the phrase "intending to be misinterpreted" applies. They are not, however, states of the kind Koťátko intended.

References

Åkerman, J. (2009): A Plea for Pragmatics. *Synthese* 170, 155-167.

Almog, J. – Perry, J. – Wettstein, H. (eds.) (1989): *Themes from Kaplan*. Oxford: Oxford University Press.

Burge, T. (1979): Individualism and the Mental. In: French, P.A. – Uehling, T.E. – Wettstein, H.K. (eds.): *Midwest Studies in Philosophy, 4*. Minneapolis: University of Minnesota Press, 73-121.

Cohen, J. – Michaelson, E. (2013): Indexicality and the Answering Machine Paradox. *Philosophy Compass*, 570-579.

Davidson, D. (1986): A Nice Derangement of Epitaphs. In: Lepore, E. (ed.): *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell.

DAVIDSON, D. (1994): The Social Aspect of Language. In: McGuinnes, B. – Oliveri, G. (eds.): *The Philosophy of Michael Dummett*. Dordrecht: Kluwer Academic Publishers. Reprinted in Davidson (2005), 109-125.

DAVIDSON, D. (2005): *Truth, Language, and History*. Oxford: Clarendon Press.

GAUKER, C. (2008): Zero Tolerance for Pragmatics. *Synthese* 165, 359-371.

KAPLAN, D. (1989a): Demonstratives. In: Almog, Perry and Wettstein (eds.) (1989), 481-563.

KAPLAN, D. (1989b). Afterthoughts. In: Almog, Perry and Wettstein (eds.) (1989), 565-614

KOŤÁTKO, P. (1998): Two Notions of Utterance Meaning. *Proceedings of the Aristotelian Society* 98, 225-239.

LEWIS, D. (1980): Index, Context, and Content. In: Kanger, S. – Öhman, S. (eds.): *Philosophy and Grammar*. Dordrecht: Reidel, 79-100. Reprinted in Lewis (1983).

LEWIS, D. (1983): *Philosophical Papers. Volume I*. Oxford: Oxford University Press.

NUNBERG, G. (1993): Indexicality and Deixis. *Linguistics and Philosophy* 16, 1-43.

PARKHURST, D. – LAW, K. – NIEBUR, E. (2002): Modeling the Role of Salience in the Allocation of Overt Visual Attention. *Vision Research* 42, 107-123.

ZOBEL, S. (2010): Non-Standard Uses of German 1st Person Singular Pronouns. In: K Nakakoji, K. – Murakami, Y. – McCready, E. (eds..): *JSAI-isAI, LNAI 6284*. Berlin – Heidelberg: Springer, 292-311.

Heidegger's Logico-Semantic Strikeback

ALBERTO VOLTOLINI

Department of Philosophy and Educational Sciences. University of Turin
Via S. Ottavio 20. 10124 Turin. Italy
alberto.voltolini@unito.it

ABSTRACT: In (1959), Carnap famously attacked Heidegger for having constructed an insane metaphysics based on a misconception of both the logical form and the semantics of ordinary language. In what follows, it will be argued that, once one appropriately (i.e., in a Russellian fashion) reads Heidegger's famous sentence that should paradigmatically exemplify such a misconception, i.e., "the nothing nothings", there is nothing either logically or semantically wrong with it. The real controversy as to how that sentence has to be evaluated—not as to its meaning but as to its truth—lies at the metaphysico-ontological level. For in order for the sentence to be true one has to endorse an ontology of *impossibilia* and Leibniz's principle of the identity of indiscernibles.

KEYWORDS: *Impossibilia* – logical form – metaphysics – nothing – ontology – *possibilia* – theory of descriptions.

1. Introduction

As is well known, in (1959) Rudolph Carnap took some sentences by Martin Heidegger (1977) to be representative of typical examples of metaphysical pseudosentences, i.e., meaningless sentences that are such not only because they contain subsentential terms that are meaningless, but also because, although they are grammatically well-formed, they are logically ill-formed, i.e., they violate logical form or syntax.

The most interesting case discussed by Carnap is a sentence Carnap ascribes to Heidegger:

(1)   The nothing nothings

for it paradigmatically exhibits both flaws.[1] According to Carnap, (1) indeed is not only meaningless because it contains the meaningless predicate "nothings", but also because from a logical point of view the (quasi-homophonous) expression "nothing" occurs in it misleadingly. For as regards logical form, this expression is not used in (1) in the only logically correct way it may be used; namely, as a quantifier, as it occurs in negative existentially quantified sentences of the kind "there is nothing that Fs", "$(\sim\exists x)(Fx)$". Rather, it pretends to be used as if it were a singular term. Thus, it precisely contributes to yielding a logically ill-formed sentence. From Carnap onwards, in some philosophical quarters that sentence or a close reformulation of it has become the paradigm of a nonsensical sentence.[2]

In what follows, it will be argued that Carnap was wrong on both counts. That is, it will be first of all claimed not only that (1) *may* be read in a logically correct way, but also that such a way is its *suggested* reading. So, once (1) is *appropriately* understood, there is no problem with its logical form. Moreover, it will be claimed that the predicate "nothings" is definitely meaningful. For the present purposes, this is enough. Yet one may even say that, if one endorses certain metaphysico-ontological views about impossible *denotata* and their identity, (1) can turn out to be not only meaningful, but also true. Admittedly, these views are rather controversial, as we will see soon below. Yet the controversy on this concern precisely regards metaphysico-ontological preferences, not semantic issues. Carnap's challenge is won if (1) is both logically well-formed and meaningful, regardless of whether it is true.

---

[1]   I say "ascribes" for Heidegger's real sentence is "the nothing itself nothings" [Das Nichts selbst nichtet]. In point of fact, Carnap (1959, 69) includes (1) as the last item in a long list of sentences from Heidegger (1977) that for him should be ranked as metaphysical pseudosentences. Yet the other sentences of the list are less representative than (1) in their being metaphysical pseudosentences. For they allegedly exhibit just the first kind of flaw—i.e., they allegedly violate logical form only. Moreover, the reason why such sentences allegedly violate logical form is for Carnap the same as with (1), i.e., they again contain the expression "nothing" (or some cognate expressions) as illogically used. Thus, the way out of this 'no logical form'-charge it will be here provided for (1) also applies to these other sentences.

[2]   See for instance the way Dennett (1987, 164) makes reference to the sentence "Nothing noths".

Before starting, a caveat. The present reconstruction of the situation at issue is meant to be theoretical, not historical. First of all, there is no aim to show that the present reconstruction of (1) displays what Heidegger had in mind in uttering it. Heidegger himself might well have endorsed Carnap's remark that, like any other relevant "nothing"-sentence he mobilizes in (1977), (1) lies beyond logic.[3] Nevertheless, it may well be the case that Carnap failed to grasp Heidegger's concern with the Nothing in its predicament of what is ruled out by whatever there really is, so that facing such a nothingness is what induces anxiety in human beings. Be that as it may, (1) will be here considered in abstraction from Heidegger's real concern with it. Rather, we will perform a theoretical exercise in order to check whether (1), taken as such, is really illogical and meaningless. Moreover, there is no intention to face the historical problem of whether Carnap might have found plausible the logico-semantic apparatus that will be here mobilized in order to retrieve both logical well-formedness and meaningfulness for (1). In conformity with our theoretical exercise, it will rather be assumed that Carnap raises a theoretical challenge concerning the very sentence (1) (and its cognates). As such, this challenge goes beyond Carnap's own logico-semantic preferences (involving e.g. his dislike for Russell's theory of descriptions). Once again, the question such a challenge raises is the following, rather abstract, one: *in itself*, is (1) really a logically ill-formed and meaningless sentence?

As a result of this way of putting things, if on behalf of a Heidegger-oriented philosopher (a certain fictional character: the Heideggerian, for short) one can show that, *qua* paradigmatic case of an alleged metaphysical pseudosentence, (1) rather is, once appropriately read, both logically well-formed and utterly meaningful, the burden of the proof is on a Carnap-oriented philosopher (another fictional character, although closer than the Heideggerean to the real philosopher that inspires it, Carnap himself: the Carnapian, for short) to show either that, appearances notwithstanding, this is still not the case or that such a reading is either just a mere, rather arbitrary, option or anyway irrelevant.

---

3   As Carnap himself underlies, by quoting again Heidegger (1977). Cf. Carnap (1959, 71).

2. Logical well-formedness

Let me start from the first problem envisaged above. Is it that clear that Heidegger's sentence is meaningless mainly because it is logically ill-formed? To begin with, suppose the Carnapian claimed that the Heideggerian logically misunderstands a sentence merely beginning with "nothing" in a subject-position by committing the Polyphemousian mistake of taking that expression as a singular term rather than as a quantifier expression. Yet this claim would contain a gross error.[4] For note that (1) begins not with the mere nominal syntagm "nothing", but with a definite description containing that syntagm—"the nothing". In terms at least of its surface grammar, this definite description works as any other such description in which a noun is prefixed a definite article (like "the book", "the dog", etc.). Given this predicament, the Carnapian might well address the Heideggerian with a Russellian accusation, namely by charging her to have misleadingly equated (1)'s grammatical form, a subject-predicate form, with its logical form, where that definite description is eliminated away.[5] Yet the Carnapian cannot charge the Heideggerian with the accusation of having produced a sentence that is logically ill-formed for it illogically mispresents a quantifier as a singular term.

Curiously enough (for what counts: see the Introduction's end), Carnap himself seems not very far from committing the above gross error. For on the one hand, he rightly reports Heidegger's *dictum* as (1), hence as starting with "*the* nothing" rather than with "nothing". Yet on the other hand, he addresses Heidegger with the joint accusation of both having produced a logically ill-formed sentence and having mistakenly maintained that "the nothing" works in it as a singular term, in such a way that the logical form of (1) coincides with its grammatical form, a subject-predicate form: "No(no)", as Carnap himself writes.[6] Yet those two things cannot stand to-

---

[4] As is well known, in *the Odissey* Ulysses leads the cyclop Polyphemus astray by letting him mean "Nobody blinded me" as a sentence containing "nobody" as a proper name and so forcing him to be misunderstood by his fellows, who correctly take "nobody" as a quantifier.

[5] Cf. Russell (1905a). As we will immediately see, however, it would be wrong to address *the Heideggerian* with this accusation. Instead, one may say that perhaps Meinong (1960), or rather the early Russell (see later), implicitly conceived "the nothing" as a singular term such that (1)'s logical form is its grammatical subject-predicate form.

[6] Cf. Carnap (1959, 70).

gether. One such joint accusation would be correct only if (1) merely contained the nominal syntagm "nothing", not the definite description "the nothing". But since (1) contains "the nothing", the only accusation that Carnap might have addressed to Heidegger is that he mistook (1)'s logical form by equating it with its grammatical form,[7] but not that (1) has no logical form at all.[8]

The above observations can be pursued a bit further. Once the Heideggerian positively endorses that a definite description occurs in subject-position in (1), she can first of all show that no logical ill-formedness occurs in (1). Moreover, by reconstructing (1)'s proper logical form, she can positively account for the idea that (1) contains an expression in subject-position that can logically function only insofar as it serves a quantifying purpose. Since this was the idea lurking behind Carnap's main reason for rejecting (1)'s well-formedness, the Heideggerian may well expect that the Carnapian charge that (1) is logically ill-formed is adequately countered. Let us see.

To begin with, on behalf of the Heideggerian suppose first that the definite description "the nothing" fits a Russellian account of definite descriptions, as any other such description. Second, suppose that the predicative element of that description expresses the property of *being identical with nothing*, or in other terms, of *being a thing such that there is no thing that is identical with that thing—$\lambda x((\sim\exists y)(y = x))$*. If this is the case, insofar as it figures in that description, "nothing" really serves a quantifying purpose.

---

[7]   That Carnap well acknowledges to (1): cf. Carnap (1959, 69).

[8]   While reading (1959), it repeatedly seems as if Carnap saw no difference in the fact that the relevant sentence merely contains "nothing" or "the nothing" in subject position. Few paragraphs later, Carnap says that the sentence "*the nothing* exists" would be meaningless even if it were logically correct that "*nothing*" in it worked as a singular term (my italics). For that sentence would simultaneously ascribe existence and non-existence to the entity allegedly denoted by the relevant singular term. As we will see later (cf. fn. 14), once (1) is appropriately understood, there is no problem in yielding "the nothing exists" an appropriate reading as well. Independently of this, however, Carnap was perhaps misled by the fact that Heidegger often capitalizes "nothing" (as we have seen, the German original for (1) is "Das Nichts (selbst) nichtet"); as if Carnap's opponent already considered "nothing" taken in isolation as a singular term, so that prefixing it by the definite article were for him simply a way to stress again that consideration. (In point of fact, we will later see that whenever a proper name "N" is prefixed by a definite article so as to get the nominal description "the N", its semantic value changes. But let us put this aside.)

Yet this is not the purpose "nothing" fulfils when it occurs alone as a nominal syntagm in any sentence of the kind "nothing Fs", namely, the purpose of letting that sentence say that there is no thing that instantiates the property expressed by the predicate "F". Rather, the quantifying purpose in (1) shows itself in the fact that the predicate in terms of which the description "the nothing" is Russelleanly analyzed contains a negated existential quantifier, so as to express the property of being a thing such, that *there is no thing* that is identical with that thing. So, when that description is analysed *à la* Russell, (1) turns out to be equivalent with:

(1R)   What is identical with nothing nothings.

This formally becomes:

(1FR) $(\exists x)((\sim\exists y)(y=x) \wedge (\forall z)((\sim\exists y)(y=z) \rightarrow (z=x)) \wedge Nx)$

where the quantifying purpose served by "nothing" is given by the *second* existential quantifier—"$\exists y$"—in the formalized sentence, the quantifier contained in the predicate by means of which the definite description "the nothing" is eliminated away. Now, once (1) is analyzed as (1FR), no mistake in logical form arises. Thus, Carnap's first claim that Heidegger's sentence is logically ill-formed appears to be ungrounded.

The Carnapian might here wonder why, even if she acknowledges that (1) contains the definite description "the nothing" rather than the mere nominal syntagm "nothing", she has to read that description as "the thing that is identical with nothing" and then to take (1) as equivalent with (1R), hence with (1FR). More explicitly, the Carnapian might reply as follows. Let us concede that (1) *may* be read as (1FR). The problem is, why *should* it be read as such. Although superficially "nothing" is a noun, logically speaking it expresses a quantifier. So, the alleged description "the nothing" is logically ill-formed. For it impossibly tries to form a nominal syntagm, "the nothing", to be accordingly analyzed *à la* Russell, out of a quantifier, "nothing", i.e., a second-order predicate, rather than out of a genuine noun, which in Russell's analysis disappears in favour (*inter alia*) of a first-order predicate.

Yet reading (1) as (1FR) is definitely not an *ad hoc* move. First of all, we may expect that the account here given for "the nothing" generalizes to other similar cases of definite descriptions involving quantifying expressions. Which is precisely the case. Consider the description "the some-

thing". Its quantifying purpose may well be accounted for by taking it as a shorthand for "the thing that is identical with something". Thus:

(2)   The something somethings

may well be read as:

(2R)   What is identical with something somethings

hence formally as:

(2FR) $(\exists x)((\exists y)(y=x) \wedge (\forall z)((\exists y)(y=z) \rightarrow (z=x)) \wedge Sx)$.

Moreover, the Carnapian should provide a justification for the thesis she implicitly assumes in her reply that an expression's contribution to the logical form of sentence in which it figures remains *always the same*, regardless of the syntactic position in which it occurs. For such a thesis grounds her claim that in the (for her logically ill-formed) description "the nothing", "nothing" works as a quantifier, as in any sentence of the form "Nothing F-s". Yet there are many cases in which an expression's contribution to the logical form of the sentence in which it figures changes according to its syntactic position in such a sentence.

Consider a proper name "N". If friends of direct reference are correct, whenever it occurs in subject position, "N"'s contribution to the logical form of the sentence in which so figures is that of an individual constant. Yet when the definite article is prefixed to "N" so as to yield the definite description "the N", the name does not logically work as an individual constant, but as a predicate, typically expressing the property of *being called 'N'*. So, if we form a definite description by prefixing the definite article to a proper name, say, if we take "Obama" and by prefixing to it the definite article we get the definite description "the Obama", once we insert that description in a sentence it is not simply the case that we *may* read that description as a shorthand for "the individual who is called 'Obama'". Rather, this is its most plausible reading.[9] *Mutatis mutandis*, therefore, the same will hold for any sentence containing the description "the nothing".

---

[9]   Some say that one such description is what the corresponding proper name abbreviates. Yet there will be here no attempt at checking whether the so-called nominal description theory of proper names (cf. e.g. Bach 1987) is the correct semantic account of such expressions.

At this point, absent a problem with (1)'s logical form, the Heideggerian can even wonder whether that description, contextually defined *à la* Russell, has a Russellian denotation, i.e., an entity that uniquely satisfies it. Surprisingly enough, the answer is far from being straightforwardly negative.

First, the existential import of (1R) may well be satisfied, once one understands the existential quantifier in a non-existentially loaded way, that is, as a mere *particular* quantifier. As many have stressed, in order for a Russellian analysis of definite descriptions to work, one does not have to require the existential quantifier involved in that analysis to range over things that exist in a substantial sense.[10] By "substantial sense" one here means that the existence involved here is a property that makes a difference for the individuals having it. As many people believing in such a property claim, this property is a non-universal first-order property, i.e., a property that, like any other genuine first-order property, divides the overall domain of what there is into entities that have it and entities that fail to have it.[11] Now, in Russell's account, to say that there is something that Fs merely means that the property of *being F* is instantiated. This says nothing as to whether in order for that property to be instantiated, whatever instantiates it must exist in a substantial sense—that is, whether the property must be an *existence-entailing* property, in the terminology inaugurated by Cocchiarella (1982). Now, this is the case precisely as regards the property of *being identical with nothing*. In order for this property to be instantiated, whatever instantiates it does not have to exist in any substantial sense. Thus, once one accepts an ontology that allows for things that do not exist in a substantial sense, it may well turn out that the existential import of (1R) is satisfied.[12]

Second, one may argue that the uniqueness import of (1R) is also satisfied. For suppose that there were *more than one* entity that had the property of *being identical with nothing*; that is, suppose that the property $\lambda x((\sim \exists y)(y=x))$ had multiple instances. Suppose further that such a property were the only property that qualifies such entities—the other features

---

[10]   As we will see soon below, Kaplan (2005) is one of the late supporters of this existential neutrality of Russell's theory of descriptions.

[11]   For this substantial sense of existence, cf. e.g. McGinn (2000). Substantial existence is what Williamson (2002) labels "existence in physical sense". If one rather sympathizes with Meinong (1960), this is the property of *having direct or indirect causal powers*.

[12]   Many people have claimed that identity properties are not existence-entailing properties: cf. again McGinn (2000), Williamson (1990).

allegedly qualifying them merely amounting to mere different descriptions that make no real difference between such entities. This indeed is a plausible supposition, for it is hard to see how entities that differ from everything, themselves included, may have other properties.[13] If this were the case, those entities would no longer be distinct. How many entities can there be that differ from everything, themselves included, if this is the only property that allegedly qualifies such entities?

Granted, in order for this argument to be really convincing, one has to endorse a Leibnizian principle of the identity of the indiscernibles. In other terms, one has to assume that for any property F, if entity x has it iff entity y has it, then x and y are the very same entity. Given that (apart from self-identity: see soon later) *being identical with nothing* is the only property such alleged entities possess, then Leibniz's principle indeed forces such entities to be the same thing. Now, it is well known that Leibniz's principle is controversial: why couldn't x and y share all their properties and nevertheless be two numerically distinct things? Be that as it may, *if* both the existential and the uniqueness import of (1R) were satisfied, it would turn out that the description "the nothing" has a Russellian denotation.[14]

For the present purposes, however, one can remain neutral not only on whether that argument is convincing, so that the uniqueness import of (1FR) is satisfied, but also on whether the existential import of (1FR) is satisfied, so that "the nothing" denotes.

To begin with, in order for the property of *being identical with nothing* to be uniquely instantiated viz. for that description to have a Russellian denotation, that denotation must be an *impossible* entity. For in order for that property to be (uniquely) instantiated, even assuming a broad domain of *possibilia* is not enough. Any merely possible entity, that is, any entity which does not substantially exist but which might have done so, does not instantiate the property of *being identical with nothing*. For it will instead instantiate the opposite property of *being identical with something*—

---

[13] Apart, as we will immediately see, the property of *being identical with something*. Consider e.g. the description "the nothing that I like". As many claim, since this description involves the intentional predicate "to like", the fact that one applies this description to x yet fails to apply it to y does not entail that x is not y.

[14] As a result, far from being meaningless because of its contradictoriety as Carnap claims (cf. fn. 8), the further Heideggerian sentence "the nothing exists" turns out to be true, once meaning "There is [in an existentially unloaded way] only one thing that is identical with nothing".

$\lambda x((\exists y)(y=x))$. The merely possible offspring of (a certain sperm of) Brad Pitt and (a certain egg of) Penelope Cruz, for instance, who does not substantially exist but might have done so, is not such that there is nothing that is identical with him; trivially, he is identical with himself. Yet consider now impossible entities, that is, entities that not only fail to substantially exist but also could not have done so. Well, only an *impossibile* can instantiate the property of *being identical with nothing*. For one such entity will be something that is made impossible precisely by the fact that it instantiates not only the property of *being identical with nothing*, but also, like any other entity whatsoever, the property of *being identical with something*.

As stated before, this is not such a big problem for the Heideggerian once she accepts a non-existentially loaded reading of the particular quantifier, or in other terms, once she accepts that the overall domain of what there is also contains entities that do not exist in a substantial sense. Yet the Heideggerian has to face a more serious problem. If the alleged impossible denotation of the description "the nothing" also instantiates, like any other entity, the property of *being identical with something*, then it is not the case that it instantiates the property of *being identical with nothing*. For if our entity is identical with something, then in point of fact it is not the case that there is nothing it is identical with! Or, if you like, it is true both that such a thing is such that nothing is identical with it and that such a thing is such that something (namely, that very thing) is identical with it. So, how can one say that "the nothing" has a denotation, let alone an impossible one, if the existential import of the Russellian paraphrase involving the description is far from being satisfied, or if you like, it is both satisfied and not satisfied?

As one may easily see, this critique is reminiscent of Russell's famous criticism against Meinong to the effect that Meinongian nonexistent objects are to be ruled out of the overall ontological domain insofar as they violate the Law of Non-contradiction. Consider two paradigmatic examples of such objects, namely the square that is not a square and the existent present king of France. The former is such that it is both the case that it is a square and that it is not a square. The latter is such that it is both the case that it exists and that it does not exist. As they violate the Law of Non-contradiction, they are no genuine entities.[15] So according to the cri-

---

[15]   Cf. Russell (1905a,b). As is well known, Russell's original example involved the case of the round square. Yet in order to avoid the problem of whether the fact that the

tique in question, the same holds of the thing that is such that both it is the case that nothing is identical with it and it is not the case that nothing is identical with it. For something is identical with it, namely itself.

As said before, there is here no particular interest in holding that "the nothing" has a denotation. On behalf of the Heideggerian, one merely has to stick to the idea that sentences containing that description, like (1), are logically well-formed. Yet the problem the above critique arises may be successfully dealt with by slightly amending the Russellian account of (1) put forward before. This amendment can be endorsed by a Heideggerian who adopts the usual device Meinongians use to dispense with Russell's criticism: an impossible entity is *ontologically* contradictory—for some pair of property *P* and its complement *non-P*, it possesses both—but it is not *propositionally* contradictory—it does not involve that both a sentence "p" and its negation "~p" are true. Let us see.[16]

Like Russell's criticism to Meinong, the above critique presupposes, as we have done here all along, that the negation involved in mobilizing the expression "being identical with nothing" is *propositional* and not *predicative*: that is, it attaches to sentences rather than to predicates. Yet, as Meinongians do to Russell, the Heideggerian may well respond to the above critique that, as far as *impossible* objects are concerned, the negation there involved is the predicative and not the propositional one. So, Meinongians say, the square that is not a square is not something such that it is both the case that it is a square and that it is not a square. Rather, it is something that is both a square and a non-square. Analogously, the Heideggerian may say, "the nothing" still has an impossible denotation. For it is not the case both that nothing is identical with such a denotation and that something is identical with it (i.e., it is not the case that nothing is identical with it). Rather, that denotation is something that is both identical with something and non-(identical with something), i.e., non-identical with everything. In this respect, like any *impossibile* for Meinong, the impossible denotation of "the nothing" violates the *objectual* version of the Law of Non-contra-

---

round square contains incompatible determinations really makes it a contradictory object, Russell's example is reformulated in terms of a case of a Meinongian object that really is in some sense (as we will see immediately below, an objectual rather than a propositional one) a contradictory object. For this common policy, cf. e.g. Thrush (2001, 160).

[16] For this Meinongian move cf. e.g. Simons (1990).

diction—for any property *P* and its complement *non-P*, an object cannot have both—but not the *propositional* version of that Law—for any sentence "p" and its negation "~p", both cannot be true, as the Russell-like critique instead presupposes. In other terms, the description "the nothing" may still denote an object, let alone an impossible one. For such an entity is impossible not because it is both the case that nothing is identical with it and that something is identical with it, but merely because it has both a property—*being identical with something*, *λx((∃y)(y=x))*—and its complement—*being non-(identical with something)*, i.e., *being non-identical with everything*, *λx((∀y)(y≠x))*.

   This response requires some adjustments in what the description "the nothing" is a shorthand for, hence in the Russellian paraphrase of (1). Once it is read *à la* Russell, the description "the nothing" has to mean the same as "the thing that is non-identical with everything" instead of "the thing that is such that it is not the case that something is identical with it". In other terms, in the relevant predicate that contributes to eliminate away the description, the negation is not attached to an existential quantifier, as it has been supposed with the Heideggerian all along—"(~∃y)"—but rather to the sub-predicate "identical to something", "(∃y)(x=y)", so as to generate the complementary predicate "non-(identical to something)", i.e., "non-identical to everything", "(∀y)(y≠x)". Thus, the quantifying import of "nothing" in that description turns out to be ascribed to a *universal* quantifier—"(∀y)". Putting all this together, for the Heideggerian that adopts the Meinongian distinction between predicative and propositional negation (1R) comes out true when read not as (1FR), but rather as:

   (1FRMH) $(∃x)((∀y)(y≠x) ∧ (∀z)((∀y)(y≠z) → (z=x)) ∧ Nx)$.

For according to such a Meinongianized Heideggerian there is just one thing that is non-identical with everything, a thing which is *objectually* impossible for it is also identical with something, and that thing nothings.

   Granted, even once the Meinongian distinction is adopted, not only endorsing the Leibnizian principle of the identity of the indiscernibles, but also allowing for *impossibilia* so as to get a Russellian *denotatum* for a description such as "the nothing" is hard to swallow. But, to stress the point once again, the problem now is no longer logical, let alone semantic; it is purely metaphysico-ontological. For the purposes of logical form, to read (1) as (1FRMH) is enough. Remember that the challenge against the Heideggerian is to show that (1) is both logically well-formed and meaningful,

not to show that it is true—for which, a necessary condition is that "the nothing" has a Russellian denotation. Yet if on the behalf of the Heideggerian one has an ontological argument in favour of *impossibilia*—as some claim they have[17]—then there is no reason not to admit *impossibilia* into the overall domain of what there is. If moreover on the same behalf one also endorses the Leibnizian principle, there is no reason not to admit an *impossibile* to be the unique instantiator of the property of *being non-identical with everything*, hence to be the denotation of the description "the nothing". As David Kaplan magistrally claimed:

> In fact, although Russell's theory of descriptions is often described as a model for avoiding ontological commitments, it is essentially neutral with respect to ontological commitment. This, I think, is one of its virtues. Meinong believed that there is a nonexistent object that is both round and square. Russell didn't. This is an ontological dispute. If Meinong is right, and nothing else is round and square, then the definite description 'the round square' denotes, and there is no way of using Russell's theory of descriptions to remove this object from the ontology. If Meinong is wrong, then the definite description doesn't denote, and that's the end of it. (Kaplan 2005, 975-976)

At this point, the Carnapian may try to launch a final attack against the Heideggerian with respect to (1)'s logicality. Let us concede, she might say, that Heidegger did not commit the Polyphemousian error of mistaking a quantifying expression—"nothing"—for a singular term. Yet, the Carnapian might go on saying, even if Heidegger deliberately mobilized a definite description—"the nothing"—this does not mean that the Heideggerian reading of it conforms to Russell's theory of descriptions. For what counts, Heidegger quite likely meant "the nothing" as a genuine singular term. But if the Heideggerian maintained that "the nothing" is a genuine singular term, then a logical mistake can still be ascribed to her with respect to (1); namely, the logical mistake Russell ascribed to anyone taking (1)'s subject-predicate grammatical form to be identical with its logical form.

Once again, let us put aside what the *real* Heidegger had in mind. Yet first of all, if in this concern by "genuine singular term" the Carnapian

---

[17]   Cf. for instance Priest (2005), Iacona (2007).

means a *Fregean* singular term,[18] the accusation bounces back against the Carnapian herself. Suppose that "the nothing" is a definite description and moreover that a definite description, as Frege (1980) wanted, has not to be eliminated away *à la* Russell, for it logically is a singular term as much as a proper name. But then, not only "the nothing" turns out to be a genuine singular term, but also (1) is logically perfect. For its logical form turns out to be a subject-predicate form, as Frege claimed for any sentence containing a definite description in subject-position.

Moreover, suppose that by "genuine singular term" the Carnapian instead means a directly referential device, i.e., a term that extinguishes its truth-conditional contribution in its referent—as logically proper names were for Russell and the early Wittgenstein, and proper names in general are for contemporary sustainers of the post-Kripkean new theory of reference.[19] But then to accuse the Heideggerian of mistaking that definite description for a genuine singular term is rather unfair. If somebody must be charged with this accusation, *this precisely is Russell himself*: not the 1905 Russell of the theory of descriptions, but the 1903 Russell of the *Principles of Mathematics*. At that time, Russell explicitly took the expression "nothing" as directly referring[20] to *a certain denoting concept*, [NOTHING].[21]

---

[18]   As a loyal Carnapian should do, given Carnap's (1949) intention of precisifying Frege's semantics. In point of fact, it would be very strange that a loyal Carnapian addressed this criticism to the Heideggerian.

[19]   To be sure, a loyal Carnapian could not take a genuine singular term as a directly referential device; she should take it in a way that does not distinguish between proper names and rigid definite descriptions, namely, as a singular term having *qua* its intension a constant function from possible worlds to individuals. But let us put this complication aside.

[20]   As is well known, Russell tends to use the term "to indicate" in order to speak of the relation between an expression and its denoting concept (1903, §51). Yet this is the same term he uses to speak of the *standing-for* relation that holds between a name and its *designatum*. Thus, following Salmon (2005) the *indication* relation may well be taken as a relation of direct reference between the expression and the item it indicates.

[21]   Cf. Russell (1903, §73). Curiously enough, in that very complicated paragraph of the book Russell puts forward the apparently contradictory theses that i) unlike most denoting concepts, the denoting concept [NOTHING] has no denotation and ii) that denoting concept is not nothing, i.e., it is not identical with its denotation (literally, it is not what itself denotes: for Russell, this is what the sentence "Nothing is not nothing", or better the proposition that sentence signifies, means). That tension may be removed

Thus, one may well take (*that*) *Russell* also to hold that the definite description "the nothing", like any other definite description, directly refers to *another* denoting concept, one having uniqueness embedded in it: a *determinative* denoting concept, [THE [NOTHING]].[22] That is, pretty much as for the 1903 Russell the definite description "the present Queen of England" directly refers to the determinative denoting concept [THE [PRESENT & QUEEN_OF_ENGLAND]], while the definite description "the present King of France" directly refers to the determinative denoting concept [THE [PRESENT & KING_OF_FRANCE]], so according to *that* Russell the description "the nothing" has to directly refer to the determinative denoting concept [THE [NOTHING]]—whatever that concept really amounts to.[23]

Now, if reading "the nothing" as a directly referential term entails endorsing the 1903 Russell's theory of denoting concepts, then the Heideggerian may well say that she declines to endorse that theory for the well known problems it raises *in general*—that is, also with respect to 'ordinary' determinative denoting concepts such as the two above [THE [PRESENT & QUEEN_OF_ENGLAND]] and [THE [PRESENT & KING_OF_FRANCE]].[24] As a result, the Heideggerian may well be inclined to espouse Russell's later

---

if one takes i) and ii) as respectively saying that such a denoting concept has no *possible* denotation and that such a concept does not coincide with its *impossible* denotation. *Mutatis mutandis*, the same would hold of the determinative denoting concept [THE [NOTHING]] that "the nothing" indicates. So, one may tentatively conclude that Russell at least adumbrated the idea, which I have hitherto reformulated in the context of his later theory of definite descriptions, according to which if the description "the nothing" had a denotation, it would be an impossible one.

[22]   The phrase "determinative denoting concept" comes from Orilia (2006, 195). See again Orilia (2003) for the notational convention I use in order to speak of denoting concepts.

[23]   Of course, this denoting concept does not come out as an absurd application of the import of the definite article to the denoting concept indicated by "nothing". Rather, that concept is something the description "the nothing" elliptically indicates. Pretty much as the denoting concept indicated by a description of the kind "the N.N.", where "N.N." stands for a proper name, which would be something like [THE [CALLED_'N.N.']].

[24]   For the purposes of this paper, the well known controversy on whether Russell (1905a) criticism of his own previous theory of denoting concepts is sound may be set apart. For that theory has the preliminary problem that, appearances notwithstanding, it prevents one from speaking of the denotation of a denoting concept by using a description that directly refers to that concept.

theory of descriptions. Which is precisely what the Heideggerian has been ascribed here all along, by supplying her with a reading of (1) as (1R) taken as (1FR), or better, in terms of the previous adjustment, as (1FRHM).

3. Meaningfulness

At this point, the Carnapian may put forward again his second claim against (1). That is, she may remark that in order for (1) to be true, it is not enough to allow the description "the nothing" to have a denotation, it must also be the case that such a denotation satisfies the predicate "nothings" contained in the sentence. But this can hardly be the case, for that predicate means nothing! Which was the second, admittedly weaker, reason, to deny (1) meaningfulness (the first reason being its alleged violation of logical form).

Apparently, the Heideggerian may have an easy reply here. It is rather curious that the same philosophical quarters that deny meaningfulness to an apparently *ad hoc* predicate such as "nothings" have admitted meaningfulness to clearly *ad hoc* predicates like "pegasizes", or even "carnapizes" and "heideggerizes", which have been invoked to yield a systematic descriptivist theory of proper names. As is well known, Quine (1961) assumed that a proper name "N.N." is synonymous with the description "the N.N.-izer", i.e., "the thing that N.N.-izes". Once such a description is available, one can form a sentence out of it in which the predicate explicitly contained in the Russellian paraphrase of such a sentence that 'eliminates away' that description also figures with the same interpretation in the properly predicative part of the sentence, as follows:

(3)   The N.N.-izer N.N.-izes
        $(\exists x)(Nx \wedge ((\forall y)(Ny \rightarrow y=x) \wedge Nx)$

So why not allow for the meaningfulness of a predicate such as "nothings" along the same line or similar ones?

Yet it would be wiser for the Heideggerian to reject this easy reply. First of all, as we have seen the Heideggerian never commits to the Polyphemousian idea that "nothing" works as a proper name in (1), so it would be odd for her to construe a predicate such as "nothings" out of a proper name, as "pegasizes" and the other infamous predicates are. Second, as far as I know, there can be only one reasonable non question-begging in-

terpretation of all the predicates of the same kind as "to N.N.-ize",[25] according to which those predicates respectively express certain *monadic properties* of the individuals they are truly predicated of; typically, monadic individual essences, i.e., non-relational properties those individuals uniquely and necessarily possess.[26] In this vein, the Heideggerian might say that the predicate "nothings" expresses a monadic individual essence of the thing the description "the nothing" denotes. Yet this would amount to explain *obscura* by *obscuriora*. There are many doubts as to what an individual essence *in general* is, that is, also as regards *bona fide* individuals like Carnap and Heidegger, Meinong and Russell, let alone a *monadic* individual essence.[27] To grasp what is a *monadic* individual essence of an impossible entity is definitely an even harder enterprise.

As a matter of fact, the Heideggerian has an alternative and more plausible reply at her disposal. Once the description "the nothing" is taken to mean the same as "the thing that is non-identical with everything", the predicate "nothings" may be analogously paraphrased as "is non-identical with everything", which is formally to be read as: "x is such that every y is non-identical with it". In other terms, (1FRHM) may be further analyzed as:

(1FRHM′) $(\exists x)((\forall y)(y \neq x) \wedge (\forall z)((\forall y)(y \neq z) \rightarrow (z = x)) \wedge (\forall y)(y \neq x))$.

So, the whole sentence may be read as saying that the only thing that is such that everything is non-identical with it is also such that everything is non-identical with it.

In point of fact, this is the most plausible interpretation of the predicate "nothings". An (admittedly colloquial) way of rephrasing Heidegger's sentence is:

---

[25]   For a criticism to an interpretation reading a predicate of the kind "N.N.-izes" as "being identical with N.N.", which obviously does not explain away the name "N.N." Quine intended to get rid of, cf. Leonardi and Napoli (1995).

[26]   Notoriously, Kaplan (1975, 722-3) proposed *haecceities* as relational individual essences like the property of *being identical with O* such that a certain object *O* is its only possible possessor. One such property however presupposes that we already have its unique instantiator at our disposal.

[27]   The only plausible candidates for individual essences are relational non question-begging properties like world-indexed individual properties (cf. Plantinga 1974), or *actualia*-dependent individual properties (cf. Rosenkrantz 1984).

(1′)   The nothing is a nothing

whose surface grammar is of the form "The F is an F". As any sentence of
this form requires that the very same predicate with the same interpreta-
tion occurs throughout it, this is the case with (1′), hence with (1).

In terms of the (admittedly controversial) hypothesis that "the nothing"
denotes, this move yields a statement that is not only logically well-formed
and meaningful, but also true. Granted, that truth is rather trivial—as
when one says that the zip inventor is a zip inventor. Yet the Heideggerian
may also read (1FRHM′) in a more interesting way, by treating the last oc-
currence of the universal quantifier "$(\forall y)$" occurring in it as contextually re-
stricted to the mere subdomain of *possibilia*. By means of this adjustment,
the sentence would say that the only thing that is such that everything is
non-identical with it, is also such that *everything possible* is non-identical
with it. Incidentally, it is likely that such a reading would be welcome to a
Heideggerian that insisted, in a very Heideggerian way, that what one talks
about via "the nothing" is not a *res*, it is beyond the realm of the *Seiende*. In
this reading, in fact, that insistence would not be a Heideggerian clumsi-
ness. For it would express something rather clear: insofar as the thing that
is non-identical with everything is no *possible* thing, it evaporates from the
only reality that counts—the subdomain of *possibilia*—it nullifies itself.
The nothing itself nihilates, as Heidegger's *dictum* is nowadays alternatively
translated.[28]

To sum up. If what has been said here is right, then "the nothing
nothings" is, *pace* Carnap, neither logically ill-formed nor anyway meaning-
less because of the meaninglessness of some of its subsentential compo-
nents, notably its predicate expression. It may also turn out that the de-
scription "the nothing", once it is supplied with a Russellian account, has
a Russellian denotation, provided i) one accepts a Meinongian distinction
between predicative and propositional negation and moreover ii) she en-
dorses the richest possible ontology one may conceive of, i.e., an ontology
also of *impossibilia*, as well as iii) Leibniz's metaphysical principle of the
identity of indiscernibles. But while appealing to i)–iii) goes toward show-

---

[28]   Cf. Heidegger (1977).

ing that (1) is true, it is not essential in order for Heidegger's vituperated sentence to regain a highly respectable status.[29]

References

BACH, K. (1987): *Thought and Reference*. Oxford: Clarendon Press.

CARNAP, R. (1949): *Meaning and Necessity*. Chicago: University of Chicago Press.

CARNAP, R. (1959): The Elimination of Metaphysics through Logical Analysis of Language. In: Ayer, A.J. (ed.): *Logical Positivism*. (Transl. from the 1932 German original by A. Pap.) Glencoe, IL: The Free Press, 60-81.

COCCHIARELLA, N.B. (1982): Meinong Reconstructed *Versus* Early Russell Reconstructed. *Journal of Philosophical Logic* 11, 183-214.

DENNETT, D. (1987): *The Intentional Stance*. Cambridge MA: The MIT Press.

FREGE, G. (1980): On Sense and Denotation. In: Geach, P.T. – Black, M. (eds.): *Translations from the Philosophical Writings of Gottlob Frege*. (Transl. from the 1892 German original by M. Black.) Oxford: Blackwell, 56-78.

HEIDEGGER, M. (1977): What is Metaphysics? In: *Basic Writings*. (Transl. from the 1929 German original by D.F. Krell.) New York: Harper & Row, 89-110.

IACONA, A. (2007): Not Everything is Possible. *Logic Journal of IGPL* 15, 233-237.

KAPLAN, D. (1975): How to Russell a Frege-Church. *The Journal of Philosophy* 72, 716-729.

KAPLAN, D. (2005): Reading 'On Denoting' on its Centenary. *Mind* 114, 933-1003.

LEONARDI, P. – NAPOLI, E. (1995). On Naming. In: Leonardi, P. – Santambrogio, M. (eds.): *On Quine*. Cambridge: Cambridge University Press, 251-266.

McGINN, C. (2000): *Logical Properties*. Oxford: Clarendon Press.

MEINONG, A. (1960): On the Theory of Objects. In: Chisholm, R. (ed.): *Realism and the Background of Phenomenology*. (Transl. from the 1904 German original by I. Levi, D.B. Terrell, and R.M. Chisholm.) New York: Free Press, 76-117.

ORILIA, F. (2003): A Description Theory of Singular Reference. *Dialectica* 57, 7-40.

ORILIA, F. (2006): Identity across Time and Stories. In: Bottani, A. – Davies, R. (eds.): *Modes of Existence*. Munich: Ontos Verlag, 191-220.

PLANTINGA, A. (1974): *The Nature of Necessity*. Oxford: Clarendon Press.

PRIEST, G. (2005): *Towards Non-Being: The Logic and Metaphysics of Intentionality*. Oxford: Clarendon Press.

QUINE, W.V.O. (1961): On What There Is. In: *From a Logical Point of View*. New York: Harper & Row, 1-19.

ROSENKRANTZ, G. (1984): Nonexistent Possibles and Their Individuation. *Grazer Philosophische Studien* 22, 127-147.

RUSSELL, B. (1903): *The Principles of Mathematics*. Cambridge: Cambridge U.P.

RUSSELL, B. (1905a). On Denoting. *Mind* 14, 473-493.

RUSSELL, B. (1905b): Critical Notice of: A. Meinong, *Untersuchungen zur Gegenstand-theorie und Psychologie*. *Mind* 14, 530-538.

SALMON, N. (2005): On Designating. *Mind* 114, 1069-1133.

SIMONS, P. (1990): On What There Isn't: The Meinong-Russell Dispute. In: *Philosophy and Logic in Central Europe from Bolzano to Tarski*. Dordrecht: Kluwer, 159-181.

THRUSH, M. (2001): Do Meinong's Impossible Objects Entail Contradictions? *Grazer Philosophische Studien* 62, 157-173.

WILLIAMSON T, (1990): Necessary Identity and Necessary Existence. In: Haller, R. – Brandl, J. (eds.): *Wittgenstein: Towards a Re-Evaluation I*. Vienna: Hölder-Pichler-Tempsky, 168-175.

WILLIAMSON, T (2002): Necessary Existents. In: O'Hear, A. (ed.): *Logic, Thought and Language*. Cambridge: Cambridge University Press, 233-251.

Frege on Knowing and Individuating Senses

FREDRIK STJERNBERG

Department of Culture and Communication. University of Linköping
Building Key 4403. Campus Valla. Linköping. Sweden
fredrik.stjernberg@liu.se

ABSTRACT: This paper discusses Frege's views on how Fregean senses, *Sinne*, should be individuated, and what chances we have of arriving at knowledge about them. There is a conflict in Frege's views. First he introduces a criterion of sameness of sense which requires that speakers are authoritative concerning senses, then holds that there are cases where no one knows the sense of an expression. If no speaker is authoritative concerning sense, then the original criterion for sameness of sense cannot be upheld. But Frege repeatedly emphasized the need for criteria of identity. This paper discusses the conflict between the different things Frege wants to say about sense, and points to possible ways out of the problems.

KEYWORDS: Frege – identity – mode of presentation – sense.

1. Introduction

We can find a conflict in Frege's characterization of sense.[1] There should be a tension between on the one hand the criterion of identity Frege proposes for senses, and on the other hand his view that there are cases where no speaker knows the sense of a given expression. It is a conflict which should have been obvious to Frege, but it seems that he never discussed it, or even appeared to view it as a problem. This lack of interest should tell us something about how to understand Frege's notion of sense. I will indicate the strands in Frege's thinking that lead up to the apparent

---

[1] I use the English terms "sense" and "reference" for Frege's German "Sinn" and "Bedeutung", respectively.

conflict, without arguing for the ideas behind them. Then I say something about how the conflict should be resolved, and what consequences this resolution can have for our understanding of Frege's notion of sense. This way of resolving the puzzle will then also show something about how Frege thought that we can arrive at knowledge of sense. Some difficulties remain for the proposed solution, but this proposal appears to retain most of what is central in Frege's work. Some tension will remain in Frege's views on any account of the matter. In the last section, I will very briefly sketch a kind of position concerning our knowledge of abstract objects which can be seen in Frege's work, a position that could be called "rationalist pragmatism". Tyler Burge has been emphasizing the rationalist pragmatism of Frege in his writings on Frege (now collected in Burge 2005), and the sketch I provide is no more than a sketch; developing this sketch is properly a subject for another paper.

The dominant interpretation of Frege's notion of sense has been to see it as something related to language, something like linguistic meaning, except for a few more or less supposedly peripheral cases, such as vague expressions, proper names and indexical expressions. It is also a notion tied to *understanding*. A person who understands a sentence is, according to this interpretation, said to do this in virtue of having grasped the sentence's sense, the thought expressed. Then this notion is used as the starting point for constructing a theory of meaning for a language.

Even if this interpretation is historically important, and clearly captures much of importance in Frege's thinking, it has always been clear that it is problematic. Frege's at times almost contemptuous attitudes towards ordinary language,[2] as it is used in communication, and the extent of the idealization of the kind of language he was interested in, have been toned down or neglected in this interpretation, and Frege's repeated claims that his intention was to say something about the structure of *thought*, not language, have been played down.

This interest in thought was prominent in Frege's work from the very beginning. In *Begriffsschrift* (Frege 1879), Frege characterized his goals in the following manner:

---

[2]   As for instance in "A person who wants to learn logic from language is like an adult who wants to learn how to think from a child" (Frege 1980b, 41).

I did not want to present an abstract logic in formulas, but rather display a content by means of written signs in a manner that was more precise and more surveyable than is possible by the use of words. (Frege 1879, 97)[3]

Begriffsschrift had the subtitle *A formula language, modeled upon that of arithmetic, for pure thought*. In an early paper, Frege quotes Leibniz with approval: "A *lingua characterica* ought, as Leibniz says, *peindre non pas les mots, mais les pensées*" (Frege 1880/1881, 13).[4] Of course, these two passages antedate the distinction between sense and reference, but there are also later passages, where it seems that Frege sees the need for a linguistic clothing of thoughts as something which distracts from his real interests. A late passage that highlights Frege's views on the relations between logic and language is from 1915:

If our language were logically more perfect, we would perhaps have no further need of logic, or we might read it off from the language. But we are far from being in such a position. Work in logic just is, to a large extent, a struggle with the logical defects of language, and yet language remains for us an indispensable tool. Only after our logical work has been completed shall we possess a more perfect instrument. (Frege 1915, 252)

Language is at times inappropriate for the true expression of thoughts, as is the case with expressions for functions (Frege 1914, 239). Footnote 4 of "Thoughts" says:

I am not here in the happy position of a mineralogist who shows his audience a rock-crystal: I cannot put a thought in the hands of my readers with the request that they should examine it from all sides. Something in itself not perceptible by sense, the thought, is presented to the reader—and I must be content with that—wrapped up in a perceptible linguistic form. The pictorial aspect of language presents difficulties. The sensible always breaks in and makes expressions pictorial and so improper. So one fights against language, and I am compelled to occupy myself with language although it is not my proper concern here. (Frege 1918a, 13)

---

[3]   See also *ibid*., VI-VII.

[4]   Leibniz's French can be translated "not depict words, but thoughts".

The idea behind the language-oriented interpretation has been that sense still could function as the main concept for a theory of meaning as a theory of understanding, even if we make room for the things Frege said about the shortcomings of natural language. There are good reasons for seeing sense as intimately connected with language. Dummett has often stressed that there is no other good explanation of what sense is, unless through bringing in some linguistic means for expressing these thoughts (for instance in Dummett 1991, chapter 12). The conflict I will be discussing below can be seen as a special case of the tension in Frege's thought between seeing sense as tied to the realm of thought, and seeing it as essentially connected with language, and thus that differences in sense are essentially tied to their linguistic expressions. In the next section, I will sketch how Frege introduces senses. The following section discusses how sameness of sense is to be determined, and the remaining sections describe how Frege's views on sense and sameness of sense lead to a conflict, and how this conflict is to be resolved.

2. Frege's introduction of sense

Frege's notion of sense is usually discussed as if the most important aspect of this notion were the introduction of sense for singular terms, proper names, and that the problem of informative identities—"Hesperus = Phosphorus"—was the central reason for introducing the distinction between sense and reference. This is historically not quite correct. Frege had, almost *en passant*, introduced that distinction in "On Function and Concept" (Frege 1891a), published before "On Sense and Meaning" (Frege 1892), although the two papers were written more or less at the same time. In "On Sense and Meaning", the claim that sentences have a *Bedeutung* was of equal importance to Frege, which can be seen from the space he allots to defending the latter claim. Beaney argues persuasively that part of the need for introducing a distinction between sense and reference stems from Frege's needs to justify aspects of his logicist project, and in particular to explain why the logicist definitions of central mathematical terms, such as "number", should be understood in precisely the manner suggested by Frege (Beaney 1996, chs. 5–8). This rationale for introducing a distinction between sense and reference will not make the introduction of sense for singular terms *the* reason for having a sense-reference distinction. I will

still, however, concentrate upon the senses of singular terms, partly for ex-pository reasons, partly because the problem at hand is more salient there. The question for Frege is then how "Hesperus = Phosphorus" can differ from "Hesperus = Hesperus", how they can be of different value for a sub-ject. Frege's answer is that the two names, "Hesperus" and "Phosphorus", have different senses, because their referent, Venus, is presented in two dif-ferent ways. The first explanation of sense is then that sense is that where-in the *mode of presentation* is contained (Frege 1892, 57/27). Frege does not straightforwardly *identify* sense with the mode of presentation of the object: all he says is that the mode of presentation is contained in the sense. It is not entirely clear what the difference might amount to. One possible inter-pretation is that mode of presentation is a more inclusive category, seen as being of relevance for perception as well, since there could be informative identities for a subject who has devised no linguistic means of thinking of an object, whereas sense might be confined to that for which there is a lin-guistic expression. Another important difference between sense, as con-ceived by Frege, and modes of presentation as naturally understood, is that Fregean sense is to some extent independent of how the reference is actual-ly apprehended, whereas modes of presentation are naturally understood precisely as features of the *actual* apprehension of the reference. But since Frege is introducing a technical term here, we should perhaps not try to strain the interpretations by reading too much into the notions from our preconceptions.

The sense of a proper name is grasped by everyone who is sufficiently familiar with the language (Frege 1892, 57/27). Sense is something which is both generally cognitive and tied with the understanding of the language.

3. The intuitive criterion of sameness of sense

Criteria of identity are required for senses. This is not just a lesson from Quine; Frege could have subscribed to the "No entity without identi-ty" slogan. In §62 of *The Foundations of Arithmetic* (Frege 1884), Frege stresses the need for criteria of identity:

> If we are to use the symbol *a* to signify an object, we must have a crite-rion for deciding in all cases whether *b* is the same as *a*, even if it is not always in our power to apply this criterion. (Frege 1884, 73)

If senses are to be accepted, there should be some criterion of identity for them. In accordance with the way in which sense has been introduced, and the point of having them, Frege provides a criterion of identity for senses.

> The intuitive criterion: If it is possible for a speaker to rationally have different cognitive attitudes towards two sentences S and S', then there is a difference in sense.

I have followed Evans in calling this the "intuitive" criterion, since it makes the subject's judgments of sameness and difference in cognitive attitudes that decide sameness and difference in senses.[5] This criterion is well entrenched in Frege's thought. It is used implicitly in "On Sense and Meaning" (see the opening pages of Frege 1892), and it is formulated explicitly in letters to Husserl and Jourdain, so it is found from 1891 to 1914. It is also put to use in "Thoughts", in 1918. A "difference in cognitive value" means that the speaker thinks that the two sentences can differ in truth value. A relevant passage is the following, from a letter to Jourdain, in 1914:

> What is expressed in the sentence "Ateb is Afla" is not at all the same as the content of the sentence "Ateb is Ateb".... In accordance with this, the sense of the sentence "Ateb is at least 5000 meters high" is different from the sense of the sentence "Afla is 5000 meters high". Someone who holds the latter to be true, is in no way forced to hold the former to be true. (Frege 1980b, 112)

For Frege, there is a normative aspect to this criterion; it is not just a question of what actual speakers actually do, when confronted with appropriate pairs of sentences, it is a question of what attitudes are rational for a speaker who knows the language. What is required for being rational or knowing the language is not spelled out in any great detail by Frege; it seems that he had a common-sense notion of this in mind. An intuitive criterion along these lines is not far-fetched, since sense for singular terms is introduced precisely in order to account for the possibility that a speaker thinks that two sentences differ in truth value, even if the only difference between them is that different terms, referring to the same things, are used in these sentences. The introduction of a notion of sense can then contri-

---

[5]   See Evans (1982, 18-19).

bute towards a construction of a theory of meaning that can serve to explain the activities of speakers, a project that was not Frege's but still is attractive.

Sense is thus introduced to account for the possibility of subject's rationally holding different attitudes towards sentences that differ only in the way something is thought about, or presented to the mind. If sense is to capture the possibility of speakers' rationally entertaining different attitudes towards such sentences, then sense would have to be exactly as finely individuated as it is possible for a subject to have the different attitudes and still remain rational. If this were not the case, the subject might hold different attitudes concerning the same sense, and the introduction of sense would not serve its purpose.[6] The intuitive criterion will then mean that the subject who has different attitudes towards different sentences must be strongly authoritative concerning her own attitudes and what senses she is grasping. If it were unclear to the subject which sense she were grasping, then the intuitive criterion could not be made to work in the way intended. So sense must be transparent to the subject. Dummett has formulated the requirement of transparency in the following way: "(A) sense cannot have any features not discernible by reflection on or deduction from what is involved in expressing it or in grasping it." (Dummett 1981b, 50) There are, however, factors that complicate this picture. One is that even if Dummett's gloss on transparency is accepted, there is much room for uncertainty—what makes a feature discernible by reflection or deduction? Is there, for instance, some upper limit on how much reflection is needed?[7]

There can be genuine uncertainty concerning what does follow when we reflect on the properties of a given concept. All the deductive consequences of an axiom are rarely available when we start thinking about something. The results concerning the properties of continuous functions that were obtained in the nineteenth century could have served as an inspiration for Frege's thought here. It took a great deal of effort from clever people to sort out what a continuous function was, and what was excluded by being a continuous function.

---

[6]   Or we might find a need to introduce a new level of metasenses, meant to explain why a subject rationally could hold different attitudes towards one and the same sense.

[7]   There has been a renewed interest in transparency lately, as for instance in Fine (2007, especially 60–64), and Sainsbury and Tye (2012). Aspects of these recent views are discussed in Stjernberg (forthcoming).

4. Do we know senses?

The realist, or Platonist, aspects of Frege's thought are well-known. Frege repeatedly insisted that there was a vast difference between being true and being held to be true. Abstract entities like senses, thoughts, concepts and numbers are imperceptible, and exist outside of time and space, independently of us and our possibility to gain knowledge about them.[8] These views of Frege are outside the scope of this paper, but we can note that Frege's kind of realism about abstract entities creates difficulties for the employment of the intuitive criterion: if the existence of senses really is independent of us and our knowledge, why should we then be credited with privileged authority in our judgements about the identity and diversity of sense? Just postulating such an ability doesn't help. There are some options for handling this problem.

One is to reformulate the intuitive criterion as a hypothetical criterion, possible to use if the subject does have a clear grasp of the relevant senses. Frege may have had something like this in mind (since the setting of his presentations of the intuitive criterion demand that the speaker *understands* the expressions).[9] The problem with this suggestion is that it requires that there is a way to know whether a subject does have this clear grasp, and it seems that the only way to do this would be to have some other criterion of identity for senses—and with such a criterion in place, there is no further need for the intuitive criterion. Some other criterion of identity would be needed, because without it, there would be nothing to distinguish between on the one hand the case where the speaker has an unclear grasp of sense A, and on the other hand has a clear grasp of a different sense A'. So this reformulation cannot be the whole story. It has been suggested before that Frege's intuitive criterion, or some version of it, should be restricted to cases where the subject fully understands the sense.[10] It would seem that

---

[8]   For the claim about their being non-temporal and non-spatial, see for instance Frege (1884, vii, 58, 61, 85, 93; 1918a, 25f; 1918b, 52). Their independent existence follows from their objectivity, which is stressed in for instance (1884, 26; 1918a, 69-70). For imperceptibility, see (1884, 85; 1924, 265). See also Burge (1992).

[9]   Frege in fact makes this demand on occasion: "vorasgesetzt wird, dass die Auffassung der Inhalte keine Schwierigkeiten macht", "provided that the grasp of contents presents no difficulties", Frege (1983, 212).

[10]   Jeshion (2001) is one example of this kind of restriction—see pp. 965–966.

Frege never thought of such a restriction as the whole story about under-standing and individuating senses—understanding is connected with the ability to *explicate* the concept. Presumably full understanding would be connected with the ability to fully explicate the concept. This still leaves us with a question concerning the identity criteria for senses: how are we to distinguish explicating one concept from explicating some other?

Another way to handle this worry would be to play down the Platonist understanding of what sense is like, and emphasize the way in which senses are to be seen as the way in which we think of references, thereby playing down Frege's view that senses are objects.[11]

There is one prominent passage in Frege, which might be taken to in-dicate that sense, unlike what is the case for some other abstract entities, perhaps is not completely independent of our activities and our opportuni-ties to gain knowledge about sense:[12]

> The being of a thought may also be taken to lie in the possibility of dif-ferent thinkers' grasping the thought as one and the same thought. In that case the fact that a thought had no being would consist in several thinkers' each associating with the sentence a sense of his own; this sense would in that case be a content of his particular consciousness, so that there would be no *common* sense that could be grasped by several people. (Frege 1918b, 35)

This passage appears to support the claim that Frege occasionally thought that senses in some way could be partly dependent on human activities. I do not think that this is the correct interpretation of this passage. From the context it is clear that Frege has something else in mind. The situation described in the quoted passage is dismissed as spurious. The reason for this dismissal is that the transmission of thought would be endangered (and Frege repeatedly insists that we *do* transmit thoughts). The discussion of this particular problem ends with:

> Our act of judgement can in no way alter the make-up of a thought. We can only acknowledge what is there. A true thought cannot be af-fected by our act of judgement. (Frege 1918b, 36)

---

[11]   For discussion of this, see Burge (1992) and Dummett (1991, ch. 12).

[12]   Weiner takes it this way, in her (1995a) and (1995b).

It is true that Frege had a too stark dichotomy between on the one hand the completely subjective, and in his view incommunicable, and on the other hand completely objective, timeless or eternal, Platonistically conceived thoughts on the other hand. There was no place left for an intermediate category of the intersubjective, and several authors have tried to show that Frege's arguments for the eternal existence of thoughts are unsatisfactory, failing to establish his Platonism.[13] It may well be that some other position would have suited Frege's needs better or just as good, but there should be little doubt what the position was.

Frege conceived of our thinking as the grasping of senses: "In thinking, we do not produce thoughts, we grasp them." (1918a, 25) He also thought that it is only thanks to the Platonic nature of sense that it is possible for mankind to have a common stock of theories and knowledge. The premiss in his argument is that mankind has a common stock of knowledge, and that there are genuine cases of communication. Frege goes on to hold that the only explanation for this is that what we communicate could not be the sole property of a single speaker. What is communicated is something that is common to all (or at least to those engaged in communication). But it is not sufficient that what is communicated is common, in the sense of being intersubjective. It has to have a more permanent basis of existence. In fact it has to be thought of in Frege's preferred Platonist manner. Closely tied with this kind of Platonism is the view that it may often be quite difficult to have a clear grasp of the Platonistically conceived abstract objects. Frege held that we may have to struggle to arrive at a clear command of a concept. As a consequence of this, Frege thinks that we should not speak of the development of a concept, or of the history of a concept, but rather of the history of our attempts to grasp or articulate a concept clearly.[14]

Apart from the fact that it follows from the intuitive criterion that we are authoritative about sense, there is little *direct* evidence that Frege thought that we always, or regularly, are authoritative about sense. Some of Frege's views about our knowledge of abstract objects can be summarised as indicating that he thought we were authoritative about senses as well. One example of this might be Frege (1884, § 105), which holds that certain entities, being the "nearest kin" of reason, are utterly transparent to it. This

---

[13]   See Carruthers (1984a) and (1984b); Beaney (1996, chapter 7, especially pp. 213–224).

[14]   Frege (1884, Introduction; 1891b, 134).

passage antedates the distinction between sense and reference, however, and may not be a blanket endorsement of the idea that senses are completely transparent to us. Frege could have had a restricted view of which abstract objects that qualify as the nearest kin of reason, and he could therefore in any case have wanted to exclude senses. Another passage which perhaps could be used to support the claim that sense must be completely known by a subject, is from Frege's letter to Jourdain in 1914. There he discards the idea that a Fregean thought—the sense of a sentence—could be (even in part) composed of physical objects: "It seems to be unreasonable that pieces of lava, *and also such that I have no knowledge of*, are parts of a thought".[15] It is clear from the passage, and his other discussions of this issue, that he sees the main problem as lying in the supposition that actual physical objects make up a Fregean thought—this would be impossible to square with the eternal existence of thoughts—but here he locates part of the difficulty elsewhere: in thoughts being made up, even in part, by things I know nothing of. Admittedly, this may be reading too much into the passage, but it is noteworthy that Frege gives the argument he does, when his other views would have sufficed to settle this particular issue directly.

In addition to the claim that we may have difficulties arriving at a clear view of abstract objects like concepts, there are passages where Frege goes one step further, and holds that we sometimes do *not* have a clear grasp of senses. Here, Frege's idea is that we can grasp a sense, but that our command of this sense is blurred. The following passages are from "Logic in Mathematics", written in 1914:

> When we examine what actually goes on in our mind when we are doing intellectual work, we find that it is by no means always the case that a thought is present to our consciousness which is clear in all its parts. For example, when we use the word 'integral', are we always conscious of everything appertaining to its sense? I believe that this is only very seldom the case ... If we tried to call to mind everything appertaining to the sense of this word, we should make no headway. Our minds are simply not comprehensive enough. (Frege 1914, 209)

---

[15]   Emphasis added. German original, again with emphasis added: "Es scheint mir aber ungereimt, dass stücke Lava *und zwar auch solche, von denen ich keine Kenntnis habe*, Teile eines Gedanken sein sollen." (1980b, 111)

How is it possible … that it should be doubtful whether a simple sign has the same sense as a complex expression if we know not only the sense of the simple sign, but can recognize the sense of the complex one from the way it is put together? The fact is that if we really do have a clear grasp of the sense of the simple sign, then it cannot be doubtful whether it agrees with the sense of the complex expression. If this is open to question although we can clearly recognize the sense of the complex expression from the way it is put together, then the reason must lie in the fact that we do not have a clear grasp of the sense of the simple sign, but that its outlines are confused as if we say it through a mist. The effect of the logical analysis of which we spoke will then be precisely this—to articulate the sense clearly. (Frege 1914, 211)

Perhaps the sense appears to both [men] through such a haze that when they make to get hold of it, they miss it. One of them makes a grasp to the right perhaps and the other to the left, and although they mean to get hold of the same thing, they fail to do so. How thick the fog must be for this to be possible! (Frege 1914, 217)

Surely no arithmetical sentence can have a completely clear sense to someone who is in the dark about what a number is? This question is not an arithmetical one, nor a logical one, but a psychological one. We simply do not have the mental capacity to hold before our minds a very complex logical structure so that it is equally clear to us in every detail. For instance, what man, when he uses the word 'integral' in a proof, ever has clearly before him everything which appertains to the sense of this word! And yet we can still draw correct inferences, even though in doing so there is always a part of the sense in penumbra. (Frege 1914, 222)

In these passages Frege makes room for the possibility that a speaker can have some grasp of sense, but that this grasp is not sufficient for full authority.[16] In that case, the intuitive criterion will not work.

---

[16]   There are other passages. Frege (1906, 197) expresses a similar idea, although formulated in terms of "content". Frege (1897, 138) explicitly says that our grasp of sense may be inadequate or blurred: "We might cite, as an instance of thoughts being subject to change, the fact that they are not always immediately clear. But what is called the clarity of a thought in our sense of this word is really a matter of how thoroughly it has been assimilated or grasped, and is not a property of thoughts."

What could inadequate grasp of sense amount to? One way to explicate this is to talk about division of linguistic labour, and place the issue in the hands of the experts in the speaker's society (Putnam 1975). Then it would at least be possible to say about a single speaker in a society that he misunderstands a specific sense *x* (and not understands some other sense *x*' adequately), just in case there are experts in that society that do have an adequate grasp of sense *x*. This is, however, not what Frege is driving at. According to Frege, there may well exist several cases, where *no one*, not even the experts with the best theories, can be said to have a clear grasp of the sense of a certain expression. Frege uses examples such as 'number', where not even the best mathematicians have a clear grasp of the sense.[17] What would it in general mean to have the clear grasp of sense we are looking for? More on this below, in the concluding section, but the main idea is tolerably clear. If a speaker has a clear grasp of the sense of an expression— if the sense is clearly articulated—then the speaker knows full well what further things he is committed to, so that all the commitments involved in grasping that sense are laid bare. As in the passage from Frege (1897) which was just quoted (p. 138), it is a matter of the assimilation of the sense.

I will not make any further attempts to motivate these Fregean claims; it is sufficient that the claims about an inadequate grasp of sense are well entrenched in his theory of abstract objects, and that it shouldn't be seen as some kind of momentary and regrettable lapse.[18]

5. Handling the conflict

There is an obvious conflict here. If senses are individuated by the intuitive criterion, then the subject's considered judgments regarding possible combinations of attitudes are authoritative. And since Frege presents no other account of the identity or diversity of senses, we are left without

---

[17]   This is of course a central theme in Frege (1884), but is also used in Frege (1914), as on p. 242.

[18]   Dummett refers to one of the just quoted passages from Frege (1914, 211), and comments: "It is to be hoped that this remark will not prompt anyone to try to reconstruct Frege's theory of the weather conditions in the space intervening between our minds and the realm of sense." (Dummett 1981b, 337) Frege had himself, however, used a fog or mist metaphor earlier in a similar context, see Frege (1884, vii-viii).

much of an account of what senses might be, at least if Frege's own strictures regarding the need for criteria of identity are to be accepted. If we stick with the only criterion of identity we are presented with—the intuitive criterion—the subject cannot have an incomplete grasp of senses. We stand without any alternative account of the identity and diversity of sense that works. This conflict is intimately connected with another issue as well. Among the cases where we are said to be ignorant of the senses, we find cases that were central to Frege's interests, such as *number*. In his earlier work, Frege had proceeded by definitionally equating numbers with extensions of concepts. This replaces a supposedly ill-understood notion with something that is precisely characterized. But if the sense isn't known, how could we then point to a definitional replacement?[19] In this section, I will consider various ways to handle the conflict, starting with suggestions that are of lesser interest, working my way towards those that may turn out to be more interesting.

The first suggestion is that Frege simply was confused: perhaps there really is an indissoluble conflict here, and Frege never realized it. Frege was not the kind of person to make mistakes of this kind. We should at least keep on looking for a more charitable interpretation of Frege.

Another idea is that Frege changed his mind, or that one of the two claims (the claim of the intuitive criterion and the claim about the unclear grasp of sense) doesn't really represent Frege's considered views on these issues. This is again possible, but there is no really good support for these suggestions. Both claims are well entrenched in Frege's thinking. The intuitive criterion can be found in the same paper as many of the passages about an incomplete grasp of sense (1914, 224ff). I think that none of the two claims can be discarded completely, although a solution of the conflict will probably have to put less weight on one of the claims.

The conflict could perhaps be avoided by stressing that there is a loophole: the intuitive criterion is about *differences* in sense, whereas the claim about incomplete grasp of sense is about knowledge of sense. There is then room for a position which holds that subjects really are authoritative about differences in sense, without always knowing what senses it is that they have authoritative knowledge about. They know whether this means the same as that, but they do not always know what it is that this or that

---

[19]   These issues are well discussed in Beaney's book, in Horty (1993) and (2007), and Tappenden (1995).

means. This is obviously not very attractive, and there is in fact no loophole available here. We could in that case simply construe senses as the equivalence classes of same-sense entities, in which case there would be no place left to talk about an inadequate grasp of senses—if a subject is authoritative about differences in senses, then he will turn out to be authoritative about grasp of senses as well.

Another suggestion might be that in cases where the sense of some expression is not known, then it is simply *impossible* to have rational thought. It seems that there are some hints at such a view in Frege: mathematics stands in need of firmer foundations than any that had been provided so far. But there are also opposing (rightly so, in my opinion) strands in his thinking. Mathematics is the crowning achievement of human thought, and if mathematical thinking were not rational, then what would be?

A further suggestion was mentioned earlier. The idea was that the intuitive criterion should be seen as hypothetical: if the subject has a clear grasp of the senses, then the intuitive criterion holds. One is almost inclined to say that this suggestion is *obviously* correct, and in a sense it is. But then, as I said above, two problems remain: what is it to grasp a sense clearly, and when *are* two senses identical? According to this suggestion, some other means of individuating senses takes precedence over the intuitive criterion, and the appeal to our intuitions concerning differences in cognitive value—the very intuitions that were used when Frege introduced senses in 1892—becomes problematic. Why should there be such a connection between sense and cognitive value? The answer that this is the function of sense for speakers with a clear grasp of sense is perhaps correct, but it is not very illuminating.

A different way of presenting this suggestion is that the claim about inadequate grasp of sense holds for actual speakers, whereas the intuitive criterion strictly applies only to ideal speakers. The passage from Frege quoted above (1914, 222) may substantiate this interpretation, since that passage describes inadequate grasp of sense as a matter of psychology, not logic or arithmetic. Given our constitution, we have to get at sense through language, featuring expressions we do not have complete command of, and the grasp of sense we get is at times distorted, hazy, or inadequate.

In another context, Frege held that there may exist beings that differ from ordinary human beings in being able to grasp sense directly, without the means of language. In fact, Frege has an even stronger point to make here: it is precisely *because* we human beings get at sense with the aid of

language—a human creation, created with a focus on other needs than displaying logical form—that we at times do not manage to get hold of sense properly:

> There is no contradiction in supposing there to exist beings that can grasp the same thought as we do without needing to clad it in a form that can be perceived by the senses. But still, for us men there is this necessity. Language is a human creation; and so man had, it would appear, the capacity to shape it in conformity with the logical disposition alive in him. Certainly the logical disposition of man *was* at work in the formation of language but equally alongside this many other dispositions—such as the poetic disposition. And so language is not constructed from a logical blueprint. (Frege 1924/25a, 269)

The possible conflict is then avoided by holding that the perfect grasp of sense required for the intuitive criterion to work as a matter of psychological fact is something which we human language-users hardly ever attain, but that the intuitive criterion works for those case where the subject has some authoritative grasp of sense.

This is probably along the right lines, but then several things have to be done. One is that some other kind of criteria of identity and diversity for senses have to be derived, another is that we expect to be told when a speaker has the right sort of authoritative grasp of sense for the intuitive criterion to work. Also, if the intuitive criterion is applicable only for ideal, authoritative, speakers, then it is problematic to use sense as a central concept in a theory about normal speakers' understanding of language. We could argue that the concept of sense is little more than a harmless idealization of something we rarely manage to achieve. But in that case, the relation between on the one hand sense as understood in this idealized manner, and on the other hand sense as essentially connected with the use of language, becomes obscure.

There is a general tension between a linguistic and a non-linguistic conception of sense here. If grasp of sense is possible only through the linguistic expression of sense, then there is no problem seeing how we could fail to notice that two senses are the same: we could, as is at times the case with stipulative definitions, simply have cases where one expression is too long and complex to be handled with ease. Then it would be a *psychological* matter that our grasp of sense sometimes is inadequate. In the context of "Logic in Mathematics" (Frege 1914), Frege appears to have thought that

the value of stipulative definitions lay precisely in providing signs as suitable receptacles for sense, simply because our minds are not sufficiently capacious to handle the complexities we sometimes have to deal with:

> If we tried to call to mind everything appertaining to the sense of this word, we should make no headway. Our minds are simply not comprehensive enough. We often need to use a sign with which we associate a very complex sense. Such a sign seems, so to speak, a receptacle for the sense, so that we can carry it with us, while being always aware that we can open this receptacle should we have need of what it contains. ... If therefore we need such signs—signs in which, as it were, we conceal a very complex sense as in a receptacle—we also need definitions so that we can cram this sense into the receptacle and also take it out again. So if from a logical point of view definitions are at bottom quite inessential, they are nevertheless of great importance for thinking as this actually takes place in human beings. (Frege 1914, 209)

This indicates Frege's abandoning the intuitive criterion for non-ideal speakers. The sense of one of the paired terms is too complex for the subject to have any attitudes about, and the whole issue of sameness of sense is simply bypassed. It is stipulated that the two expressions, the simple and the complex, have the same sense. And here we also have an implicit characterization of the plight of the non-ideal speaker—fettered by complexity considerations. Yet, as I said, Frege upheld the intuitive criterion even in this paper.

6. How we know senses

Frege held on to both the intuitive criterion and the view that we at times can—and in the interesting cases usually do—have an inadequate grasp of sense. This highlights a tension in his conception of sense. Senses cannot at the same time be the objects of propositional attitudes and be individuated by the subjects' attitudes towards these objects, unless we are dealing with ideal speakers. An ideal speaker is never side-tracked, bored, etc., and is in no way hindered by complexity. Hence Frege's demands on ideal speakers are very strong. Ideal speakers must have infinite minds. For finite minds, complexity will always have to be factored in. Ideal speakers are correct about the objects of their propositional attitudes, and if they

hold different attitudes towards two senses, these senses are not identical. In general, we don't arrive at that state (though there may be some cases where we have that kind of knowledge). Far from being anomalies, the possibility of an inadequate grasp of sense is in fact of central importance for Frege's philosophical project.

The project of giving a firm foundation for arithmetic forced Frege to face the issue how his proposed reconstruction of arithmetic was connected with the starting point, arithmetic as it was known and practiced. It was natural for Frege to conceive of the knowledge imparted by his reconstructions as a gradual unfolding of a sense already given, but imperfectly understood. The task of the proposed reconstruction was to better articulate and understand the notions involved. But Frege never succumbed to some kind of Cartesian scepticism concerning mathematical knowledge. He had the working mathematician's faith that they were at least getting *something* right, even if the various paradoxes and conceptual revolutions of nineteenth century mathematics showed that there was foundational work to be done. This attitude can be seen in a letter to Hilbert from 1895, where Frege discusses the merits of, and need for, symbolization:

> A purely mechanical formalization is dangerous 1 for the truth of the results, 2 for the fruitfulness of the science. It appears that the first danger can be dealt with completely by the logical perfection of the symbolization. As for the second, the science would be brought to a standstill if the formulaic mechanism would take over to such an extent that it suffocated the thought completely. I would still not like to think of such a mechanism as completely useless or harmful. On the contrary, I think it is necessary.[20]

From this passage it is clear that Frege thinks that it is fundamentally possible to have some degree of access to the objects of mathematics before we start formalization. Formalization is perhaps necessary for full understand-

---

[20]   Frege, letter to Hilbert, Oct 1, 1895, Frege 1980b, 4f. Translation by the author. German original: "Ein bloss mechanisches formeln ist gefährlich 1. für die Wahrheit der Ergebnisse, 2. für die Fruchtbarkeit der Wissenschaft. Die erste Gefahr lässt sich wohl fast ganz durch die logische Vervollkomnung der Bezeichnung beseitigen. Was die zweite betrifft, so würde die Wissenschaft zum Stillstande gebracht, wenn der Formelmechanismus so überhand nähme, dass er den Gedanken ganz erstickte. Dennoch möchte ich solchen Mechanismus keineswegs als ganz unnütz oder schädlich ansehen. Im Gegenteil glaube ich, dass er notwendig ist."

ing and sharp explication of the concepts, but ordinary practices do not leave mathematicians in the dark.

Frege's view on our knowledge of abstract entities was not tied with the usual Platonist emphasis on intuition, or using *seeing* as the guiding metaphor. Frege basically thought of such knowledge in more practical terms, as a kind of *doing*.[21] And in handling, as it were, the senses, it is possible to step by step arrive at ever clearer articulations of the sense—giving up the ideal of such access to abstract entities as a kind of glorified seeing tends to remove the temptation to think of such access as infallible. The temptation to think of such access as giving us direct insight into the nature of these entities is also reduced. The passage from Frege 1897, already referred to, is both an endorsement of the view that senses can be inadequately grasped and that it is not inherent in the nature of senses that they are transparent:

> what is called the clarity of a thought in our sense of this word is really a matter of how thoroughly it has been assimilated or grasped, and not a property of a thought. (Frege 1897, 138)

To think of senses as entities that normally, or even invariably, obey the ideal of transparency of mental content that is inherent in the intuitive criterion reverts to a recognizably Cartesian conception of the inner, and makes Frege's views harder to understand. This does not prevent the intuitive criterion from being useful as a kind of ideal at the end of inquiry.

The connection between the problem at hand and the position I called "rationalist pragmatism" in the first section is pretty straightforward, now that we have come this far. Making the grasp of sense a matter of what we can *do* with the sense relieves much of the tension that should have troubled Frege, and it also introduces views and themes that it would not be unfair to call "pragmatist". A more comprehensive treatment of this strand in Frege's thought lies well beyond the scope of this paper, so here I will just give a very brief sketch of what I mean by this, and offer some textual backing for my claims.

---

[21] The differences between the traditional, vision-based, conceptions of knowledge of abstract entities and Frege's more pragmatic views are real enough, but there are elements of a vision-based conception in Frege as well. Since Frege repeatedly used the fog and mist metaphors to characterize our inadequate grasp of senses, he must have found it natural to think of the process in terms of visual metaphors.

The tensions in Frege's conception of sense display one central strand in his views concerning the point of his philosophical enterprise. A part of this enterprise was devoted to the logicist program of providing firm, logical, foundations for mathematics. This meant that he had to show that certain entities deep down were something they on the face of it were not, as when he held that numbers were the extensions of concepts. A proposed *analysans* would have to have at least something in common with its *analysandum*, or else the point of the project was left uncertain. To leave the reference intact would be insufficient, and to leave the sense intact would lead us directly to the paradox of analysis, if the intuitive criterion were upheld: how could the proposed analysis be of any value, if all it did was to conserve the sense? Frege's thought was that, when doing the kind of foundational work he was engaged in, we present a reconstruction of the problematic notions, rather than an outright sense-preserving analysis. An ill understood notion is replaced by a streamlined and better articulated notion. This is quite close to Carnap's ideas concerning explication and Quine's views on regimentation. But if the success of the proposed reconstruction is not to be judged solely by the intuitive criterion, how is it to be judged? Here Frege's answers are somewhat sketchy, but in the early period, he tended to stress *fruitfulness* as the central criterion:

> All these concepts have been developed in science and have proved their fruitfulness. For this reason what we may discover in them has a far higher claim on our attention than anything that our everyday trains of thought might offer. For fruitfulness is the acid test of logic, and scientific workshops the true field of study for logic. (Frege 1880/81, 33)

In the early period, Frege's demand is that definitions should be fruitful (1884, 100f), and much of his early thought, at least, displays a general outlook that has some clearly pragmatist features. Peirce took the "Principle of Pragmatism" to be the following principle about meaning:

> In order to ascertain the meaning of an intellectual conception we should consider what practical consequences might conceivably result by necessity from the truth of that conception; and the sum of these consequences will constitute the entire meaning of the conception. (Peirce, quoted in Audi 1996, 566)

Guiding in the pragmatist conception of meaning is the idea that differences in meaning always must entail possible differences in experience. Prag-

matism is thus historically usually allied with some form of empiricism. In Frege, however, we find a kind of rationalist version of this idea. For logic, Frege says the following:

> One must always hold fast to the fact that a difference is only logically significant if it has an effect on possible inferences. (Frege 1880/81, 33fn)

Taken in isolation, this quotation is admittedly rather weak as evidence for a pragmatist view, since it is hard to understand what other grounds there might be for logical differences, apart from differences in inferences. Frege uses this view in arguing against Wundt's representation of logical inferences. What is more important for the present issue is that Frege's views on the basic nature of our knowledge of senses, and on a priori thought in general, diverge from the standard Platonist picture of knowledge as a kind of seeing. The pragmatist idea is that our knowledge of meaning is not so much a matter of *seeing* things in the abstract realm—it is better seen as a question of *doing*, of making the right transitions, of connecting this with that, of assimilating a concept into a network of conceptual connections.[22] And it is not surprising if it should turn out that this command is not something which the single speaker could display all at once, not even to himself.[23] For Frege, knowledge was always a matter of *doing*, which his choice to talk of *grasping* senses indicates. Knowing something about abstract objects is not tied to intuiting them, or seeing them in a third realm—it is tied to knowing what to do with them. A typical early example is the following:

> [W]hat is common to lengths and surfaces, escapes our intuition. This comes out most clearly in the case of an angle. No beginner will get a correct idea of an angle if the figure is merely placed before his eyes. this is what has occasioned numerous attempts to give an explanation of an angle, even though the situation is at bottom exactly the same in the case of a length, except that this idea is more familiar to us from our

---

[22] Frege's consistent rejection of appeals to intuition, and his insistence that we ground knowledge of something by providing proofs, is in line with this general view.

[23] Brandom discusses this feature of Frege's thought, and dubs it "tactile Fregeanism", in Brandom (1994, 579ff), and in (2008, 51–52).

ordinary lives. *If a beginner is shown how to add angles, then he knows what they are*. (Frege 1874, 56, emphasis added)

If this is Frege's most basic conception of what knowledge of the third realm amounts to, we should not be surprised to find that our knowledge can be improved, that there is no guarantee that it is authoritative or immediate.[24] It would not be too wide of the mark to call this kind of view *rationalist pragmatism*. It also clearly has parallels with Brandom's inferentialism.[25] What we *do* with stuff is more important than what it represents. Knowledge-how is more basic than knowledge-that in this area.

Frege's views here have some central aspects in common with classical pragmatism, but steer clear of the empiricism with which pragmatism is usually associated. It sees our current position as a state of knowledge, but knowledge that can and should be improved, so a wholesale scepticism is beside the point. This is a kind of position which has echoes in Peirce and in Neurath's celebrated boat metaphor.[26] Knowledge can be improved and sharpened, and put on a firmer basis. This firmer basis is provided by the new tools that the *Begriffsschrift* can provide for the mathematician, and the kind of full knowledge of abstract objects that is the mark of full understanding is probably attainable *only* by use of the symbolic tools of the *Begriffsschrift*. Knowledge in this area is a kind of knowing *how*: knowing how to add angles, how to explicate concepts fully, how to draw connections in the web of concepts making up the abstract world of mathematics, a world which is independent of us, but possible to arrive at knowledge about. So even if pragmatism is a sort of hazy and much abused term, I think the term fits many of Frege's deeply held beliefs about our knowledge of abstract objects. This position is still little explored, but merits further study.[27]

---

[24] Belabouring Frege's tactile metaphors, one could say that there is at times *groping* involved in trying to *grasp* a sense.

[25] See Brandom (1994), (2000) and (2008).

[26] In Frege (1924/25b), he states: "I, for my part, never had any doubt that numerals must designate something in arithmetic, if such a discipline exists at all, and that it does is surely hard to deny." Frege is not doubting arithmetic and the existence of numbers from the bottom up; he is trying to find a clearer and more perspicuous foundation for something.

[27] Burge (2005) has done much to investigate this kind of position

References

AUDI, R. (ed.) (1996): *The Cambridge Dictionary of Philosophy*. Cambridge: Cambridge University Press.

BEANEY, M. (1996): *Frege: Making Sense*. London: Duckworth.

BELL, D. (ed.) (1991): *The Analytic Tradition*. Oxford and New York: Oxford University Press.

BRANDOM, R. (1994): *Making it Explicit*. Cambridge, Mass.: Harvard University Press.

BRANDOM, R. (2000): *Articulating Reasons. An Introduction to Inferentialism*. Cambridge, Mass.: Harvard University Press.

BRANDOM, R. (2008): *Between Saying and Doing. Towards an analytic pragmatism*. Oxford: Oxford University Press.

BURGE, T. (1991): Frege on Sense and Linguistic Meaning. In: Bell (ed.) (1991), 30-60.

BURGE, T. (1992): Frege on Knowing the Third Realm. *Mind* 101, 633-650; reprinted in Burge (2005), 299-316.

BURGE, T. (2005): *Truth, Thought, Reason. Essays on Frege*. Oxford: Oxford University Press.

CARRUTHERS, P. (1984a): Eternal Thoughts. In: Wright (ed.) (1984), 1-19.

CARRUTHERS, P. (1984b): Fragmented Senses. *Mind* 93, 351-369.

DUMMETT, M. (1981a): *Frege: Philosophy of Language*. 2nded. (1st ed. 1973). London: Duckworth.

DUMMETT, M. (1981b): *The Interpretation of Frege's Philosophy*. London: Duckworth.

DUMMETT, M. (1991): *Frege and Other Philosophers*. Oxford: Oxford University Press.

EVANS, G. (1982): *The Varieties of Reference*. Oxford: Oxford University Press.

FINE, K. (2009): *Semantic Relationism*. Oxford: Oxford University Press.

FREGE, G. (1874): Methods of Calculation Based on an Extension of the Concept of Quality. In: Frege (1984), 56-92.

FREGE, G. (1879): *Begriffsschrift. Eine der arithmetischen nachgebildete Formelsprache des reinen Denken*. Halle.

FREGE, G. (1880/1881): Boole's Logical Calculus and the Concept-Script. In: Frege (1979), 9-46.

FREGE, G. (1884): *Die Grundlagen der Arithmetik*. Breslau. English translation, *The Foundations of Arithmetic* by J.L. Austin, 2nd ed., Oxford 1953.

FREGE, G. (1891a): On Function and Concept. In: Frege (1980a), 21-41.

FREGE, G. (1891b): On the Law of Inertia. In Frege (1984), 123-136.

FREGE, G. (1892): On Sense and Meaning. In Frege (1980a), 56-78, second pagination refers to German original, Über Sinn und Bedeutung, *Zeitschrift für Philosophie und philosophische Kritik*, 1892, 25-50.

FREGE, G. (1893): *Die Grundgesetze der Arithmetik*. Vol. I. Jena.

FREGE, G. (1897): Logic. In: Frege (1979), 126-151.

FREGE, G. (1906): A Brief Survey of my Logical Doctrines. In: Frege (1979), 197-202.

FREGE, G. (1914): Logic in Mathematics. In: Frege (1979), 203-250.

FREGE, G. (1915): My Basic Logical Insights. In: Frege (1979), 251-252.

FREGE, G. (1918a): Thoughts. In: Frege (1977), 1-30.

FREGE, G. (1918b): Negation. In: Frege (1977), 31-54.

FREGE, G. (1924): Number. Reprinted in: Frege (1979), 265-266.

FREGE, G. (1924/25a): Sources of Knowledge of Mathematics and Natural Sciences. In: Frege (1979), 267-274.

FREGE, G. (1924/25b): Numbers and Arithmetic. In: Frege (1979), 275-277.

FREGE, G. (1977): *Logical Investigations.* Edited by P. Geach. Translated by P. Geach and R. H. Stoothoff. Oxford: Blackwell.

FREGE, G. (1979): *Posthumous Writings.* Edited by H. Hermes, F. Kambartel, F. Kaulbach. Translated by P. Long and R. White. Oxford: Blackwell.

FREGE, G. (1980a): *Translations from the Philosophical Writings of Gottlob Frege.* Edited by P. Geach and M. Black. 3rd ed. (1st ed. 1952). Oxford: Blackwell.

FREGE, G. (1980b): *Gottlob Freges Briefwechsel mit D. Hilbert, E. Husserl, B. Russell sowie aus gewählte Einzelbriefe Freges.* Edited by G. Gabriel, F. Kambartel and C. Thiel. Hamburg: Felix MeinerVerlag.

FREGE, G. (1983): *Nachgelassene Schriften und Wissenschaftlicher Briefwechsel.* Edited by H. Hermes, F. Kambartel and F. Kaulbach. Hamburg: F. Meiner Verlag.

FREGE, G. (1984): *Collected Papers on Mathematics, Logic, and Philosophy.* Edited by B. McGuinness. Oxford: Oxford University Press.

HORTY, J. (1993): Frege on the Psychological Significance of Definitions. *Philosophical Studies* 72, 223-263.

HORTY, J. (2007): *Frege on Definitions. A Case Study on Semantic Content.* Oxford: Oxford University Press.

JESHION, R. (2001): Frege's Notions of Self-Evidence. *Mind* 110, 937-976.

PUTNAM, H. (1975): The Meaning of 'Meaning'. Reprinted in: Putnam, H. *Collected Papers.* Vol. I. Cambridge: Cambridge University Press.

SAINSBURY, M. – TYE, M. (2012): *Seven Puzzles of Thought.* Oxford: Oxford University Press.

STJERNBERG, F. (forthcoming): Attacks on Fregean Transparency.

TAPPENDEN, J. (1995): Extending Knowledge and 'Fruitful Concepts'. *Noûs* 29, 427-467.

WEINER, J. (1995a): Burge's Literal Interpretation of Frege. *Mind* 104, 585-597.

WEINER, J. (1995b): Realism *bei* Frege. *Synthese* 102, 363-382.

WRIGHT, C. (ed.) (1984): *Frege: Tradition and Influence.* Oxford: Blackwell.

Meaning Normativism: Against the Simple Argument

KATHRIN GLÜER

Department of Philosophy. Stockholm University
SE – 10691 Stockholm. Sweden
kathrin.gluer@philosophy.su.se

ÅSA WIKFORSS

Department of Philosophy. Stockholm University
SE – 10691 Stockholm. Sweden
asa.wikforss@philosophy.su.se

ABSTRACT: The paper examines a central argument in support of the thesis that meaning is essentially normative. The argument tries to derive meaning normativism from the fact that meaningful expressions necessarily have conditions of correct application: Since correctness is a normative notion, it is argued, statements of correctness conditions for an expression have direct normative consequences for the use of that expression. We have labeled this the 'simple argument', and have argued that it fails. In this paper we elaborate on our objections to the argument in response to Daniel Whiting's recent attempt to rescue it. We argue, first, that statements of correctness conditions simply allow us to categorize the applications of an expression into two basic kinds (for instance, the true and the false) without this having any normative implications; and, second, that the normativist has not provided any reasons to think that some further, normative notion of semantic correctness is essential to meaning.

KEYWORDS: Anti-normativism – correctness – meaning.

Since Kripke (1982) first suggested that meaning is essentially normative, the thesis has been subject to much scrutiny and criticism. It has been argued that the thesis fails, and that whatever norms are associated with language are extrinsic to meaning. Nevertheless, meaning normativism is still with us, and several authors have recently attempted to revive some

version of it. Among the arguments employed by normativists one stands at the center. This argument tries to derive meaning normativism from the claim that meaningful linguistic expressions necessarily, or essentially, have conditions of *correct application*. Since correctness is a normative notion, the argument goes, statements of correctness conditions for an expression have *direct* normative consequences for the use of that expression. In a nutshell: no meaning without correctness conditions and no correctness conditions without normative consequences.[1]

We have labeled this 'the simple argument' (Glüer and Wikforss 2009). It is simple in the sense that it does not require any substantive semantic commitments beyond the idea that correctness conditions are required for an expression to have meaning. Naturally, the notion of 'conditions of correct application' is a place holder, and one can debate what to fill it with, but all parties agree that some such notion is needed to account for the basic semantic relation between meaningful expressions and the world. If, therefore, meaning normativism just rests on the assumption that meaningful expressions necessarily have correctness conditions, the thesis would seem to be beyond reproach. Moreover, it would mean that the thesis could have the function Kripke (and many following him) assigned to it: It can serve as a constraint on any acceptable of meaning, to be used as a weapon against every attempt to naturalize meaning.

It is therefore easy to see the attraction of the simple argument. However, we have argued that it fails (Glüer 1999a; 1999b; 2000; Glüer and Wikforss 2009; Wikforss 2001). In this paper we would like to defend and elaborate on our objections to the simple argument in response to some recent attempts to rescue it. In particular, we would like to consider Daniel Whiting's efforts, in a string of recent papers, to save the argument (2007; 2009 and 2013). We shall argue that Whiting fails to defend the simple argument against the objections he considers. There are other arguments in support of meaning normativism (for an overview and discussion see Glüer and Wikforss 2009), but the prospects of finding a quick, theory neutral argument, look increasingly bleak.

---

[1]   See, for instance, Boghossian (1989, 513), and Blackburn (1984, 281f), for versions of this argument. The argument is clearly present in Kripke (1982, 37), although he also provides other considerations in support of meaning normativism. Whiting (2013) characterizes normativism based on this argument as the 'orthodox interpretation' of the slogan that meaning is normative.

1. The simple argument: round one[2]

To say that meaning is essentially normative is to make a claim about the *nature* of meaning: No meaning without norms. The norms in question are supposed to follow from nothing but the nature of meaning—they are genuinely *semantic*, and distinct from other types of norms (moral, prudential, epistemic, etc.). According to the strongest interpretation, it is both metaphysically and conceptually necessary that meaning is normative: Meaning normativism is a conceptual truth.[3] Those who rely on the simple argument tend to adopt this stronger claim. Moreover, they share the standard assumption that the relevant notion of normativity is that of *prescriptivity*, involving genuine, action-guiding 'oughts'. Whiting formulates the position along these lines: "Facts about meaning, according to it, are inherently action-guiding or prescriptive; they have implications for what a subject *may* or *should* (not) do" (2009, 536).

We have suggested that there are two distinct interpretations of meaning normativism, what we have labeled (ME)-normativism and (MD)-normativism (Glüer and Wikforss 2009). According to the first interpretation, statements of what an expression means have immediate *normative consequences*: The normativity is 'engendered by', or consequent upon, meaning. According to the second interpretation, it is the norms that 'engender' meaning, and normativism is a metasemantic thesis—meaning is *determined* by the speakers' following certain norms, or by their being in force for them. The simple argument is used in support of (ME)-normativism, since the idea is precisely that meaning statements have immediate consequences for how the speaker ought (or may) use an expression.[4]

---

2   Parts of this section are taken straight from our earlier paper "Against Normativity Again: Reply to Whiting" (unpublished MS). Whiting (2009), responds to this text (it was available on the web), and in section 2 we respond to this response.

3   Normally, the conceptual claim is taken to imply the metaphysical claim. Not everyone agrees however. For example, Gibbard (2012) argues that while the *concept* of meaning is normative, *meaning* is not. His view, therefore, does not actually qualify as meaning normativism in the sense characterized here.

4   Defending the simple argument, Whiting makes clear that his concern is with (ME)-normativism: "to say that meaning is a normative notion is to say that a statement of what an expression means is, or immediately implies, a statement about what

Those who have objected to this argument have not denied the truism that meaningful expressions have correctness conditions. We shall go along with Whiting here and take this to amount to a commitment to the following principle:

(C) w means $F \rightarrow \forall x(w$ applies correctly to $x \leftrightarrow x$ is $f)$,

"where 'w' is a word, 'F' gives its meaning, and 'f' is that feature in virtue of which w applies" (2007, 134). What the anti-normativist denies is that (C) has any direct normative consequences. She denies that we can move from a statement such as (C_1) to (N_1):

(C_1) For any speaker S, and any time t: if 'green' means *green* for S at t, then it is correct for S to apply 'green' to an object x iff x is green at t.

(N_1) For any speaker S, and any time t: if 'green' means *green* for S at t, then S ought to apply 'green' to an object x iff x is green at t.[5]

In support of their claim, anti-normativists have pointed out that a) 'correct' can be used in normative and in non-normative ways, and b) the relevant notion of correctness in (C) is the notion of *semantic correctness*. What that precisely amounts to depends on the choice of basic semantic concept; the main contenders are *truth* and *warranted assertibility*. Either way, the anti-normativist submits, the notion of semantic correctness is non-normative in precisely the sense that no statements about what we ought (not) to or may (not) do with w directly follow from (C).

Putting this point more positively, we have argued that what statements of correctness conditions, such as (C), give us is nothing more than the conditions for the application of the basic semantic concept to applications of the word w.[6] Nothing in (C) shows that this has to amount to anything

---

we ought (not) to or may (not) do with that expression" (2007, 134). See also Whiting (2013, 4).

[5]   It has been much debated how the relevant norm is to be formulated. The 'iff' formulation seems too strong since we clearly do not have an obligation to apply 'green' to all green things. At the same time, simply replacing it with an 'if then' is too weak to support standard normativism (Bykvist and Hattiangadi 2007). Whiting's proposal is that the 'iff' can be retained if the 'ought' is replaced with a 'may' (2009, 544-545). Since we deny that (C) implies norms of any form, we shall not engage with this debate.

[6]   See especially Glüer (2001, 60f); Wikforss (2001, 205ff).

over and above the possibility of *categorizing*, or *sorting*, applications of *w* into two basic semantic kinds; for instance, the true and the false. Nothing in (C) shows that correct applications of *w* are those that ought to, or may, be made of *w*.

Let's call a categorization that has no direct normative consequences a "non-normative categorization". Sorting things into tables and non-tables should clearly be non-normative in this sense. Now, saying that a categorization is non-normative is not the same as saying that it cannot be used to derive normative consequences. Indeed, *any* categorization can be used to derive normative consequences. But not directly. Any categorization of things into *A*s and *non-A*s, be they actions or not, can be used to derive normative consequences *if a suitable norm is in force*. Take tables. If a suitable norm is in force, for instance the norm that tables under all circumstances ought to be kicked, normative consequences can be derived from something's being a table. But not directly. Things can be categorized into tables and non-tables without any such norm being in force.

The normativity thesis must therefore not be mixed up with the claim that normative consequences can be derived from semantic categorization. That would utterly trivialize the thesis. Normative consequences can be derived from any categorization. But not every categorization is such that they can be derived directly. The anti-normativist claims that semantic categorization is like sorting objects into tables and non-tables: No immediate normative consequences ensue.[7]

In his 2007 paper, Whiting defends the simple argument against similar objections put forth by, among others, Hattiangadi (2006). In support of his argument, Whiting tries to hijack an analogy provided by Hattiangadi. The example is that of a minimum height requirement for going on a certain ride in a theme park. Hattiangadi observes that whether a child meets this "standard" is a "straightforwardly non-normative, natural fact" (2006, 224). Whiting agrees that in order for the child to meet the standard certain descriptions concerning her height must be true of her. However, he argues, given that the "standard is in force" (136), the fact that the child

---

[7]   There might, of course, be *other* reasons for why meaning statements has to be loaded with normative consequences. For instance, Koťátko (1998) argues that the very concept of utterance meaning is to be analyzed in terms of the utterance's normative consequences. What is relevant here is simply that this does not *immediately* follow from an expression's having conditions of semantic correctness but requires further, substantive semantic commitments.

meets this standard has immediate implications for whether she *may* (or *should not*) go on the ride: "If she were to do so incorrectly, *with the norm in place*, sanctions or criticism ... would be appropriate" (136, emphasis ours). A similar conclusion, he contends, holds in the case of meaning. If '*w* means *F*' does indeed imply that there are conditions for the correct application of the term, then this equally has implications "for whether it may or should be used in certain ways" (ibid.).

As far as we can see, this simply illustrates the point that a given non-normative categorization (here, of children as having or not having a certain height) can be used to derive normative consequences—*if* a norm to that effect is in force. Here, this is the norm that children under a certain height may not go on the ride in question. That does nothing to show that we would not be able to sort children by height if no such norm were in force. Rather, it seems perfectly obvious that we can do that. Our point is precisely that (C), by itself, gives us no reason to think that the same does not hold for semantic categorization: Sorting applications of *w* by, for instance, truth and falsity is possible without any norms being in force.

Notice that the notion of a standard, just like the notion of correctness, has non-normative and normative uses. Whiting seems to have the latter use in mind when he speaks of a standard's *being in force*.[8] It is of course trivial that meeting the standard in this latter, normative sense has normative consequences for whether the child may (or should not) go on the ride. Similarly, if 'semantic standard' is construed normatively, the fact that a given use "meets the standard" is, as Whiting puts it, "clearly a normative or evaluative matter" (2007, 135). But what we wanted to know was why the notion of a semantic standard (or, as we have put it, the notion of a semantic category) should be construed normatively in the first place; not what follows if it *is* construed that way.

---

[8]   The talk of "being in force", notice, applies to *norms*: It is the idea that the norm is valid for a subject. Hattiangadi suggests that standards can be "in force" in the purely descriptive sense "that they are accepted within some relevant community and are enforced by sanctions" (2009, 57). But it is very difficult to see how standards, in a non-normative sense, can be accepted and enforced by sanctions. Whether the standard "ought to be in force", as Hattiangadi puts it, is a further matter—even norms that ought not to be in force are norms (in the sense relevant to this discussion).

2. The simple argument: round two

Whiting (2009) sets out to defend normativism against recent criticisms. In particular, he aims to show that the objections to the simple argument fail. He has two lines of defense: First, he suggests that it is problematic to hold that 'correct', in (C), expresses a non-normative notion, when it normally does express a normative notion. The normativist, according to Whiting, postulates an ambiguity that is not independently motivated. Second, returning to Hattiangadi's fairground example, he argues that it does not provide an example of a standard by which one can judge that a certain act is correct and which has no normative implications.

Let us discuss the second point first. In her response to Whiting (2007), Hattiangadi suggests that we consider what kind of standard is operative in the fairground case. She invites us to compare (2009, 56):

(S$_1$) S is permitted to ride if and only if S is over one meter tall.
(S$_2$)   Ride X is safe for S if and only if S is more than one meter tall.

The mere fact that a child meets a certain standard, Hattiangadi argues, does not in itself have any normative consequences, as illustrated by (S$_2$): That Vikram is over one meter implies that it is safe to ride, but does not in itself imply anything about whether he ought (not) or may (not) go on the ride (cf. 2009, 57).

Whiting responds by suggesting that Hattiangadi's argument can be used to turn the tables on the anti-normativist. If indeed the fairground standard is along the lines of (S$_2$), Whiting argues, then it is no longer possible to derive from it, and from the fact that Vikram is over one meter, that his going on the ride would be correct. Indeed, Whiting suggests, this supports the view that correctness is a normative affair, and that we can infer from (C$_1$) that S should not apply 'green' to x if x is not green. After all, that such an application would be incorrect is accepted by normativists and anti-normativists alike (2009, 542).

However, this misses the crucial claim the anti-normativist starts out from: that the notion of semantic correctness used in (C) is a *placeholder*, to be replaced by your favorite basic semantic concept. If your favorite basic semantic concept is that of truth, and if 'green' means *green*, you'll agree that applying 'green' to an object o that's not green is incorrect—where that means that o does not satisfy the predicate 'is green'. The fairground analogy assumes that the relevant placeholder notion of correctness, *fair-*

ground correctness, has already been replaced—in the example by the notion of a ride's safety. If you take both these things together, it does indeed follow that going on the ride *is* correct—*fairground correct*—if you are one meter tall. But that by itself has no normative consequences because it only means that going on the ride is safe. Whether you ought (not) or may (not) go is a different matter entirely.

This of course is precisely the move that Whiting's first line of defense is supposed to undermine. According to Whiting, the claim that the notion of semantic correctness used in (C) is merely a placeholder for your favorite basic semantic concept is implausible; 'correct' in (C) *must* be interpreted normatively. Since the shared starting point here is (C), *this* dispute does *not* concern the *extension* of the notion of semantic correctness—the normativist and the anti-normativist count precisely the same applications as correct (and as incorrect). The dispute thus concerns the very *concept* of semantic correctness, not its extension.

Take an anti-normativist who is a fan of truth-conditional semantics. She thinks that the basic semantic concept is that of truth. Against her suggestion that 'semantically correct' just provides a theory-neutral way of talking about the basic semantic features of sentences (or predicates), Whiting in effect argues that even though 'truly' and 'correctly' can, in a context like (C), be substituted *salva veritate*, they cannot be substituted *salva intensione*.

To make this plausible, he appeals to a distinction stressed by Gideon Rosen, between correctness (a normative notion) and the correctness-making feature, the non-normative property something must have in order to count as correct. To say that someone is playing the Moonlight Sonata correctly is not just to make a claim about the notes played, but to make a higher order claim that the performance possesses the feature that "makes for correctness in acts of that kind" (Rosen 2001, 620). Similarly, Whiting argues, we should not identify correctly applying an expression with truly applying it. Therefore, "even if one agrees with the anti-Normativist that the pertinent 'word-world relation' is not normative, this does not undermine the view that the property of correctness—possessed in this instance in virtue of the 'word-world relation'—is normative" (2009, 539). What the anti-Normativist must show, Whiting continues, is not just that from (C) one can derive a non-normative statement about when an expression truly applies "but that one cannot *also* derive normative statements about what a subject may, should or has reason (not) to do" (2009, 540).

Of course, if you think that the notion of correctness used in (C) is a normative notion, you will think that you can derive normative consequences directly from (C). The considerations just rehearsed show that employing a Rosen-style "higher order" notion of correctness makes it possible for the normativist to hold on to that claim *even if he concedes that the basic semantic notion itself is not normative*. This is a very important concession for the normativist to make, and we shall come back to that.

But first, we would like to ask how showing that a Rosen-style construal of the notion of correctness *can* be used to hold on to the claim that (C) has direct normative consequences, is supposed to demonstrate that 'correct' in (C) *must* be interpreted normatively (Rosen-style)? As far as we can tell, it simply does not do that. Rather, we are at a conceptual impasse again: The normativist uses the Rosen-style construal of 'correct' to absorb the insight that the basic semantic notion itself is not normative and shows that 'correct' *can* nevertheless be interpreted normatively in (C). But this does nothing to prevent the anti-normativist from countering that 'correct' *does not have to* be interpreted that way in (C)—she is perfectly free to interpret it in terms of the non-normative concept she takes semantic correctness to be.

Faced with such a dispute, it can be concluded either that the disputants operate with different concepts, or that one of the parties is conceptually confused. We do not think it is our business to accuse people of conceptual confusion and will therefore simply grant the normativist that he can absorb the insight that the basic semantic concept is not normative if he interprets 'correct' in (C) Rosen-style. To get beyond this impasse, however, substantive *further* argument would be needed.

Whiting seems to rest his case simply on the idea that interpreting 'correct' in (C) non-normatively does not cohere with ordinary usage. He appeals to Ralph Wedgwood's complaint that it is "surely implausible" to suggest that the word 'correct' is ambiguous, and suggests that this places the burden on the anti-normativist to "provide reason to think that 'correct' behaves in the way he suggests, when appearances suggest otherwise" (2009, 538).

We have two things to say in reply. First, it is important to remember that the notion of semantic correctness *as used in (C)* is *not* an everyday notion, but has its place in semantic theory. The question of whether 'correct' is ambiguous in natural language is therefore of only limited relevance when it comes to the plausibility of the claim that *as used in (C)*, 'correct' is

just a placeholder for the basic semantic concept. But secondly, it seems rather clear to us that 'correct' is indeed used in different ways in natural language, some of them normative, others not.[9] In fact, a look in the dictionary confirms this. Here are the three main entries Merriam-Webster lists for the adjective 'correct': 1) true or accurate, agreeing with facts, 2) having no errors or mistakes, 3) proper or appropriate in a particular situation. We thus remain unconvinced that 'correct' in (C) has to be interpreted normatively.

Is the upshot then that both the normativist and the anti-normativist can simply stick to their guns? Not quite. Remember that Whiting's normativist has conceded that the basic semantic concept is *not normative*. This concession saddles him with a considerable task if he still aims to provide further argument to the effect that 'correct' in (C) not only can, but must be interpreted normatively. He would have to argue that the anti-normativist fan of truth-conditional semantics, for instance, is missing something essential to meaning by interpreting 'correct' in (C) as a placeholder for 'true'—even though the anti-normativist construal of (C) *captures the basic semantic relation perfectly well* (the "word-world relation" Whiting talks about). But given this concession, what element essential to meaning could possibly be missing from an anti-normativist account of semantic correctness? Doesn't the very need to construe normative correctness as a "higher order" feature, a feature merely "surfing on" the basic semantic relation, testify to its inessentiality? If truth indeed is the basic semantic concept, is there any reason to think that having truth conditions would not amount to being meaningful unless these truth conditions are also correctness-making conditions in a normative sense? Given the concession that the basic semantic concept itself is not normative, the need for construing the notion of semantic correctness as normative has become ever so much harder to motivate. Pending further argument, we cannot help but conclude once more that the normativist notion of semantic correctness is nothing but an idle wheel in the theory of meaning.

---

[9]   Whether these amount to ambiguity is a difficult question that we do not want to pronounce upon here.

References

BLACKBURN, S. (1984): The Individual Strikes Back. *Synthese* 58, 281-301.

BOGHOSSIAN, P. (1989): The Rule-Following Considerations. *Mind* 98, 507-549.

BOGHOSSIAN, P. (2005): Is Meaning Normative? In: Beckermann, A. – Nimtz, C. (eds.): *Philosophy – Science – Scientific Philosophy*. Paderborn: Mentis.

BYKVIST, K. – HATTIANGADI, A. (2007): Does Thought Imply Ought? *Analysis* 67, 277-285.

GIBBARD, A. (2012): *Meaning and Normativity*. Oxford: Oxford University Press.

GLÜER, K. (1999a): *Sprache und Regeln. Zur Normativität von Bedeutung*. Berlin: Akademie.

GLÜER, K. (1999b): Sense and Prescriptivity. *Acta Analytica* 14, 111-128.

GLÜER, K. (2000): Bedeutung zwischen Norm und Naturgesetz. *Deutsche Zeitschrift für Philosophie* 48, 449-468.

GLÜER, K. (2001): Dreams and Nightmares. Convention, Norms and Meaning in Davidson's Philosophy of Language. In: Kot'átko, P. et al. (eds): *Interpreting Davidson*. Stanford, CA: CSLI.

GLÜER, K. – WIKFORSS, Å. (2009): The Normativity of Meaning and Content. In: Zalta, E.N. (ed.): *The Stanford Encyclopedia of Philosophy (Summer 2009 Edition)*. URL = <http://plato.stanford.edu/archives/sum2009/entries/meaning-normativity/>.

GLÜER, K. – WIKFORSS, Å (Unp. Ms.): Reply to Whiting. http://people.su.se/~kgl/Reply to Whiting.pdf

HATTIANGADI, A. (2006): Is Meaning Normative? *Mind and Language* 21, 220-240.

HATTIANGADI, A. (2009): Some More Thoughts on Semantic Oughts: A Reply to Daniel Whiting. *Analysis* 69, 54-63.

HORWICH, P. (1995): Meaning, Use, and Truth. *Mind* 104, 355-368.

KRIPKE, S. (1982): *Wittgenstein on Rules and Private Language*. Cambridge MA: Harvard University Press.

KOŤÁTKO, P. (1998): Two Notions of Utterance Meaning. *Proceedings of the Aristotelian Society* 98, 225-239.

ROSEN, G. (2001): Brandom on Modality, Normativity and Intentionality. *Philosophy and Phenomenological Research* 63, 611-623.

WEDGWOOD, R. (2007): *The Nature of Normativity*. Oxford: Oxford University Press.

WHITING, D. (2007): The Normativity of Meaning Defended. *Analysis* 67, 133-140.

WHITING, D. (2009): Is Meaning Fraught with Ought? *Pacific Philosophical Quarterly* 90, 535-555.

WHITING, D. (2013): What Is the Normativity of Meaning? *Inquiry*, forthcoming (published online November 20, 2013).

WIKFORSS, Å. (2001): Semantic Normativity. *Philosophical Studies* 102, 203-226.

Performatives and the Role of Truth in Semantics

STEPHEN BARKER

Department of Philosophy. Faculty of Arts. University of Nottingham
University Park. Nottingham, NG7 2RD. United Kingdom
Stephen.Barker@nottingham.ac.uk

ABSTRACT: According to Austin, in uttering I hereby X in a performative we are neither asserting nor saying anything true/false-assessable about what we are doing, our Xing. Still in producing the performative utterance we can be said to say we are Xing. So, we have the production of a declarative sentence, that is perfectly meaningful and not lacking in content in any way, that is nevertheless not produced in an assertion nor open to evaluation as true or false, despite the fact that it says something. In this paper, I argue that Austin's claim about performatives is correct. I then argue that Austin's thesis about performatives has radical implications for received ideas about role of truth and truth-conditions in the explanatory enterprise known as 'semantics'.

KEYWORDS: Austin – performatives – semantics – truth-conditions.

In *How to do Things with Words* Austin saw himself as attacking the idea that language is fundamentally a system for describing reality. By *describing* we can take Austin to mean an activity that is either asserting how things are or the production of sentences that can be said to be either true or false of the world, depending on how the world is. So he is against the idea that in using language we are always asserting or holding up truth/false-assessable sentences—sentences that are open to truth or falsity assessment. Austin proposes much of the activity we undertake is not describing in this sense. Naturally, enough people knew before Austin that there were orders and questions, which look neither like assertions nor like productions of truth/false-assessable sentences. So, what's the interesting insight is Austin offering? It lies in the phenomenon of the *performative utterance*. The canonical form of a performative utterance is:

I hereby X.

Here *X* will be a phrase whose main verb denotes a kind of linguistic performance, like ordering, declaring, stating, promising, and so on. Some performatives are: *I hereby pronounce you man and wife*, *I hereby name this ship Baggins*, *I hereby condemn you to death*, and so on. Take just about any non-assertoric illocutionary act and you can transform it into a performative utterance. Take an order, *Leave!*. Then its explicit performative form is *I hereby order you to leave*. Take a question: *Who did it?* Its explicit performative might be: *I request that you tell me who did it*. Take an assertion, *Smith did*. Then there is the performative: *I say Smith did*.

So what's interesting about performatives? It's this. According to Austin, in uttering *I hereby X* in a performative we are neither asserting nor saying anything true/false-assessable about what we are doing, our *X*ing. Still in producing the performative utterance we can be said to say we are *X*ing. So, we have the production of a declarative sentence, that is perfectly meaningful and not lacking in content in any way, that is nevertheless not produced in an assertion nor open to evaluation as true or false, despite the fact that it s*ays* something. Instead of asserting, what we do with *I hereby X* is *say*, in a non-assertoric, true/false-assessable sense, that we are *X*ing and we *X*.

Austin's idea of performatives then requires that there be declarative sentences that describe how things are, they have propositional content that encodes propositions that match reality, speakers want audiences to recognize that these states of affairs obtain, but they are neither assertions of these states of affairs, nor are they truth/false-assessable tokenings of declarative sentences. Evidently, performatives are *descriptions* in the sense that they have meanings that correspond to how things are. We know which performative is being undertaken from the words used by the speaker. *I hereby pronounce you man and wife,* deploys the performative verb *pronounce*. Because of the celebrant uses this verb we know that a pronouncing is going on. The adjective *pronounce* is satisfied by the activity undertaken. Suppose you have produced the performative, *I hereby pronounce you man and wife*. The next day you can assert, using the same words but with some grammatical modifications: *I hereby pronounced you man and wife yesterday*. That is not a performative, but it is used with effectively the same words as the performative utterance, and clearly can be used to assert what the performative merely indicates.

So there is, and Austin admits this, a correspondence between the words used in the performative locution and the performance undertaken. But this involves no assertion. Austin's idea of saying without asserting is not totally obscure when we consider, for example, orders. In uttering *George, leave!*, I use the term *leave* to specify an act, *George is leaving*. But I am not asserting that this act will happen. So here we have another kind of saying that is not a describing or asserting of fact. So maybe performatives are like this (though we shall see that one crucial sense, there is a way that they are not).

According to Austin, instead of being truth/false-assessable, performatives are only felicity- and infelicity-apt. The felicity is correctness of performance, in which all the conditions for the correct performance of the speech act are met. Infelicity is a defective performance. Felicity conditions are not truth-conditions.

Why exactly is Austin's contention about performatives interesting? It is at odds with a very well entrenched paradigm about meaning and truth and the link between them. The thesis RT sums up one aspect of the link:

> RT A sentence S is true/false-assessable iff S has propositional content, it is a representation of how things might be.

Of course, not everyone accepts the correspondence theory of truth, but the bland thesis invoked in RT does not amount to a theory of the property of truth but just of what is necessary and sufficient for possessing truth, whatever truth turns out to be. RT conveys a semantic idea of truth: the thesis that meaning is given by truth conditions. But Austin's contention about performatives is inconsistent with RT and so inconsistent with truth-conditionalism about meaning. That's because he is claiming declarative sentences with propositional content may not be true/false-assessable, contra RT.

The vision of language that Austin was attacking with his performatives is alive and well. It's just the standard view. Despite the attractions of the standard view, I think Austin is right, in his contention, and his views about performatives have quite radical implications about how we are meant to approach the issue of meaning.

To defend Austin, I will argue that performatives contra theorists like Bach and Harnish (1979) and Searle (1989)—are not assertions, unless, that is, they are of the form *I assert/say/state that P*. I will then argue that they are not true/false-assessable either. I then consider the implications of these conclusions.

1. Unasserted

Bach (1975) and Bach and Harnish (1979) think that Austin is wrong that performatives are not assertions. According to them, in producing a performative one asserts that one Xs and one Xs. The performative involves an indirect speech acts. An example of an indirect speech act is this. You ask me: *Are you going to the party tonight?* I answer: *I have to work.* I indirectly answer your question—I indirectly perform the assertion of *I am not going*, answering your question, through another speech act, which I explicitly perform. Bach and Harnish see performatives as doing just this: I assert, *I hereby X*, and I am asserting that I am Xing, but I indirectly X.

Bach and Harnish's picture is integrated into a Gricean framework of conversational implicature driven by speakers' following conversational norms. In taking in speaker U's production of a performative, an audience H is meant to reason as follows (drawn from Bach 1975, 234):

(1)   U is saying 'I X'.
(2)   U is stating that she is Xing.
(3)   If true, then U is Xing.
(4)   If U is Xing, then it must be this utterance that constitutes the Xing.
(5)   Presumably, U is speaking truly.
(6)   Therefore, in uttering 'I hereby X' U is Xing.

Their view then is that performatives are assertions about an illocutionary act that the speaker indirectly performs.

Should we accept Bach and Harnish's view? I think not. Consider the reasoning we as audiences are meant to undertake in interpreting a speech act as a performative. Step (4) is meant to be an inference to utterance of 'I hereby X' constituting the performative. That inference looks open to doubt. If I look you in the eye and assert, *I am ordering you to leave*, there is still some question about the order itself. When was it given? Is it somehow being implicitly given? Is this an announcement that an order will be issued at some stage? All these questions arise. But why should the assertion itself be taken to be the order? If someone asserts: *I am adjourning this meeting*, there is some question about whether the meeting is adjourned. For example, one could say, without redundancy: *I am adjourning this meeting: meeting adjourned*. So, asserting *I am adjourning this meeting* does not *constitute* adjourning the meeting as such. But saying *meeting adjourned* or

I hereby adjourn the meeting can constitute adjourning the meeting. In re-
sponse to *I am adjourning this meeting*, an audience could always ask, *Is it
adjourned?* Whereas, they cannot pose the same question in response to
I hereby adjourn the meeting. In the latter case they can dispute that it is a
good time to adjourn, but not that the directive to adjourn the meeting
have been issued.

Bach and Harnish's theory that performatives are assertion used to per-
form indirectly non-assertoric illocutionary acts does not explain this basic
fact. The issue is the step (4) in the reasoning above. Why conclude that
the act of assertion is also the act of *X*ing? It seems all the audience can
conclude is that *X*ing will go on sometime.

One might object: what secures the inference in (4) is the fact that in
performatives we use *hereby*. I say, *I hereby X*, as in *I am performing with this
very utterance a curse on you*. But there are several things to say here. First,
it strikes me that in saying, *With these very words I am cursing you*, I invite
the reply: *You are cursing me?* or *Is that a curse?* Secondly, and perhaps more
importantly, we do not have to put *hereby* in with performatives to carry
out the performative. I say, *I curse you*, and that is my curse. But change
slightly the tense, to make it clear I am making an assertion, as in *I am
cursing you*. My assertion of *I am cursing you* is not as such a curse. Perhaps
an audience can infer that I am inwardly cursing you, but my utterance as
such is not the curse. That is despite the fact that my audience can
through inferential mechanisms infer that I have certain dispositions to
undertake cursing. But no act of cursing has been undertaken. Whereas ut-
terance of *I curse you* does not leave any of this open and does constitute
a curse.

According to the Bach and Harnish line, we infer that U's utterance
constitutes an *X*ing. But what is it to utter a sentence and directly *X*? It is
presumably to produce a sentence *S* with such and such intentions. That's
what constitutes the *X*ing. So if I curse you, I produce an utterance with
intentions that my deep hatred with what you have done is manifest, and
I make no assertion. Cursing cannot be a secondary inferred speech act!
Cursing is upfront, primary, overt, and direct, otherwise it is not cursing.

This suggests another problem with the indirect-speech-act view. Why
would speakers want the extra illocutionary act to get in the way of the
*X*ing, since *X*ing is the main point. What we want is that the phrase, the
description, *I hereby X*, indicates what I am up to, but we don't want any-
thing as heavy duty as an assertion. Perhaps, *I hereby X* remains unasserted,

but plays the role of describing. Describing can go on without asserting. This may be exactly the kind of describing that goes on when I utter, in an order, *George, leave the room!*. In this case, the speaker describes a possible state of affairs, the one they want realised, but no assertion goes on. Still, with my utterance of *You are leaving!* I directly perform an order, as much as I do with *I order you to leave*. This does not quite capture what is going on in the performative. The utterance of the phrase *I curse you* is a non-assertive description of my act of cursing and it does correspond to something that is the case. But it's just an indication of a state of affairs, somewhat like the way we indicate something through *conventional implicature*—see Barker (2003). In saying, *Even Granny got drunk* I communicate that it is comparatively surprising for me that Granny got drunk, but I do not assert that it is so.

I think that shows that performatives are not indirect speech acts but primarily performances of orderings, commandings, adjournings, that is, of *X*ings, and not acts that are constituted by assertions allowing us to infer that a performance of *X*ing is going on.

Searle (1989) agrees that performatives are not indirect speech acts. But he thinks they are assertions as well as *X*ings. His argument that they are assertions is simply that their utterance brings about a state of affairs that makes the propositional content of the performative true. My utterance of *I hereby X* brings about my *X*ing and that corresponds to the content of the sentence I utter. But it does not follow from this fact that performatives are assertions. The fact that a sentence has content that corresponds to how things are—in other words, descriptive content—does not make self-standing utterance of the sentence an assertion.

We have positive reasons to think that performatives are not assertions. Here's a test for assertiveness. If I really produce an assertion that *P*, I can append my claim with *that's my belief*. Take the following case. I can say: *I thereby pronounced them man and wife: that's my belief*. There is no issue here of reflexivity being the problem. I can say: *I am now commanding you to leave: or that's my belief*. But try this with the performative: *I hereby command you to leave: or that's my belief*. That's odd. So, it seems that in producing a performative, the speaker is not expressing belief with the sentence. Or at the very least, we have some evidence that this is the case.

Another test for assertion is that along with an assertion we can insert a sentence that comments the state of affairs that the assertion commits us to. If I really assert *P*, then I can follow it up with some commentary on

the state of affairs that we are asserting is the case, such as *But the people did not like it*. So, one can say something like this: *I thereby declared the game over. But people didn't like it.* One can say: *I will declare/have just declared the meeting over. But people won't like it.* But one cannot say (this is the performative): *I hereby declare the meeting over. But people won't like it.* In this case, the question of what *it* in the commentary sentence is referring to arises. But why should that be if there is an assertion going on? The fact that one is doing something else as well should not be an issue. So why cannot one comment on the state of affairs to which the assertion commits us? If we can make no comment that would suggest that we have made no assertion.

We cannot explain the inability to make comments of this kind in relation to performatives with the hypothesis that we cannot comment on secondary assertions. Even if an assertion is secondary, say in a non-restrictive relative clause, it looks accessible to commentary. Take this case: *The Queen, unlike the PM, likes people. That fact won't please the PM's wife.* We can access the secondary assertion, that the PM does not like people, with the right context. But there seems to be no way to access the assertion in the performative, as we have seen above. But that suggests that no assertion is being made with a performative.

Another test for assertion involves inference. Imagine this dialogue. *A: Are you going to come to a movie? B: I am at the office. A: You are working too hard.* That's perfectly fine. In other words, *B*'s primary assertion involved in an indirect speech act can be the basis for an inference by *A*, that is, that *B* is working too hard. A secondary assertion produced in a relative clause can also be the basis of an inference: *The Queen, unlike the PM, likes people. So, the Queen differs from the PM.* It does not matter if assertions are ensconced in some broader context as secondary assertions, or if they're deployed in indirect speech-acts. As assertions they can be the basis for inferences. But it is odd to say: *I curse you, curse you. So, I am angry with you.* On the other hand, one can infer: *I am cursing you right now—under my breath—so I am very angry with you.* One cannot say: *I wonder if he is mad. So, I am puzzled.* Whereas, one can say: *I was wondering, then, if he was mad, so I was puzzled.*

Not all sayings are assertions, even if the sayings are used to indicate that some state of affairs obtains. I think there is a genuine motivation for speakers to want to exploit non-asserted indications or sayings of how things are. We do not want all such indications of how things are to be assertions, since indications that are assertions bring what should otherwise be background to the foreground. To make an assertion is to treat a given

subject matter as a topic of a conversation. But in a performative, the fact that one is performing a specific kind of illocutionary act shouldn't be the topic of conversation when the performative is issued. That would be self-defeating.

Others have made this point about the nature of assertion being in conflict with the idea that performatives are not asserted. Jarry (2007), invoking Barker (2004), proposes that assertions are *defensive*, and democratic, whereas, other speech-acts, like orders, are autocratic. The point of assertion is to raise a commitment to the level of dialectical engagement. That means that the purpose of assertion is to manifest a disposition to display reasons for a commitment. This is an approach in the spirit of Brandom (1998). In contrast, other speech-acts, like orders, are acts whose purpose is to manifest a state, like desire. The order produced by uttering *Leave the room!* involves the speaker manifesting a desire that the audience leave the room. This is distinct from an assertion of *You ought to leave the room*. This is an utterance in which a speaker defends the state of desiring that the audience leaving the room. Its purpose is to invite a potential response from the audience to provide reasons for such a desire in themselves. They can reject the desire—not forming the intention to leave—and express that rejection by affirming that the speaker's utterance is false. Truth/falsity-assessment is to the fore because the utterance is an assertion, and indeed, the speaker in uttering *You ought to leave the room*, is implicitly inviting such response, in contrast to the issuing of the order, *Leave the room!* One can say precisely the same thing for *I order you to leave the room*.

If this point about identification of assertion with a certain dialectical purpose is correct, then we don't want to think of performatives as assertions. Assertion gets in the way, since it invites audiences to have a dialectical response to possession of the kind of mental state the speaker is expressing with the whole performative, which is to say, an epistemic state about the speakers *X*ing. We have no interest in thinking of performatives as assertions if we think of assertions in this way. This amounts to wasting the specific power of assertion, which has one job to do, whereas performatives have another—each job depending on what the *X*ing is. Assertion could only get in the way. This explains some of the phenomena we have observed above.

Indeed, a certain class of performatives, which we have not yet looked at, only confirms the no-assertion thesis being made here. These are performatives of the form:

I assert/state/claim etc. that P

So an example is:

I state that I have never been in the vicinity of Prague.

Here it seems we have all the marks of an assertion of *I have never been in the vicinity of Prague*. Someone can respond to the above performative with, *You liar! You have been there*. Or with *And that's a fact*. Or, *Yes, I believe you: you have never been there*. And so on. The content of the assertion is not that the speaker is making an assertion, that which the whole performative sentence describes, but the content corresponding to the complement sentence, namely, that the speaker has never been to Prague. We don't have a double assertion.[1]

This confirms the idea that the act the speaker is in fact performing with the performative is what they are describing themselves as doing. In the case of an utterance of *I order you to leave* they are performing an order, as they describe. In the case of utterance of *I state that I have never been to Prague*, it is a statement, that is, the assertion that the speaker has never been to Prague. In other words, in uttering the performative, *I hereby X*, the speaker performs an *X*ing and describes themselves as going so. However, the act of describing is not an assertion.

2. Unasserted but true/false-assessable?

Let us suppose then that performatives are not assertions. So, we can agree with Austin that in uttering a performative, *I hereby X*, the speaker says something with the whole sentence but that this saying something is not an assertion that they are *X*ing. This concession, however, does not mean in itself that the sentence *I hereby X* is not true/false-assessable—Austin's second thesis about performatives. Being unasserted does not imply not being true/false-assessable. For example, constituent sentences of a logical compound are not asserted, but they are true/false-assessable. In other words, we can ask in relation to unasserted sentences whether they are true or false: *The sky is blue. (Unasserted) True or false?* The

---

[1]   The very phenomenon we are examining here is the performadox of Boër and Lycan (1980) which they see as a paradox for compositional semantics.

speaker is not asserting the sentence concerned, but the enquiry about its truth makes sense as does an answer of *true* or *false*. So, the question in play is whether we should think of performative sentences *I hereby X* and so on, as being in the class of unasserted but nevertheless truth/falsity-assessable sentences.

Here is a phenomenological argument against truth/falsity-assessability. Performatives just don't feel like they are true/false-assessable. If someone intones, *I curse you*, the response, *false*, looks inappropriate. In relation to the utterance of *I hereby adjourn this meeting*, the response: *true* looks wrong. Or, take this dialogue: Smith asks Jones: *say something true*. Jones replies: *I order you to stop making that request*. Smith cannot say that he has got Jones to do what he wants, namely, to say something true. Suppose, in a ceremony, you intone: *I appoint you president of the club*. Do I say something false, if you are already president? I think not.

These phenomenological arguments for non-true/false-assessability may not move you at all. Instead of trading intuitions about the appropriateness of true/false-assessment, we can move to a prior question: why wouldn't performatives be true/false-assessable? Given the representationalist thesis *RT* above, surely they are. But maybe, contra *RT*, being a sentence with propositional content does not make you true/false-assessable. Just consider declaratives used to perform non-assertoric illocutionary acts. Take the following case. I am very angry with you for having taken some money from my piggy bank with the intention of spending it on a beer. I say: *You are going to put that money back*. The response: false (or true) is inappropriate. To reply, *false* involves a subtle misunderstanding of what I am up to in making the utterance. I am ordering you to put the money back and not making a predication about what you will do. Indeed, an audience can enquire: *Is that an order or a prediction?* Depending on how the speaker responds, the judgment of true/false-assessability follows accordingly. Or take a question. I can ask: *Fred has a lot of time on his hands*, with a certain rise in tone at the end of the sentence indicating a question. Clearly I am posing a question. A response *true* or *false* is not appropriate. Or I say *This meeting is adjourned* (banging a gavel). The response, *That's false*, is inappropriate. In all these cases we have sentences that correspond to states of affairs that are the case (or not), but true or false are inappropriate. But according to *RT* they are all true/false-assessable.

Here is a general line of argument for the thesis that performatives are not true/false-assessable. Take any sentence that is produced with a non-

assertoric force. In any such case the sentence may be a declarative sentence used with non-assertoric force—say, *You are leaving* (as an order)—or non-declaratives, *George, leave!*. In all these cases we can see the sentence as providing a representations of how things are, even if there is no explicit declarative form in the sentence produced that expresses the representation. (See Price's 1988.) For example in the case of an order, *George, leave!*, there is a representation, say, that I want George to leave. For clearly, in producing the order one is representing the fact that one wants someone to leave. But that is a representation with propositional content. So why isn't the resulting sentence, which is a production of that representation, true/false-assessable? One also represents a state of affairs that George leaves. If that state of affairs obtains, which it may do, then there is a fact, represented by the sentence, that obtains. If so, why isn't it true?

We have then three classes of sentences to think about:

1. Non-declarative sentences used in non-assertoric speech acts, as in *Leave, George!*;
2. Declarative sentences used in non-assertoric speech acts, as in *You will leave!*;
3. Sentences used in performatives, that is sentences that can be said to offer descriptions of illocutionary acts being formed with those very sentences, for example, *I order you, George, to leave.*

I think we want to say that class 1 sentences are not true/false-assessable. Similarly, class 2, orders and questions performed with declaratives, are not true/false-assessable. Given that class 2 sentences are not truth/falsity-assessable, then we have reason to believe that performatives are not truth/falsity-assessable, since they too are just unasserted declarative sentences with non-assertoric force. Class 2 sentences describe possible states of affairs, but their doing so does not mean they are truth/falsity-assessable. Similarly, class 3, the performatives, describes states of affairs, but that in itself cannot mean they are true/false-assessable as such.

Pan truth-conditionalism is the idea that all sentences are true/false-assessable. The principle RT implies a pan truth-conditionalism. Such views have been held. Lewis's (1970) conception is a paradigm example of such pan truth-conditionalism. According to Lewis, all speech-acts are like performatives, and all are truth/false-assessable. They all have the form:

I hereby X that P.

On this view, assertion isn't the specialised act, one amongst others, that it seems to be. Rather, it has a privileged place. It is the representational act par excellence. On the Lewis view all speech-acts are this kind of act. Lewis's position, however, just looks like defiance in the face of evidence to the contrary. Indeed, it may be the view that all the evidence to the contrary is just in the sphere of the don't-cares. But this is a vast sphere, and surely ordinary speakers cannot be so wrong about the nature of their own language. Any view of meaning as constrained by use cannot agree with Lewisian high-handedness.

Davidson's view is a more tempered version of pan truth-conditionalism. Davidson's (1979) conception is very similar insofar as all declarative sentences are true/false-assessable. For Davidson, mood-modified sentences, like *You, leave the room!*, are really two sentences,

(S)   You will leave the room.
(R)   This is an order.

Davidson's analysis is meant to explain hesitation to assign truth or falsity to these sentences, even if they are basically true/false-assessable. The hesitation is meant to be there because we are disinclined to assess sequences of sentences, *S*, *R*, for collective truth or falsity, unlike, say conjunctions, (*S* & *R*). However, Davidson's conservative light touch explanation is not that convincing. Why shouldn't we assess sequences of sentences for truth? Cannot I say in response to a text involving more than one sentence that it is all true? If someone issues a sequence of sentences like the above, one can say: *True, you have given me an order*. But that kind of response cannot be made to an actual order, produced through a sentence in the imperative. Davidson's position looks untenable.

3. The tripartite analysis and compositional semantics

That is my assembled evidence for the non-truth/false-assessability of performatives as a general class. I think Austin's original intuition is right. Performatives are not assertions and not true- or false-assessable. But in producing a performative one says something. But the saying is just the production of a declarative sentence with content, indeed, representational content. Although we speakers intend to say how things are, intending to say how things are is not asserting or producing true/false-assessable words.

Why are these matters of real concern? Why not just admit, as surface considerations and some general considerations about assertion seem to show, that performatives, and the other sentences we have looked at, are not assertions, and not true/false-assessable as such?

In fact, there is a general, theoretical reason to be very concerned with the question of the true/false-assessability of such sentences. In *How to do Things with Words?* Austin begins with the constative/performative distinction. He sees constatives as contrasted with performatives. Performatives are doings in a way in which constatives are not doings. But that distinction begins to break down, as he proceeds in his lectures. The supposed characteristics of performatives also seem to be shared by constatives. Constatives are also speech acts, doings, but just a different kind. That in fact is confirmed by the fact that the *X*ing that a performative can describe is asserting, as in *I say you are such a fool*.

That leads to a revision in which performatives lose their place as a fundamentally different kind of activity with language. What replaces it is Austin's analysis of speech-acts. Self-standing speech acts with sentences have a tripartite structure of three nested acts. The first act is the production of a grammatical sentence. The second and third acts are the *locutionary* and *illocutionary* acts respectively:

> *Phrastic act*: utterance of a grammatical sentence.
> *Locutionary act*: utterance of a sentence with reference and linguistic meaning and a propositional content.
> *Illocutionary act*: utterance of a sentence with a certain propositional content and a force.

By *force* we mean a specific use made of a proposition in relation to an audience. The standard view goes roughly (but not necessarily accurately) as follows: assertive force is intending others to come to believe P, commissive force is intending others see to it that P is made true, and so on. There is also a fourth act: the perlocutionary act, which is the securing of an effect in the audience. We shall not be closely concerned with it here.

The view about performatives we are contemplating as Austin's does not sit well with this general speech-act analysis. The problem is the status of the saying that goes on in the production of a performative *I hereby X*. In that act, the speaker U is meant to be describing the illocutionary act, the particular *X*ing, U is undertaking, but that describing is a saying that is neither an asserting nor true/false-assessable. It is like the kind of saying

that goes on in production of *You will F* produced as an order. But since this kind of saying is not an illocutionary act, we might, in the light of our tripartite analysis, want to say it is a locutionary act. But sentences tokened in locutionary acts, we might suppose, are true/false-assessable. At least, that is the orthodox view, which is, more or less summed up in *RT*, and the representationalist idea of true/false-assessable sentences. If so, performatives must be true/false-assessable, contrary to our recent arguments and Austin's contentions.

However, I have argued, following Austin that performatives are not true/false-assessable. Neither are all the sentences performed in non-assertoric illocutionary acts. But if that is right, sentences produced in the performance of locutionary acts cannot be true/false-assessable. But now we face the crunch. The idea that true/false-assessable sentences are just sentences produced in locutionary acts is central to truth-conditional approaches to meaning. The truth-conditional approach to meaning claims that the meaning of a sentence is given by its truth-conditions. Meaning here means *propositional content*: the meaning a sentence has excluding its force, which is to say, the meaning possessed by a sentence in a locutionary act. In other words, to possess a propositional content is to be a sentence performed in a locutionary act, which means having truth-conditions, and so being true/false-assessable.

The sense/force distinction

The tripartite analysis is at one with the sense/force-distinction. The latter is often seen as the heart of compositional semantics. It's claimed that the only way we can give a systematic account of the content is by separating content from deployment or use of content. Take three mood-modified sentences, as below:

Lucky jumps.
Does Lucky jump?
Lucky, jump!

All involve the same basic vocabulary but combined differently. The content that *Lucky jumps* shares with the other sentences must be independent of its assertoric use, since the other sentences are not, apparently, associated with assertion. Thus the common content must be assertion-independent. What is this assertion-independent content if it is not propositional con-

tent? Let's denote it by ⟪Lucky jumps⟫. If we accept this hypothesis, then, the structure of the three sentences is given as below, where A, Q and O are the three distinct forces:

Lucky jumps	A⟪Lucky jumps⟫
Does Lucky jump?	Q⟪Lucky jumps⟫
Lucky, jump!	O⟪Lucky jumps⟫

That there is a common content seems to be confirmed by the fact that the truth-conditions for the declarative sentence, affirmative-answer conditions for the interrogative, and compliance conditions for the imperative can all be specified by the same condition: that Lucky jumps. It's this common content that is the object of semantic investigation, and whose assignment to sentences is a function of the semantic composition of those sentences.

If this is right, all sentences have a truth-conditional component, which is ⟪Lucky jumps⟫. But this picture of compositional semantics and the force/sense-distinction gives rise to two problems:

Problem 1: A force-operator, supposedly, is not like a logical operator like negation. If a speaker U utters *Not-S*, then one cannot say that U has said something true, namely that S, or something false, that S. In short, S is not true/false-assessable in an utterance of *Not-S* despite the fact that in uttering *Not-S*, U utters the sentence S and expresses a proposition that S thereby. The reason, obviously, is that negation prefixes S. Because negation contributes to truth-conditional content, it seals off S from true/false-assessment in the sentence *Not-S*. But force operators don't contribute to truth-conditional content. They don't incorporate the content in their scope within some larger truth-conditional content. So, they don't seal it off from true/false-assessment. If that is right, then why can't we access all illocutionary acts in terms of truth and falsity, given that all of them involve a component of their meaning, which is the locutionary act, supposedly, a true/false-assessable act?

Problem 2: We are assuming, as seems to be right, that a force operator is not making any truth-conditional contribution to a sentence. But now consider *grammatical mood*, the linguistic indicator of force. Grammatical mood in a sentence indicates that a force is in place. Why isn't mood a truth-conditional contributor to the content of a sentence? In other words why can't the mood operator be seen as the Lewisian suggests, as being like a performative verb, so that the whole mood-modified sentence

comes out as: *I X that P*, where *X* is a force verb? In which case, why aren't all mood-modified sentences true/false-assessable?[2]

One cannot reply here that it is just a matter of stipulation that mood-operators are not truth-conditional content contributors. Natural languages are not the result of stipulation, like artificial 'languages'. In other words, we have to come up with a general theory of what it is for an expression in a language to contribute to truth-conditional content. We need in other words a theory of what truth-conditions are as a phenomenon. Clearly, we don't want mood operators to come out as truth-conditional operators. But now note. Any explanation of why mood-operators are not contributors to truth-conditions is probably also going to have the result that in performatives, the indicator of performance, 'I hereby X' is not going to be a truth-condition contributor. In other words, performative sentences are not going to come out as truth/false-assessable sentences. That is, we should accept:

> Mood-modifiers are not truth-conditional operators iff performatives are not true/false-assessable.

This naturally enough supports all the arguments we rallied above for the Austinian thesis that performative are not true/false-assessable.

So here is our problem: why aren't mood-modifiers truth-conditional?

Problem 3: If the objects of attitudes are propositions, and propositions are the primary truth-bearers, in the sense of locutionary-contents of the tripartite analysis—then why are beliefs truth/falsity-assessable whereas desires are not truth/falsity-assessable, given that both are mental states with propositions/thoughts/locutionary-contents as their objects? So, we accept *B* but we don't accept *D*:

> *B*: My belief that the world will get better is true
> *D*: My desire that the world will get better is true.

There is a close connection between our current concerns with illocutionary acts and this problem about attitudes. Orders if anything correspond to desire states, and assertions correspond to beliefs. This is a bit crude, of course, but it has some validity. This correspondence is reflected in speech acts. Saying *I want you to leave* to someone is as good as a request, a kind of commissive,

---

[2]   In effect this is the question explored in Price (1989).

that they leave. I have expressed, indeed, I have represented the mental state that I want you to leave. It seems we explain why beliefs are truth-apt whereas desires are not, by appeal to direction of fit. But as an explanation, that assumes that it's not propositions, thoughts, that are truth/false-assessable, but mental states with a certain relation to the world, the state-world direction of fit. But the position of the tripartite analysis is that it is sentences expressing thoughts that are truth/falsity-assessable.

The Predicament in Outline. These three problems are all related in different ways to the thesis that there are *locutionary acts*—force-neutral acts performed with sentences—and that these acts, or the propositions they express, are the primary objects of true/false-assessment. Truth, in short, cannot play the theoretical role that truth-conditional semantics proposes that it does, as the key property that allows us to theorize locutionary content. So where can we go from here?

One obvious line of solution is to find an alternative understanding of locutionary acts. We need another property that sentences can have, which isn't truth, which defines locutionary content, that is, a substitute for truth-conditions.

4. Truth: divide and conquer

There is a simple proposal on how to do this. That's to divide truth into two kinds. So, Stenius (1967) and Sadock (1985), propose that there are two kinds of truth, one applying to the proposition—the force neutral content—and one applying to assertions or sentences thought of as being potential assertions. Call the second *folk-truth*. The idea then is that we cannot apply folk-truth to performatives, and orders, and so on, because they are not associated in the right way to assertion, whereas we can apply theory-truth to them.

The problem with this approach is that we have no reason to think that the term *true* is ambiguous in this way. We never say things like: *This sentence is true, but it's not true* in the way that we can say *This is a bank, but it's not a bank*, pointing at the shore of a river, indicating thereby that the sentence should be interpreted: *This is a bank (river edge), but it's not a bank (financial institution)*. One response to this problem is that *theory-truth* is not really a term in ordinary parlance. It simply isn't part of the non-theoretical lexicon. But then we have to ask what grounds we have for bes-

towing the term *true* on this theoretical property, supposedly playing a theoretical role? In other words, no interpretative theory can justify the thesis that *theory-true* is a disambiguation of the term *true*. I don't think the ambiguity theory can be right for that reason.

One might object that this is a minor concern. We can forget about who owns the term *true*. Why not propose that semantics just describes in a compositional way how sentences get a certain, theoretically defined property in terms of the semantic properties assigned to sentence constituents and their mode of composition. The property is not *having-such-and-such-**truth**-conditions*. Let's suppose rather that it is *having-such-and-such-**representation**-conditions*. So, all sentences, let's say, are used to represent how things might be, but this does not make them true/false-assessable. Being true/false-assessable is not merely about having representational content. Sentence content—locutionary-act content—is linked to representation conditions and not to truth-conditions.

What are representation conditions? For example, using the tools of possible-worlds semantics, we could identify the representational conditions of a sentence with, say, a set of worlds. We propose then that a sentence matches the world @, when @ is in the set of worlds. This used to be called truth, in possible worlds semantics. But now we are rejecting that identification. We are just saying it is *matching*, or *describing*, where these are purely theoretical terms. We then work out a compositional semantics in which sentences are assigned worlds in a systematic way, based on assignments of reference to their ultimate constituents.

Which set of worlds gets assigned to a sentence? Here we encounter a slightly embarrassing issue for the possible worlds approach. We cannot say that the worlds are those in which the sentence is true, since that is to bring in truth and implicitly the idea of a truth-bearer. So our analysis would be, at best, circular. We need to find something common to all works, apart from the truth of *S*—that can tie them all to *S*. The obvious idea is that what's common to all the worlds is a state of affairs, in the sense of an abstract, non-concrete one, and the sentence is tied to this state of affairs. But if we have to bring in states of affairs in this way, we ought to drop worlds—they have now become entirely redundant.

Think then of representation conditions in terms of abstract, non-concrete states of affairs. States of affairs are structured entities that correspond to the form of a concrete (actual) state of affairs, in the sense of Armstrong (1997). We can then say that the sentence *describes* the

world @ if and only if the state of affairs it represents is realized in @ by some fact. This is very much in the terrain of the correspondence theory of truth. But again, I emphasise, we are not proposing at this point to identify truth with correspondence.

What is this relation we have called *representation* that holds between a sentence and a state of affairs? It's easy to think we can say *representing a state of affairs* and think we know what we are talking about. But that appearance of clarity may be mere appearance. What is *representing* here? It cannot be *denoting*. We don't want the sentence to be a referring term, of the form: *the state of affairs that S*. Sentences don't denote states of affairs. Of course, we can see some expressions as denoting states of affairs, phrases like, *the state of affairs that S*. But such phrases are not sentences.

You might say that the representational content of S is fixed compositionally. That S represents a state of affairs is a function that of S's parts and their mode of combination. We might suppose the representational content of a predicate is its denotation, a property, and that of a singular term, is an object. In which case, where S is an atomic sentence, we get:

The sentence O *is* F represents the state of affairs ⟨x is y⟩, in which O denotes x and F denotes y.

But would specifying all such axioms required for a compositional account of sentence representation tell us what representation is? No. It will not since such an enterprise assumes the very relation of sentential representation we are trying to illuminate. It just tells us which states of affairs are represented. It does not tell us what representation of a state of affairs *is*.

We might see parallels between the present proposal and Wittgenstein's Tractatarian view of sentences as word-pictures of states of affairs. Sentences picture states of affairs, which is a kind of isomorphism between the sentence with its grammatical constituents and the state of affairs. Indeed, this might give us our representing. S represents just in case constituents in S denote constituents in ⟨s⟩, and the order of S mirrors the order in ⟨s⟩. This might work, as long as we have a theory of constituenthood of sentences and that of states of affairs. But there are well-known problems with such ideas.

We might insist that the relation of representing a state of affairs is a primitive. In other words, there is just a kind of language-world relation—distinct from denotation—that sentences have to states of affairs. This is perhaps objectionable, and would be seen as a retrograde step back

into the obscurities of past theory. Moreover, there are all sorts of ontological concerns about this kind of approach, and indeed, about states of affairs as an ontological category. I will not dwell on these here, since there is a more serious concern, residing in how exactly this approach is going to solve our problem of true/false-assessability.

We see the problem when we focus on what the story of true/false-assessable is going to be. Supposedly, the view is this. Given S represents a state of affairs, S is true/false-assessable just in case it is asserted, or thought of as a potential assertion. (In short, the illocutionary act of assertion fixes what true/false-assessability is, and not the locutionary act.) For this account to work, we need to give an account of assertion. An assertion, say, drawing on the Brandomian picture, is the following kind of act. U utters S with purpose of defending a mental state that is a commitment to S's *describing* reality. That means U is expressing a disposition to offer reasons for a mental state, commitment to S's matching how things are, that is, S represents a state of affairs that is realized.

In uttering a performative, goes the line of thought, one is not doing this. Rather, one is merely expressing the commitment that S describes how things are. The dialectical element of intending to defend this commitment is missing in the second case. Hence performatives are not truth-apt. Thus although performative sentences, *I hereby X*, describe the act of *X*ing that is going on, they are not assertions about it, nor are they true or false. In other words, we have sentences that describe reality, in our rather technical sense of *describing*, but which are not truth-falsity-assessable. They are simply not being used with the purpose of being viewed in that way.

That something is not entirely resolved in the present proposal comes out when we look again at our treatment of assertion. The current idea is that a speaker U asserts that S then U utters S with the purpose of defending a mental state, namely: Commitment to S describing how things are. This commitment looks like a propositional attitude. You may not think of commitment as belief, strictly speaking, but it looks like an attitude towards *a thought*. The thought is: *S describes reality*. But if this is a thought then our account of true/false-assessability already presupposes that there is something prior to assertion that is true/false-assessable, namely, the objects of attitudes: *thoughts*. But if this is right, then it seems it will be difficult for us to deny that sentences that simply express thoughts in this sense are true/false-assessable. But then we shall be firmly back in semantic

orthodoxy and unable to explain why performatives are not true/false-assessable.

In essence, the problem is that we are treating assertion as an act whose mental antecedents are states—commitments to thoughts—that already have content of the very kind we are meant to be explicating, viz, true/false-assessable content. We need, it seems, to deny that the mental antecedents of assertion are mental states that have content, in the sense of true/false-assessable thoughts, as their objects. What theory of assertion could meet this condition? Two approaches are:

(*i*) Brandom (1998) treats assertion as expressing commitment not to sentences representing how things are, but commitment to engaging in activity with sentences, which lack content as such. Rather, the activity with the sentences somehow bestows content on them. Needless to say, Brandom calls this activity inferential activity. In short, inferential behaviour is prior to sentence content.

(*ii*) We treat assertion as expressing mental antecedents, but give up the idea that we should think of them in terms of content or even in terms of commitment. One way of doing this is to generalize certain basic ideas of expressivism to all assertion. Moreover, instead of giving a theory of sentence content, we provide a kind of expressivism about content. This approach has been sketched in Barker (2007; 2014).

It may be with one of these approaches we can deal with our three problems, articulated in the last section above. Both ways, however, undermine the tripartite analysis to some degree. Both are closer to the idea that we explain assertion first, and can then talk of true/false-assessable sentences afterwards. Both seriously undermine the idea that truth and truth-conditions have a central role to play in the theory of meaning.

I will not explore the prospects of such approaches here. Perhaps there are other possibilities to explore. We have faith that some kind of theory is out there somewhere. But my intention here is not to arrive at that theory. It is rather to show the radical kind of critique of received ideas about sentence content implicit in Austin's ideas about performatives, and why they are still here to challenge us now.[3]

References

ARMSTRONG, D.M. (1997): *A World of States of Affairs*. Cambridge: Cambridge University Press.

AUSTIN, J.L. (1962): *How to Do Things with Words*. 2nd Edition. Ed. J.O. Urmson and M. Sbisá. Cambridge, MA: Harvard University Press.

BACH, K. (1975): Performatives are Statements. *Philosophical Studies* 28, 229-236.

BACH, K. – HARNISH. R. (1979): *Linguistic Communication and Speech Acts*. Cambridge: MIT Press.

BARKER, S.J. (2003): Truth and Conventional Implicature. *Mind*, 112, 1–33.

BARKER, S.J. (2004): *Renewing Meaning*. Oxford: Oxford University Press.

BARKER, S.J. (2007): Semantics Beyond the Distinction between Sense and Force. In: Tsohatzidis, S. (ed.): *Illocutions, Institutions and Intentionality: Essays on Themes from the Philosophy of John Searle*. Oxford: Oxford University Press.

BARKER, S.J. (2014): Pure versus Hybrid Expressivism and the Enigma of Conventional Implicature. In: Fletcher, G. – Ridge, M. (eds.): *Having it Both Ways: Hybrid Theories and Modern Metaethics*. Oxford: Oxford University Press.

BOËR, S. – LYCAN, W.G. (1980): A Performadox in Truth-Conditional Semantics. *Linguistics and Philosophy* 4, 71-100.

BRANDOM, R. (1998): *Making it Explicit*. Harvard: Harvard University Press.

DAVIDSON, D. (1979): Moods and Performances. In: Margalit, A. (ed.): *Meaning and Use*. Dordrecht: Reidel.

JARY, M. (2007): Are Explicit Performatives Assertions? *Linguistics and Philosophy* 30, 207-234.

LEWIS, D. (1970): General Semantics. *Synthese* 22, 18-67. Reprinted in: Davidson, D. – Harman G. (eds): *Semantics of Natural Language*. Dordrecht: Reidel 1972.

PRICE, H. (1988): *Facts & The Function of Truth*. Oxford: Basil Blackwell.

SEARLE, J.L. (1989): How Performatives Work. *Linguistics and Philosophy* 12, 535-558.

SADOCK, J.M. (1985): On the Performadox, or a Semantic Defence of the Performative Hypothesis. *University of Chicago Working Papers in Linguistics*, Vol. I, 160-169. Chicago: University of Chicago Department of Linguistics.

STENIUS, E. (1967): Mood and Language-Game. *Synthese* 17, 254-274.

A Puzzle about Rigid Designation

MARIÁN ZOUHAR[1]

Institute of Philosophy, Slovak Academy of Sciences
Klemensova 19, 811 09 Bratislava, Slovak Republic
marian.zouhar@gmail.com

ABSTRACT: Proper names are usually supposed to be rigid *de jure*. Given this claim and some other usual assumptions—namely that properties are explicated as intensions of a sort and that various possible worlds have various universes—one may derive the following inconsistent pair of conclusions: (*i*) for all properties P and for all possible worlds w it holds that an object, o, exemplifies P with respect to w only if o exists in w; and (*ii*) there is at least one property P and at least one possible world w such that o exemplifies P with respect to w even though o fails to exist in w. The aim of the present paper is to show how the problematic pair of conclusions is derived, spell out its background (most notably the idea of rigidity *de jure*) and review possible ways out.

KEYWORDS: Intension – linguistic convention – possible world – proper name – rigid designation – rigidity *de jure*.

*To Petr Koťátko,
a close friend, witty companion
and excellent philosopher*

1. Introduction

Possible worlds are a powerful theoretical tool widely employed in the current philosophy of language. They can be used to show, for example,

---

that certain kinds of singular terms are *rigid designators*; an expression is supposed to be rigid provided it designates the same object with respect to all possible worlds (in which the expression designates something or in which the object in question exists).[2] This holds, primarily, for all (or almost all) proper names and some definite descriptions.[3] Since the ways proper names and definite descriptions designate the targeted objects are widely different, one may expect they are rigid in different senses—definite descriptions are usually said to be rigid *de facto* while proper names are supposed to be rigid *de jure*.[4]

Now despite being well motivated, the possible world apparatus together with the idea of proper names as rigid *de jure* may be used to derive a pair of problematic conclusions. More specifically, if (*i*) proper names are rigid *de jure*, (*ii*) properties are exemplified by objects relative to possible worlds only, and (*iii*) different possible worlds assume different universes, i.e. sets of objects inhabiting the worlds, then it could be inferred both (*iv*) that for all properties P and for all possible worlds w it holds that an object, o, exemplifies P with respect to w only if o exists in w and (*v*) that there is at least one property P and at least one possible world w such that o exemplifies P with respect to w even though o fails to exist in w. Obviously,

---

[2]   It is often admitted that some general terms are rigid designators as well (cf., for example, Kripke 1980, Ch. 3; Putnam 1975). There are numerous papers and books dealing with general terms and their rigidity, some of the most interesting ones being Devitt (2005), Gómez-Torrente (2006), LaPorte (2004; 2013), Martí (2004), Soames (2002). However, it is by no means settled which kinds of general terms could be rigid and in which sense they could be said to be rigid. I tried to disentangle these problems in Zouhar (2009). As for now, however, I deal merely with rigid designation as applied to singular terms, leaving general terms aside.

[3]   The bracketed qualification concerning proper names is in place provided one is willing to make room for some special kinds of proper names. It could be claimed, for example, that the so-called descriptive names *à la* G. Evans' 'Julius' (cf. Evans 1982, 31) are non-rigid. (This is not to say, of course, that Evans himself would take them as non-rigid; in fact, descriptive names were rigid for him (cf. Evans 1982, 60-61).) Anyway, since rather marginal with respect to the present paper, I shall ignore complications of this kind throughout the paper. I shall ignore also complications brought about by the names of non-existent entities such as 'Vulcan' (*qua* a name of the purported intra-mercurial planet) as well as the names of fictional entities, whatever they are, such as 'Sherlock Holmes'.

[4]   This distinction has been introduced by S. Kripke in Kripke (1980, 21).

claims (*iv*) and (*v*), though both resulting from apparently unproblematic assumptions, are mutually incompatible.

I try to show how the widespread and more or less acceptable assumptions (*i*)–(*iii*) bring about the apparently unacceptable pair of conclusions (*iv*) and (*v*). In Sections 2–4, I set the stage by introducing the relevant ideas concerning possible worlds, proper names and rigid designation. Section 5 presents, in some detail, the two lines of reasoning leading to the problematic conclusions. Finally, possible ways out are outlined in Section 6.

2. A possible world framework

Possible worlds are often taken to be an effective tool for representing a wide variety of semantic as well as metaphysical features of natural languages. To select just some examples at random, they can be used to explain the truth-conditions of modal sentences such as 'It is necessary that *p*' or 'It is possible that *p*', where *p* is an indicative sentence of a given language; they can be used to explain the truth-conditions of counterfactual conditional sentences; they can be used to explain why certain sentences are merely contingently true while some other sentences are necessarily true. And most importantly for us, they are used to explain why certain singular expressions designate objects rigidly and some other ones designate them non-rigidly.

Possible worlds can be understood as maximal and consistent collections of states-of-affairs. There are various kinds of states-of-affairs, the most notable ones being those in which an object exemplifies a property or those in which a tuple of objects exemplifies an *n*-ary relation. The first kind of states-of-affairs can be described by the sentences of the form 'α is Φ', where α is a singular term and Φ is a predicate expressing a property, while the second kind of states-of-affairs can be described by the sentences of the form '$\beta_1, ..., \beta_n$ are Ψ', where $\beta_1, ..., \beta_n$ are singular terms and Ψ is a predicate expressing an *n*-ary relation.[5] A collection of states-of-affairs is max-

---

[5]   I take '$\beta_1, ..., \beta_n$ are Ψ' to be a form instantiated by sentences such as 'Bill and Bob are brothers' or 'Bill is taller than Bob' rather than by sentences such as 'Bill and Bob are bachelors' or 'Bill and Bob are tall' (the latter being, instead, sentences of the form 'β_1 is Φ and β_2 is Φ').

imal provided for every object and for every property it is specified whether the object exemplifies the property or not and for every tuple of objects and for every n-ary relation it is specified whether the tuple of objects exemplifies the relation in question or not. A collection of states-of-affairs is consistent provided it does not involve a state-of-affairs according to which an object exemplifies a property (or a tuple of objects exemplifies an n-ary relation) together with a state-of-affairs according to which the object in question does not exemplify the property in question (or the tuple in question does not exemplify the n-ary relation in question).

Within the possible world framework, properties and relations are nicely explicated as intensions, i.e. functions mapping possible worlds to extensions; properties are explicated as (total) functions mapping possible worlds to sets of objects while relations are explicated as (total) function mapping possible worlds to sets of tuples of objects.[6] As a result, objects are allowed to exemplify properties in possible worlds only; the same holds for tuples of objects exemplifying relations. In fact, exemplifying a property (or an n-ary relation) in a given possible world is the same as being a member of the set of objects (or the set of tuples of objects) that is assigned to the possible world by the property (the n-ary relation). It makes no sense to say that an object exemplifies a property, or that a tuple of objects exemplifies a relation, independently of any possible world whatsoever.

The set of all objects inhabiting a possible world is a universe. If the idea of a constant universe across all possible worlds is adopted, then an object exists in one possible world if, and only if, it exists in all possible worlds as well. Alternatively, universes can be construed as variable; for every possible world there is a universe such that different possible worlds assume different universes. In such a case, if an object exists in one possible world, it need not exist in all possible worlds. The philosophers who study natural languages in terms of possible world semantics often opt for variable universes (cf., for example, Smith 1987, 84). This enables them to admit that a speaker may truly assert about a particular object o that o does not exist in a certain possible world; it enables them also to admit that a speak-

---

[6]   If there is no object exemplifying a certain property in a given possible world or if there is no tuple of objects exemplifying a certain n-ary relation in a given world, then the respective functions assign the empty set to such a world.

er may assert of *o* that *o* exists in a certain possible world without saying that the assertion is utterly trivial.[7]

In what follows, it is unnecessary to assume a particular version of the possible world semantics. What we need is just the general view about properties as intensions and the assumption of different universes for different possible worlds. This gives us claims (*ii*) and (*iii*) mentioned in Section 1 as premises of the inferences leading to the problematic pair of conclusions (*iv*) and (*v*). It remains to flesh out the claim (*i*); this is done in Sections 3 and 4.

3. Rigid designation

To begin with, let us introduce the notion of rigid designation. There are several definitions of rigid designation on the market; some of them are equivalent in that they depict the same notion of rigidity while other definitions are non-equivalent because there are slight, though in effect substantial, differences between the respective notions defined.

The core idea common for all definitions of rigid designation can be captured in the following partially negative manner:

> An expression, ε, is a rigid designator if, and only if, (*i*) there is an object *o* and a possible world *w* such that ε designates *o* with respect to *w* and (*ii*) there are no objects *o* and *o*′ (where *o* ≠ *o*′) and no possible worlds *w* and *w*′ (where *w* ≠ *w*′) such that *a*) ε designates *o* with respect to *w*, and *b*) ε designates *o*′ with respect to *w*′.

The condition (*i*) guarantees that ε designates something, i.e. that it is a designator. The condition (*ii*) is essential for rigidity—if an expression fails to satisfy this condition (while satisfying the condition (*i*)) it is a nonrigid designator. This definition of rigid designation is neutral in that the

---

[7]   If the idea of the constant universe is adopted, one may say that an object does not exist only in the sense that there is no object in the universe exemplifying a certain property. As a result, when one wants to say truly that Bill does not exist in a particular world, one has to mean, at most, that the world in question is not inhabited by an object uniquely exemplifying a certain property. There is no way how a person might truly say of a particular object *o* she has in mind that this very object does not exist in a certain possible world (or in any other possible world, for that matter).

core idea given in the negative condition (*ii*) is consistent with all positive formulations one may dig out in the literature. There are at least two possibilities how to flesh this idea out:

> An expression, ε, is a rigid designator if, and only if, there is an object *o* such that (*i*) there is a possible world *w* with respect to which ε designates *o* and (*ii*) for any possible world *w'* it holds that if *o* exists in *w'*, then ε designates *o* with respect to *w'*.[8]

> An expression, ε, is a rigid designator if, and only if, there is an object *o* such that (*i*) there is a possible world *w* with respect to which ε designates *o* and (*ii*) for any possible world *w'* it holds that if ε designates anything with respect *w'*, then ε designates *o* with respect to *w'*.[9]

Though closely connected, the two definitions are non-equivalent and the notions of rigidity defined therein are different.[10] Since the former definition is usually employed in the literature, I will stick to it in what follows.[11]

There is one question that pops immediately into one's mind: What happens with respect to those worlds in which the object rigidly designated by the expression fails to exist? The most natural response seems to be that the expression designates nothing with respect to such worlds. For, being

---

[8]   This is my restatement of Kripke's claim that "a designator rigidly designates a certain object if it designates that object wherever the object exists" (Kripke 1980, 48-49); for an alternative formulation of the same idea see also some of the numerous works by S. Soames devoted to reference, e.g., Soames (2002, 4; 2006, 16).

[9]   This is my restatement of H. Putnam's claim that a designator is called "'rigid' (in a given sentence) if (in that sentence) it refers to the same individual in every possible world in which that designator designates" (Putnam 1975, 231). Obviously, the conditional "if ε designates anything with respect *w'*, then ε designates *o* with respect to *w'*" could be strengthened to the biconditional "ε designates anything with respect *w'* if, and only if, ε designates *o* with respect to *w'*". Anyway, I consider the simpler version in the main text.

[10]   Without going into details I just hint that some expressions are rigid according to the latter definition without being such according to the former definition. For more about the differences between the two notions of rigidity see Zouhar (2012).

[11]   The choice between the two definitions is inessential for our purposes because proper names remain rigid *de jure* whichever definition is considered. As a result, the line of reasoning leading to the pair of inconsistent conclusions outlined in Section 1 (and further elaborated in Section 5) is independent of a definition selected.

a rigid designator, the expression cannot designate anything else. And since the object in question does not exist in such worlds, there is nothing with which the expression could enter the designation relation. As a result, the expression should be expected to designate nothing with respect to such possible worlds.

This reply seems to be fully satisfactory in the case of definite descriptions. Obviously, a definite description 'the Φ' designates an object o with respect to a given possible world w only provided o satisfies, with respect to w, the descriptive condition expressed by 'the Φ'; o may satisfy the descriptive condition with respect to w only provided o *uniquely* exemplifies, in w, the property expressed by the predicate part 'Φ' of the description 'the Φ'. As a result, 'the Φ' is a rigid designator only provided the descriptive condition expressed by 'the Φ' is satisfied by the same individual, namely o, with respect to all possible worlds relative to which the descriptive condition is satisfied by anything. In other words, being a rigid designator, 'the Φ' uniquely describes o with respect to all possible worlds in which o exists and fails to describe uniquely anything with respect to those possible worlds in which o does not exist. And since nothing satisfies the descriptive condition expressed by 'the Φ' relative to those worlds, 'the Φ' does not designate anything with respect to them. Kripke introduced the term 'rigidity *de facto*' to name this kind of rigid designation.[12]

---

[12] Obviously, I assume that definite descriptions (as well as some other expressions) are allowed to designate, or to refer to, something. This seems to contradict the view according to which expressions cannot be properly said to refer to, or designate, anything; it is only speakers who can be said to refer to, or designate, something by uttering expressions. This view has been forcefully presented by those inspired by Strawson's (1950) criticism of Russell's theory of descriptions; cf., in particular, Linsky (1963), Searle (1969) as well as Koťátko (1993; 1995; 2009) and his numerous works written in Czech (including the 2006 book). Of course, I do not wish to claim that definite descriptions are capable of performing referential speech acts. Anyway, this fact need not prevent us from assuming another notion of reference (called "designation" in this paper) as depicting a relation between expressions and extra-linguistic entities; the relational notion of reference, i.e., designation, differs from the speech act notion of reference (or, better, referring). In the case of definite descriptions, an expression designates some object with respect to a possible world only if the object satisfies, in that world, the descriptive condition expressed by the description; this relation holds independently of all acts of referring anyone might carry out by uttering the description. See, e.g., Kallestrup (2012, 11ff.) for a recent employment of the relational notion of reference.

It is easy to see that a definite description may rigidly designate some-
thing with respect to all possible worlds without exception only provided
the object designated exists in all possible worlds.[13] Such an object would
be a necessary existing object. Typical examples abundant in the literature
involve descriptions designating mathematical entities such as 'the sum of 2
and 3' or 'the even prime number'.[14] However, this is debatable because
numbers and mathematical entities in general cannot be properly said to
exist in possible worlds; they are, rather, altogether independent of all
possible worlds. As a result, it is highly contentious to claim that 'the sum
of 2 and 3' designates 5 *with respect to* some possible world or other. One
should better say that mathematical descriptions designate entities *indepen-
dently* of possible worlds. Since the notion of rigid designation has been in-
troduced in terms of possible worlds, one should also better abandon the
idea that mathematical descriptions are rigid designators. In short, this
conceptual apparatus is not designed to capture expressions designating
entities which are not ordinary objects, i.e. members of the universe.[15]

So, putting these purported examples aside it is debatable whether there
really are definite descriptions that would rigidly designate something with
respect to all possible worlds. If objects are understood in their mundane

---

[13]   In Kripke's terms, such a designator is *strongly rigid*; cf. Kripke (1980, 48).

[14]   See, for example, Kripke (2011a, 9) where he discusses the definite description 'the
square root of 25' as an example.

[15]   Of course, one might attempt to introduce another notion of rigid designation
which could be used to show that mathematical descriptions are rigid designators after
all. We might assume that there is (in some sense of 'is') the "world of numbers" con-
sisting of "arithmetical states-of-affairs" and that mathematical descriptions designate
"inhabitants" of this "world". If this is the case, however, mathematical descriptions
would be rigid in a fairly trivial sense. The reason is that there is just one such "world",
i.e. there are no possible alternatives to this single "world of numbers". (The "world of
numbers" should not be confused with the various arithmetic systems—the systems are
just representations of the single "world".) Consequently, it makes no sense to mull over
what would happen had things been different with the "world of numbers". Obviously,
there is also no way how to effectively differentiate rigid mathematical descriptions from
non-rigid ones (there are none such expressions) and, thus, the very notion of rigidity
appears to be useless in this realm. Anyway, if someone wants to retain the notion of ri-
gid designation for mathematical expressions as well, she has to bear in mind that this
would be a *different* notion completely dissociated from the one defined in terms of
possible worlds. Be that as it may, in our present sense, mathematical descriptions are
not rigid designators; they are not, however, non-rigid designators either.

sense, we should, perhaps, respond in the negative. Thus, we might take it as a plausible hypothesis that for rigidly designating definite descriptions it holds that there are possible worlds with respect to which the descriptions designate the same individual as well as possible worlds with respect to which they fail to designate anything at all.

4. Proper names and rigidity

Now let us turn to proper names. They are usually supposed to be paradigmatic examples of rigidly designating expressions. Concerning designation, however, proper names are quite dissimilar from definite descriptions. The main semantic differences between proper names and definite descriptions stem from the simple fact that proper names designate objects in a completely different fashion than definite descriptions. While definite descriptions, if rigid, are rigid *de facto*, proper names are—to use Kripke's locution again—rigid *de jure* meaning that "the reference of a designator is *stipulated* to be a single object, whether we are speaking of the actual world or of a counterfactual situation" (Kripke 1980, 21).

Now, how it happens that proper names are rigid *de jure*? Since his remarks on rigidity *de jure* are very sketchy, Kripke offers no direct response to such question. Anyway, it is easy to devise one on the basis of his overall theory of proper names.

Proper names are usually introduced into the language by certain baptismal acts in which depicted objects are assigned linguistic items as their appellations.[16] When an object is assigned a linguistic item as its proper

---

[16]   To my knowledge, the most thorough considerations concerning baptisms can be found in M. Devitt's earlier book Devitt (1981). Though the term 'baptism' might suggest that introducing a name into language is a rather formal and ceremonious procedure, in many cases it is by no means so. Name introduction is often very informal; sometimes a name simply takes hold somehow in the community of speakers (cf. nicknames, for example). Anyway, we may imagine, as a highly idealized situation, that an act of baptism follows certain rules; retrospectively, we might simulate there being an act of baptism for virtually every proper name. Be that as it may, nothing of importance in this paper rests on how baptisms are modeled.

What is important is that proper names designate their bearers on a conventional basis, as claimed below in the main text. Even the picture of baptismal acts as giving raise to certain kinds of linguistic conventions might be, however, taken as highly idealized and simplified. I do not doubt it. There are long-running discussions concerning

name, a new linguistic convention is being introduced into the language. So, baptismal acts can be viewed as forging linguistic conventions associating linguistic items with extra-linguistic entities. In short, baptismal acts establish name-bearer relations; the relation between proper name and its bearer is thus merely conventional. As a result, whenever one utters a name (intending to refer to whatever is its semantic reference)[17] she refers to the particular object that is assigned to the name as its bearer on the basis of a particular linguistic convention established during the act of baptism. The difference between the referential behavior of proper names and those of definite descriptions is immense: while definite descriptions designate objects because the objects satisfy certain descriptive conditions, proper names designate their bearers because of the linguistic conventions introduced during the baptismal acts. As a result, an object needs not satisfy any descriptive condition to be designated by a given proper name. This is an important feature because there is, strictly speaking, virtually no property such that the object designated by a proper name has to exemplify in order to be designated by the name in question. So, the name designates the object irrespective of virtually any property the object exemplified or might have exemplified. Obviously, this is just another way of saying that the name designates the object with respect to all possible worlds in which that object exists, i.e. that the name is a rigid designator.[18]

---

the nature of linguistic conventions, but I do not wish to take here a definitive stand on this question. For our purposes it suffices to admit that, in the case of proper names, linguistic conventions—whatever they are and however they are constituted—can be taken as sanctioning the link between the names and their bearers. (The nature of linguistic conventions and their role in communication and meaning determination is one of the pivotal topics of Petr Koťátko. He authored numerous papers devoted to this topic. Moreover, they play an important role in his two philosophical books written in Czech; cf. Koťátko 1998, 2006.)

[17]   On the notion of semantic reference see Kripke (2011b).

[18]   The considerations from this paragraph suggest that there is a close connection between being a *de jure* rigid designator and being a directly referring term (in one particular sense of direct reference). For details see Zouhar (2011); for some related considerations see also Pendlebury (1990). According to the relevant notion of direct reference, there is no conceptual mediation between a name and its bearer; to my knowledge, this notion has been introduced by R. Barcan Marcus (cf. Marcus 1993a, 11; 1993b, 203). Yet another notion of direct reference is propounded by D. Kaplan (cf. Kaplan 1989a; 1989b): a directly referring term is one which supplies just its referent to the propositions expressed by sentences featuring the term. Kaplan's notion of direct

However, we should take one step further. Proper names should be allowed to designate their bearers even with respect to those possible worlds in which the objects designated fail to exist.[19] Why? Well, because proper names designate their bearers conventionally. The conventionally established link between a name and an object has to persist with respect to all possible worlds that can be described in the language involving the name. And, of course, the worlds in which the name's bearer fails to exist do belong to those worlds which can be described in the language.

To use N. Salmon's terminology, the suggestion implies that proper names are best conceived of as being *obstinately* rigid designators as opposed to being merely *persistently* rigid ones. An obstinately rigid designator designates "the same thing with respect to every possible world, whether that thing exists there or not" (Salmon 1981, 34); on the other hand, a persistently rigid designator "designates the same thing with respect to every possible world in which that thing exists, and which designates nothing with respect to possible worlds in which that thing does not exist" (Salmon 1981, 33-34). Obviously, definite descriptions are persistently rigid, while proper names are obstinately rigid. So, proper names should be both rigid *de jure* as well as obstinately rigid. In fact, obstinate rigidity seems to be just the other side of rigidity *de jure* since both features can be explained in terms of the conventionally determined link between proper names and their bearers.[20]

---

reference concerns the semantic content of (the sentences involving) certain kind of expression; in the present paper, however, the semantic content is not at issue. For further details see, for example, Martí (1995; 2003) and Kallestrup (2012, 35-37).

[19] This idea has been ingeniously defended by D. Kaplan several decades ago. He claimed: "Some have claimed that though a proper name might [designate] the same individual with respect to any possible world ... in which he exists, it certainly cannot [designate] him with respect to a possible world in which he does not exist. With respect to such a world there must be a gap in the name's designation, it designates nothing. This is a mistake. There are worlds in which Quine does not exist. It does not follow that there are worlds with respect to which 'Quine' does not [designate]. What follows is that with respect to such a world 'Quine' [designates] something which does not exist in that world" (Kaplan 1973, 503). Cf. also Kaplan (1989a, 492-493; 1989b, 569).

[20] The claim that proper names are obstinately rigid is adopted by, e.g., Almog (1986), Branquinho (2003), Salmon (1981), Smith (1984; 1987) (as well as Kaplan and many others). There are also some dissenting voices according to which proper names are better viewed as persistently rigid; cf., most notably, Murday (2013). Murday's primary tar-

To put the same idea differently, let us try another course. Possible worlds can be used to explain certain features of languages. To simplify things to a considerable degree, the language is defined once its vocabulary involving simple expressions associated with linguistic conventions is provided and the grammatical rules used to generate compound expressions are given. It should be, thus, plain that linguistic conventions enter the picture at the language forming stage. Now, once we have the language at our disposal, we may describe its features using, *inter alia*, the possible world apparatus. What is important, however, is that the language should be there for our disposal first and foremost. And since linguistic conventions for proper names fix the name-bearer associations, it should be assumed that proper names are assigned their bearers *prior* the apparatus of possible worlds is invoked for whatever reasons.[21] In other words, the name-bearer relation is not dependent on any possible world whatsoever. As a result, the relation has to hold for all possible worlds describable in the language at hand without any exception; it has to hold with respect to all possible worlds including those in which the bearer of the name does not exist.

Refusing this idea amounts to saying that the linguistic convention associating the name with its bearer holds no more with respect to those possible worlds in which the name's bearer does not exist. However, this is unacceptable because it means that the language comprising the linguistic convention in question has been revised somehow—the linguistic convention was removed from it. It means, strictly speaking, that we have another language in place of the original one.[22] However, since the possible world apparatus was intended to be used in explaining certain features of the original language, the new language is utterly irrelevant. So, if we want to stick with the original language we cannot but accept the idea that the name is

---

get is the usual requirement that proper names must be obstinately rigid if they are to be directly referring (in Kaplan's sense; cf. footnote 17). He argues, however, that this is not the case and that the idea of direct reference is better combined with persistent rigidity.

[21]   As claimed by Smith, "[w]ith names, designata are specified in the base clauses of our semantic theory antecedently to the running of the possible-worlds machinery" (Smith 1987, 87).

[22]   For the sake of simplicity, I assume here a synchronic view of language. However, nothing important rests on this assumption. We could easily switch to the diachronic approach to language; what would be required are just some minor reformulations.

an obstinately rigid designator, i.e. that it designates its bearer even with respect to those possible worlds in which the bearer fails to exist.

Summing up, proper names designate objects on the basis of linguistic conventions that are established independently of possible worlds. Therefore, whatever possible world is described by the language, the name at hand designates its conventionally assigned bearer with respect to it. The name is, thus, a rigid designator. Because of its conventional nature, the link between the name and its bearer warrants that names are rigid *de jure* rather than *de facto*. And as far as I can see, the fact that a proper name designates its bearer even with respect to those possible worlds in which its bearer does not exist is a simple consequence of the conventional nature of nominal designation.

5. The puzzle

Now we are ready to jump on the puzzle advertised in Section 1. Given the previous considerations, we may provide two lines of reasoning, which are both acceptable and consistent with what we have just said, but lead to incompatible conclusions. Let x be a variable ranging over objects, P be a variable ranging over properties (of objects), w be a variable ranging over possible worlds, e be a variable ranging over expressions and c be a variable ranging over linguistic conventions. Let us further assume that o is an object and α is a proper name such that there is a linguistic convention associating α with o; given this linguistic convention, α is a name of o and o is the bearer of α.[23]

Given these assumptions, the first line of reasoning may be summarized in the following way:

---

[23]   The indefinite article in "α is a name of o" suggests that o may have more than one proper name while the definite article in "o is the bearer of α" implies that α has just one bearer. The latter assumption might be taken as a simplification, though an innocuous one. We might admit that α has more than one bearer, but still it would hold that, *relative to a particular linguistic convention*, α may have only one bearer, namely o. In other words, there have to be as many linguistic conventions associated with a given name as there are bearers of the name; it cannot happen that a name designates various objects relative to *the same* linguistic convention. If this were the case, the name should have assigned various objects during one and the same baptismal act.

1. For any object x and any property P it holds that x exemplifies P only in the sense that there is a possible world w such that x exemplifies P in w.
2. Thus, for any property P it holds that o exemplifies P only in the sense that there is a possible world w such that o exemplifies P in w.
3. For any object x, any property P and any possible world w it holds that x exemplifies P in w only provided x exists in w.
4. Thus, for any property P any possible world w it holds that o exemplifies P in w only provided o exists in w.
5. There are possible worlds w and w' such that o exists in w and fails to exist in w'.
6. So, if there is a possible world w' such that o does not exist in w', then no property P is such that o exemplifies P in w'.

Claim 1 is based on our common understanding of possible worlds as collections of states-of-affairs, where a state-of-affair might consist of an object exemplifying a property or of a tuple of objects exemplifying an n-ary relation. It respects the fact that properties are explicated as intensions, i.e., functions from possible worlds to extensions (sets of objects or of tuples of objects). Claim 2 is a particular instantiation of what is involved in claim 1. Claim 3 is, again, an unproblematic assumption that is based on the understanding of the exemplification relation outlined in Section 2. Similarly, claim 4 is a particular instantiation on claim 3. Claim 5 is an assumption that is based on the idea of variable universes for different possible worlds. Conclusion 6 follows from the above claims. It suggests that there is no possible world with respect to which it would hold both that o does not exist in that world and that o exemplifies some property or other relative to that world. As far as I can see, this argument is rather unproblematic and its conclusion is justified by its premises.

The same can be said about the second argument. It can be put forth in the following way:

1. For any expression e it holds that if e is a proper name then e is a *de jure* rigid designator and there is an object x and a linguistic convention c such that e designates x on the basis of c.
2. For any linguistic convention c and any possible world w it holds that c is in force regardless of how things are in w.
3. Thus, if there is a linguistic convention c such that α rigidly designates o on the basis of c, then for any possible world w it holds that

α rigidly designates *o* with respect to *w* regardless of how things are in *w*, i.e., *inter alia*, regardless of *o*'s existence or non-existence in *w*.

4. There are possible worlds *w* and *w′* such that *o* exists in *w* and fails to exist in *w′*.

5. Thus, if there is a linguistic convention *c* such that α rigidly designates *o* on the basis of *c*, then for any possible world *w′* such that *o* fails to exist in *w′* it holds that α rigidly designates *o* with respect to *w′*.

6. If there is a linguistic convention *c* such that α rigidly designates *o* on the basis of *c*, then α exemplifies the property of *naming o* and *o* exemplifies the property of *being named by* α.

7. Thus, if there is a possible world *w′* such that *o* fails to exist in *w′*, *o* still exemplifies the property of *being named by* α in *w′*.

8. So, if there is a possible world *w′* such that *o* does not exist in *w′*, then there is at least one property *P* such that *o* exemplifies *P* in *w′*.

Claims 1 and 2 are assumptions based on what we have said in Section 4. Everyone who believes in rigidity *de jure* should accept them without much ado. Claim 3 is a particular instantiation derived from the above claims. Claim 4 is an assumption based on the idea of variable universes for different possible worlds; it serves as an introduction of a possible world in which a particular object does not exist. Claim 5 presents just a restatement of the consequent of claim 3 applied to those possible worlds in which a particular object does not exist. Claim 6 is an assumption but, again, an unproblematic one. The reason is that if we assume that there is a relation between a name and an object, both the name and the object have to exemplify the properties of being in the relation in question with the other entity. Claim 7 is derived from claims 5 and 6 and applies to those possible worlds in which a particular object does not exist. Finally, conclusion 8 is a mere generalization of the previous claim. It suggests that there is a possible world in which it holds both that *o* does not exist in that world and that *o* exemplifies some property relative to that world.

6. Possible ways out

The set of assumptions used in the arguments from the previous section, though innocuous at first sight, lead to the mutually incompatible conclu-

sions. This fact implies that at least one of the assumptions should be given up. All in all, there are three important claims occurring in the arguments that can be blamed for the derivations of the problematic conclusions:

1. It is impossible for an object to exemplify any property independently of any possible world (because properties are explicated as intensions, i.e. functions defined on possible worlds).
2. It is possible for an object to exist in some possible worlds without existing in all possible worlds (because different possible worlds are allowed to have different universes).
3. Proper names are *de jure* rigid designators (because they designate their bearers merely on the basis of linguistic conventions established during baptismal acts).

The puzzle could be blocked when any of the assumptions 1–3 is rejected. Assumption 1 is a crucial thesis backing the first line of reasoning, so denying assumption 1 amounts to refusing the first part of the argument. Assumption 3 is a crucial premise of the second line of reasoning, so denying assumption 3 amounts to refusing the second part of the argument. Assumption 2 is important in the derivations of both conclusions because it permits to take into account worlds in which a particular object does not exist. Consequently, if we refuse assumption 1 and/or assumption 2, the first part of the argument would be blocked, and if we refuse assumption 3 and/or assumption 2, the second part of the argument is blocked.

In what follows, I discuss some options that are available when the above assumptions are item-by-item denied. Denying each of the assumptions opens up various routes one may take but, needless to say, I cannot discuss all of them here.

To begin with, let us consider the possibility of withdrawing assumption 2. This would mean that there was one and the same universe for all possible worlds and that if an object existed in one possible world, it would exist in all worlds indiscriminately. As a result, there could not be an object which did not exist in certain possible worlds and, yet, exemplified some property or other in such worlds. There is, however, a price to be paid.

Firstly, the sentences of the form 'α exists' (where α is a proper name of an object) would be necessarily true, if true at all. If α has been successfully introduced into the language as a proper name of something and, thus, a particular object has been named by α, the object designated must be a necessarily existing entity. In such a case, 'α exists' is a necessarily true

sentence. Yet, we hardly take natural language sentences of the form 'α exists' to be true of necessity (if true at all). This conclusion could be obviated if it were denied that proper names designate objects on the basis of linguistic conventions. If proper names were associated, instead, with some kind of (descriptive) condition that is to be satisfied by objects designated, α would designate something only if the object satisfied the condition in question. In such a case, the object, though necessarily existing, needs not be designated by the name with respect to certain possible worlds and the corresponding sentence of the form 'α exists' would be false relative to such worlds (provided nothing else satisfied the condition in question). This suggestion would, however, undermine also the idea that proper names are rigid *de jure*.

Secondly, it seems that the sentences of the form 'α exists' would be, if false, necessarily false. This is, again, rather unintuitive with respect to natural language sentences of the form 'α exists'. There is, however, something even more puzzling. It seems that a sentence of the form 'α exists' (where α is a proper name and, thus, designates something on the basis of a linguistic convention) could be false only provided something went wrong during the act of introducing α into the language. For such a sentence could be false only if no object has been designated by α. Obviously, this might happen only if no object has been assigned to α during the baptismal act which means that α was not introduced as a full-blooded proper name. As a result, whenever one would come across a false sentence of the form 'α exists', she would learn something about the stock of names—or purported names—we have in our language instead of something about the extra-linguistic world itself. This would be rather far-fetched. Anyway, these seem to be unpleasant consequences to be met by everyone who would like to eliminate the puzzle by discarding assumption 2.

Another route to eliminate the puzzle is denying assumption 3 according to which proper names are *de jure* rigid designators. If proper names were not rigid *de jure*, there would be no reason to admit they designate something also with respect to those possible worlds in which their bearers do not exist.[24] So, supposing that α is a proper name of an object *o*, α would designate *o* with respect to those possible worlds in which *o* exists, though, with respect to the remaining worlds, it would designate nothing at all. There

---

[24] Obviously, this option amounts also to denying that proper names are obstinately rigid. An ingenious argumentation to this effect can be found in Murday (2013).

would be, thus, no possible world such that *o* does not exist in it and, still, *o* does exemplify some property relative to it. So, no puzzle would arise.[25] There are, however, certain problematic consequences of this option.

If one wants to deny that proper names are rigid *de jure*, one has to deny also that proper names designate their bearers on the basis of certain linguistic conventions introduced during baptismal acts. Since the name-bearer relation is no more conventionally warranted, proper names should designate their bearers on a different basis. It might be suggested, for example, that an object has to satisfy some condition or other in order to be designated by a proper name.[26] This would amount to admitting some kind of descriptivism concerning proper names.[27] This is a complicated topic and I have no space to pursue it further in this paper. I should add, nevertheless, that adumbrating descriptivism need not be problematic in itself; descriptivism might be problematic only provided it could not offer satisfactory responses to the arguments devised against it.

What is worse, however, is that denying assumption 3 goes against an established empirical fact. It seems to be an obvious empirical fact that objects receive their names mainly on the basis of conventionally driven decisions undertaken during baptismal acts (or some other acts more or less resembling baptisms). The link between a name and its bearer is, therefore, best supposed to be conventional. If an object were determined to be a bearer of a name not on the basis of a linguistic convention but on the basis of, let us say, satisfying certain conditions (descriptive or other), then there would be no point in saying that the object has been assigned to the name during an act of baptism. It means, on the other hand, that if bapt-

---

[25]   Recently, P. Baumann attacked the view that proper names are *de jure* rigid designators in Baumann (2010). His argument is based on (*i*) denying that proper names *qua* types can be said to designate anything and on (*ii*) J. Katz's ideas concerning multiple bearerhood that is typical for ordinary proper names (cf. Katz 2001). Without going into details I just point out that both points can be contested such that claim 3 from the main text remains untouched.

[26]   As a result, proper names, if rigid, could be, at most, rigid *de facto*.

[27]   Obviously, a special kind of descriptivism would be required according to which the descriptive condition associated somehow with a proper name determined which object is designated by the name. This holds for Fregean versions of descriptivism ('Fregean' being used here in a very broad sense). On the other hand, non-Fregean versions—such as the one developed by J. Katz in a number of works; see, e.g., Katz (1992; 1994)—are not suitable in this connection.

isms are to be decisive for relating names and their bearers, satisfying con-
ditions (descriptive or other) must be irrelevant to this purpose.[28] So, if
one wants to discard baptisms as sources of conventionally established
name-bearer relations, one has to explain somehow away the above empiri-
cal fact.

Anyway, we cannot retain both the idea of conventionally established
name-bearer relations and the idea of name-bearer relations being deter-
mined such that the bearer of a name satisfied some kind of condition. So,
when one decides to drop assumption 3, one has to cope somehow with
the above consequences and provide an alternative (non-conventional) ex-
planation of the link between proper names and their bearers.

The final option consists in refusing assumption 1. In such a case the pa-
radox would not arise because properties would not be explicated as certain
intensions defined on possible worlds and, thus, objects could, if properly ex-
plicated, instantiate properties independently of possible worlds. As a result,
it could be feasible for an object to exemplify a certain property also with re-
spect to such a possible world in which the object failed to exist (provided, of
course, the new construal of properties admits such a possibility).

Now it seems that this effect could be achieved even though assumption
1 is not refused in its entirety; it merely suffices if it is restricted to a cer-
tain degree. We might distinguish ordinary properties such as *being
a mammal* or *being red* from properties such as *being a bearer of α*, where
α is a proper name. The former properties can still be explicated as ordi-
nary intensions; it is a necessary condition for an object to exemplify them
in some possible world that the object existed in the world in question. As
a result, when we confine the term 'property' for these kinds of attributes,
assumption 1 can be retained in a restricted form. So, the final option
might consist in refusing to take the attributes like *being a bearer of α* as

---

[28]   The incompatibility of the idea of conventionally established name-bearer relations
with the idea of satisfactionally guaranteed name-bearer relations can be summarized al-
so in the following way: If, to be a bearer of a name, an object has to satisfy some kind
of condition, the name-bearer relation would hold only with respect to those worlds in
which the object does satisfy the condition; on the other hand, since linguistic conven-
tions are supposed to hold for all possible worlds that can be described in the language
at hand, a proper name would designate its bearer with respect to all worlds and irres-
pective of any property its bearer exemplifies in those worlds. So, there is an insur-
mountable conflict between the two ideas.

properties explicated in terms of possible world intensions of a sort. This would suffice to deal with the puzzle.

The main challenge, however, would be to determine the dividing line between the properties capable of being explicated in terms of intensions and the other attributes. It might, perhaps, suffice to say that if an attribute is such that an object has it on the basis of how things are with our language and linguistic conventions that are in force in our language, then it cannot be explicated as an intension while all other attributes can be so explicated. This suggestion seems to be quite natural because the attributes an object has in virtue of linguistic conventions are somewhat special. As I have already pointed out, linguistic conventions are introduced regardless of how things are in some particular possible world or other. The language with all of its linguistic conventions is a device used to describe the actual world as well as all the worlds that are possible with respect to the actual one. At the same time, the language is not supposed to be an object inhabiting those worlds. So, all relations between linguistic items as well as all relations the linguistic items bear to anything else are supposed to be independent of possible worlds. Consequently, all the attributes anything has on the basis of the above relations should be also taken as independent of possible worlds. It means that the attributes like *being a bearer of α* or *being named by α* should be exemplified by objects independently of possible worlds, as required.[29]

To sum up, the last option seems to be the least demanding one because it permits to preserve all the above assumptions 1 – 3 almost untouched. What is required is just a suitable restriction of assumption 1. Of course, the other ways eliminating the puzzle could be also viable, though they would call for more radical changes than the last one and would require more ingenious arguments than those provided in this paper on behalf of the final option.

---

[29]   This suggestion, though somewhat unorthodox, can be extended to other cases as well. For example, mathematical entities can hardly be said to exemplify mathematical properties relative to possible worlds. Number 2 exemplifies the property of *being the even prime number* or the property of *being an even number* regardless of any possible world whatsoever. The reason does not consist in that number 2 is even or is the even prime number with respect to all possible worlds but in that numbers (or mathematical entities in general) do not belong to the universe of the actual world (or any other possible world). So, the above properties cannot be explicated as ordinary intensions defined on possible worlds; they must be attributes in some other, non-intensional, sense.

References

ALMOG, J. (1986): Naming without Necessity. *Journal of Philosophy* 83, No. 4, 210-242.

BAUMANN, P. (2010): Are Proper Names Rigid Designators? *Axiomathes* 20, 333-346.

BRANQUINHO, J. (2003): In Defense of Obstinacy. *Philosophical Perspectives* 17, No. 1, 1-23.

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

DEVITT, M. (2005): Rigid Application. *Philosophical Studies* 125, 136-165.

EVANS, G. (1982): *Varieties of Reference*. Oxford: Oxford University Press.

GÓMEZ-TORRENTE, M. (2006): Rigidity and Essentiality. *Mind* 115, 227-259.

KALLESTRUP, J. (2012): *Semantic Externalism*. London – New York: Routledge.

KAPLAN, D. (1973): Bob and Carol and Ted and Alice. In: Hintikka, J. – Moravcsik, J. – Suppes, P. (eds.): *Approaches to Natural Language*. Dordrecht – Boston: Reidel, 490-518.

KAPLAN, D. (1989a): Demonstratives. In: Almog, J. – Perry, J. – Wettstein, H. (eds.): *Themes from Kaplan*. Oxford: Oxford University Press, 482-563.

KAPLAN, D. (1989b): Afterthoughts. In: Almog, J. – Perry, J. – Wettstein, H. (eds.): *Themes from Kaplan*. Oxford: Oxford University Press, 565-614.

KATZ, J.J. (1992): The New Intensionalism. *Mind* 101, No. 404, 689-719.

KATZ, J.J. (1994): Names without Bearers. *The Philosophical Review* 103, No. 1, 1-39.

KATZ, J.J. (2001): The End of Millianism: Multiple Bearers, Improper Names, and Compositional Meaning. *The Journal of Philosophy* 98, No. 3, 137-166.

KOŤÁTKO, P. (1993): Frege, Tichý, and the U.S. President. *From the Logical Point of View* 2, 28-42.

KOŤÁTKO, P. (1995): Definite Descriptions: The Nicest Vehicle of Reference I Know. In: Hill, J. – Koťátko, P. (eds.): *Karlovy Vary Studies in Reference and Meaning*, 184-214.

KOŤÁTKO, P. (1998): *Význam a komunikace* [*Meaning and Communication*]. Prague: Filosofia.

KOŤÁTKO, P. (2006): *Interpretace a subjektivita* [*Interpretation and Subjectivity*]. Prague: Filosofia.

KOŤÁTKO, P. (2009): Definite Descriptions Again: Singular Reference, Quantification and Truth-Evaluation. *Organon F* 16, No. 4, 552-568.

KRIPKE, S. (1980): *Naming and Necessity*. Cambridge (Mass.): Harvard University Press.

KRIPKE, S. (2011a): Identity and Necessity. In: Kripke, S.: *Philosophical Troubles. Collected Papers*. Vol. 1. Oxford: Oxford University Press, 1-26.

KRIPKE, S. (2011b): Speaker's Reference and Semantic Reference. In: Kripke, S.: *Philosophical Troubles. Collected Papers*. Vol. 1. Oxford: Oxford University Press, 99-124.

LAPORTE, J. (2004): *Natural Kinds and Conceptual Change*. Cambridge: Cambridge University Press.

LAPORTE, J. (2013): *Rigid Designation and Theoretical Identities*. Oxford: Oxford University Press.

LINSKY, L. (1963): Reference and Referents. In: Caton, C. (ed.): *Philosophy and Ordinary Language*. Urbana University of Illinois Press, 74-89.

Marcus, R.B. (1993a): Modalities and Intensional Languages. In: Marcus, R.B.: *Modalities*. Oxford: Oxford University Press, 3-23.

Marcus, R.B. (1993b): Possibilia and Possible Worlds. In: Marcus, R.B.: *Modalities*. Oxford: Oxford University Press, 189-213.

Martí, G. (1993): The Essence of Genuine Reference. *Journal of Philosophical Logic* 24, No. 3, 275-289.

Martí, G. (2004): Rigidity and General Terms. *Proceedings of the Aristotelian Society* 104, 129-146.

Martí, G. (2005): The Question of Rigidity in New Theories of Reference. *Noûs* 37, No. 1, 161-179.

Murday, B. (2013): Names and Obstinate Rigidity. *The Southern Journal of Philosophy* 51, No. 2, 224-242.

Pendlebury, M. (1990): Why Proper Names Are Rigid Designators. *Philosophy and Phenomenological Research* 50, 519-536.

Putnam, H. (1975): The Meaning of 'Meaning'. In: Putnam, H.: *Mind, Language, and Reality. Philosophical Papers*. Vol. 2. Cambridge: Cambridge University Press, 215-271.

Salmon, N. (1981): *Reference and Essence*. Princeton: Princeton University Press.

Searle, J.R. (1969): *Speech Acts*. Cambridge: Cambridge University Press.

Smith, A.D. (1984): Rigidity and Scope. *Mind* 93, No. 370, 177-193.

Smith, A.D. (1987): Semantic Considerations on Rigid Designation. *Mind* 96, No. 381, 83-92.

Soames, S. (2002): *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford: Oxford University Press.

Soames, S. (2006): *Reference and Description: The Case against Two-Dimensionalism*. Princeton – Oxford: Princeton University Press.

Strawson, P. F. (1950): On Referring. *Mind* 59, No. 235, 320-344.

Zouhar, M. (2009): On the Notion of Rigidity for General Terms. *Grazer philosophische Studien* 78, 207-229.

Zouhar, M. (2011): Are There Directly Referring Non-Rigid Designators? *Prolegomena* 10, No. 1, 87-100.

Zouhar, M. (2012): On a Broader Notion of Rigidity. *Kriterion* 26, 11-21.

Can Texts Be Read? The Anatomy of a Paradox

ANDERS PETTERSSON

Department of Culture and Media Studies. Umeå University
901 87 Umeå. Sweden
anders.pettersson@littvet.umu.se

ABSTRACT: According to standard philosophical analyses of the ontology of texts, texts are abstract objects. However, that analysis has the paradoxical consequence that texts cannot be read: one cannot read something that lacks material existence. The essay resolves the paradox by introducing a distinction between the ordinary conception of a text—the supposed object of the philosophical analysis of the concept of a text—and the philosophical conception of a text, the end-product of that analysis. It is demonstrated by means of examples that a text as ordinarily conceived is at once physical and immaterial (and so can certainly be read, because of its physical side), and at once one and many, while the philosopher, uninterested in the actual, obviously illogical character of the ordinary conception of a text, turns the text into one abstract object by fiat—an object which cannot, of course, be read. It is argued that the ordinary concept of a text is handy in practical contexts precisely because it ignores troublesome distinctions, while the philosophical concept of a text serves no genuine purpose, since the theoretical understanding of verbal communication requires a partly different intellectual framework.

KEYWORDS: Literary aesthetics – ontology of art – text.

According to standard contemporary thinking in philosophy, texts cannot be *read*, at least not in a literal sense of the word.[1] Texts are thought to be abstract entities, and one cannot, of course, read something that lacks material existence.

---

[1] By "a text" I refer to any linguistic composition in communicational use: an oral remark, an email message, a novel, an encyclopaedia in twenty-four volumes, and so on. Texts can thus be short or long, literary or non-literary, written or oral. My essay could just as well have been called "Can Speeches Be Listened to?".

The counter-intuitive consequences of the idea that texts are abstract entities—and of the same idea concerning musical works, whose ontology is often discussed in parallel with texts—were pointed out by Richard Rudner in an article from 1950 which still haunts the ontology of art. As Rudner put it, if artworks are abstract entities, then they are "non-experienceable" (1950, 386). Nor, he observed, can they be created: "a counterintuitive consequence of the position is that Beethoven did not compose [his] *Fifth Symphony*" (ibid., 384). In the philosophical debate about the ontology of art, the question of the possibility of creating abstract objects has drawn more attention than the question of the possibility of experiencing them.[2] In this essay, however, I will concentrate wholly on the paradox, or seeming paradox, that texts cannot be read if we accept standard contemporary philosophical thinking. It is perhaps worth emphasizing that, for me, that paradox has nothing to do with aesthetics specifically but concerns texts in general.

It may appear evident that texts can, quite literally, be read—are you not in fact, quite literally, reading a text right now?—and equally evident that one cannot read something that has no physical existence, so the additional premise that texts are abstract entities does create a conundrum. Nor is the paradox an isolated phenomenon. Many human artefacts confront the philosopher with basically the same problem: it appears natural to understand them as abstract objects, but if one does, counterintuitive consequences make themselves felt. Take the Czech flag as a simple example. No physical object or aggregate of objects can very well be identified with the Czech flag, nor any mental state or occurrence, so the Czech flag will perhaps have to be considered to be an abstract object. But that solution, too, hardly complies with common sense: for example, it will rob the Czech flag of all colours. If the flag is an abstract entity it may no doubt have colours in a manner of speaking, but the flag itself cannot actually possess any colours. (Nor, it seems, can the Czech flag ever have been created.[3]) The pa-

---

[2]   That is well illustrated by the recent collection of articles *Art and Abstract Objects*, Mag Uidhir (2013a). See esp. the editor's introduction, Mag Uidhir (2013b).

[3]   Many will be quick to point out that the Czech flag, just like a text, is a *type*, with concrete Czech flags (in the case of the text: copies or exemplars) as its *tokens*. That is true, but not relevant in the present context. Identifying flags or texts with types only produces a classification, a statement about what kind of concept the concepts of a flag and of a text are. The paradoxes remain: you still cannot read a text itself, and the Czech flag itself still has no colours.

radox that texts cannot literally be read (or, for that matter, listened to)[4] thus forms part of a more extensive group of paradoxes around so-called "abstract artefacts".[5] I believe that what I have to say about my chosen paradox has rather obvious bearing on the whole issue of the ontology of abstract artefacts, but I will make no attempt to draw out such consequences here.

The philosopher can attempt to neutralize the counterintuitive consequence that texts cannot be read by pointing out that texts can clearly be read in a manner of speaking, since there will be physical manifestations related to a given text—"copies" or "exemplars", or entities with a similar function—and it will be possible to read those physical manifestations. That is the spirit in which Stephen Davies answers Rudner's challenge. Although a symphony is an abstract object, he says, a symphony can be noisy and triumphant at its close: its being noisy and triumphant at its close means that its "well formed instances" must be noisy and triumphant at its close.[6] Adopted to my paradox, that attitude would amount to saying that one will be able to read copies of a text but never the text itself, just as one will be able to listen to well formed instances of a musical piece but never to the musical piece itself.

To me, that answer does not make the paradox, or the impression of a paradox, go away. My own view is very different: the standard philosophical idea about what a text does indeed create a paradox. The paradox in itself is trivial and easily resolved, but the forces behind its emergence are important and problematic. My essay will concern the paradox, its solution, and the mechanisms responsible for the coming into being of the paradox. To some extent, I will even touch on certain substantive questions regarding linguistic communication that also belong in the picture.

The manifest paradox goes away as soon as one realizes that it relies on an equivocation. There are two concepts of a text in play: on one hand the ordinary, pre-theoretical concept, on the other, the concept that comes out of its philosophical analysis. As we will see, if one adopts the standpoint of

---

[4]   See note 1.

[5]   Amie Thomasson's term; cf. Davies (2007, 116).

[6]   Davies (2003, 169-170): "holding that Beethoven's Fifth Symphony is abstract does not entail that the work cannot be noisy and triumphant at its close. Instead, it requires that these descriptions of the work are true if those properties are prescribed for its well formed instances."

the ordinary conception of a text the paradox disappears: texts can be read, and it is simply not true that texts are abstract entities. If, on the other hand, one truly accepts the philosophers' analysis the paradox also vanishes: it is simply true that texts are abstract entities, and they certainly cannot be read, at least not literally. The impression of a paradox arises when one fails to keep the two perspectives apart.

If that view of the paradox is correct, and I hope to demonstrate beyond reasonable doubt that it is, the paradox as such has been successfully dissolved. Yet one is bound to ask oneself what standard contemporary philosophical thinking has done to the ordinary concept of a text, and whether the philosophical operations were for the good or for the bad. That would be interesting to know, for the idea of a text is a vital part of our general workaday understanding of the functioning of linguistic communication.

I will begin by examining the ordinary concept of a text and then go on to a description and an evaluation of the philosophers' analytical reconstruction of the concept. I will in fact be voicing some criticism both of the ordinary concept of a text and of the standard philosophical reinterpretation, and will therefore also point to a third way of thinking about texts, different from both.[7]

1. The ordinary conception of a text

The ordinary conception portrays a text as an individual unified object. A text has a physical side, but the text also exhibits a complex of signs (a "text" in another sense of the word) and a meaning. Very often, texts exist in many exemplars.

That description of a text may sound self-evidently true, but a text, such as the ordinary conception depicts it, is in fact clearly a contradiction in terms.[8] The text is a physical object (for one can read it or listen to it)

---

[7]   I have written several times on the ontology of texts, from an early work in Swedish wholly dedicated to the problem, Pettersson (1981), to the chapter "Conceptions of the Text" in my latest book, Pettersson (2012, 145-162). In the present essay I do not depart from the stance taken in my latest book, but I look at the whole issue from a very different angle.

[8]   That fact was demonstrated long ago by Michael Reddy in his well-known study of a basic system of metaphors underlying our thinking about linguistic communication; see Reddy (1979).

containing non-physical elements (for there are words and meanings in the text, and words and meanings are abstract entities), so a text is at once physical and non-physical. There is also a strong suggestion that a text is both one and many, for it is well known that one and the same text can be found in its entirety in many places simultaneously. The physical/non-physical ambiguity is pervasive in talk of texts, and language users often appear completely unaware of that duplicity. The one/many ambiguity seems to me to be less firmly entrenched: I believe that people are more or less conscious of the ambiguity and fully prepared to make a distinction between the one text and its many exemplars.

It is easy to show that texts are indeed being spoken of and written about in the manner just described; examples could be multiplied ad infinitum. Let us first look at two passages about literary texts taken from David Damrosch's *How to Read World Literature* (2009). The first quote refers to Voltaire's *Candide*:

> Interestingly, Voltaire actually presented his book as a translation from the outset. Rather than publish his religiously and sexually scandalous tale under his own name, he had the title page declare that *Candide, ou l'optimisme* was "Traduit de l'Allemand de Mr. le Docteur RALPH" (translated from the German of Dr. Ralph). (2009, 68)

In the quote, *Candide* comes across as being at once physical—as a material book, provided with a title page—and non-physical—as containing words ("Traduit de l'Allemand …") and meaning (being a "religiously and sexually scandalous tale"). The citation nicely illustrates the tripartite nature of a text, such as texts are ordinarily conceived of.

Concerning the one/many ambiguity, consider the following remark by Damrosch about Orhan Pamuk's *Snow* (*Kar*):

> Without translation, the novelist Orhan Pamuk would be unknown outside his native Turkey; thanks to translations, his haunting novel *Kar* can be found in Mexico City airport under the title *Nieve*, bought in Berlin bookshops as *Schnee* and ordered from Amazon.com in its English version, *Snow*. (ibid., 65)

One and the same novel can obviously be found in its entirety in many places at once.

The citations above refer to written, literary texts, but talk of oral texts and non-literary texts follows the same pattern. As a last example I will use

President Obama's 2014 State of the Union address. Barack Obama was speaking to a joint session of the United States Congress on 28 January. On the next day, Paul Steinhauser wrote on the CNN web page:[9]

> Less than half of those surveyed who watched President Barack Obama's State of the Union address reacted very positively to it, a new poll showed.
>
> And [a] while the President emphasized in his speech on Tuesday night that he's willing to bypass Congress and take executive action to accomplish his goals, a CNN/ORC International survey also indicated that only three in 10 said Obama should make unilateral changes to deal with major issues.
>
> During his address, the President told lawmakers that "I'm eager to work with all of you. But America does not stand still, and neither will I. So wherever and whenever I can take steps without legislation to expand opportunity for more American families, that's what I'm going to do."
>
> But two-thirds of those questioned in the poll said that Obama should seek bipartisan compromise when dealing with major issues, with just 30% advocating the President make unilateral changes.

The now familiar pattern emerges here too. The speech by President Obama is a physical event: it can be seen and heard by the members of Congress and by television viewers. But the speech also contains words, like those cited verbatim in the block quote, and it has a meaning: for instance, the President declared himself willing to bypass Congress and take executive action to accomplish his goals. The original speech took place on 28 January 2014, but the speech still exists: it can be found in its entirety in many different places, videotaped or as a written text, and various avatars of the speech can be heard, viewed, or read by different people in various places but at the same time.

---

[9]   Paul Steinhauser, "CNN Poll: Speech Watcher Reaction Less Positive Than Previous Years". http://edition.cnn.com/2014/01/29/politics/cnn-poll-sotu-response/index.html?iid=article_sidebar (16 March 2014)

2. The ordinary conception, the cluster conception, and the philosophical conception

As I already pointed out, the ordinary conception of a text is illogical: nothing can be at once physical and non-physical, one and many. Yet that observation is by no means intended as a criticism of Damrosch or Steinhauser, for the ordinary concept of the text is very practical in everyday contexts like theirs, and that is what gives the ordinary concept its raison d'être. I would not want to see that concept banished—except, certainly, from ambitious theoretical analyses of matters having to do with linguistic communication.

It is, however, possible, and sometimes highly motivated, to distinguish clearly between physical exemplars, complexes of signs, and meanings, thus removing all physical/non-physical and one/many ambiguities. For example, my latest book, *The Concept of Literary Application: Readers' Analogies from Text to Life*, was produced in 300 copies. Those copies are physical objects. No matter what copy you open, you will be able to construe the same complex of signs—the same sequence of sentences in English, not to put too fine a point on it—and the same meaning. Different readers may not arrive at an exactly identical meaning, but I will not address the complexities of textual meaning here. For present purposes, we can simply say that there are, right now (things may change), 302 objects with which one potentially has to count when referring to my latest book: the 300 physical copies, the immaterial complex of signs, and the immaterial meaning. When speaking of my latest book, one is in reality speaking of that cluster of objects—of the cluster as a whole, or of parts of it, or of aspects of it. An analogous approach can be adopted to any text (although many cases are far more complicated than my chosen example). Let us call that way of thinking about what a text is "the cluster conception of a text".

The cluster conception differs substantially from the ordinary conception of a text. The ordinary conception achieves its attractive simplicity by neglecting a number of distinctions carefully respected by the cluster conception. The distinction between different copies is not being heeded: viewed through the optics of the ordinary conception, my latest book is *one* object—as if one physical copy, no matter which, were standing in for them all. Further, the copy-no-matter-which is conceived of as a physical object but also as being provided with text (a sequence of sentences) and a meaning: such distinctions as those between patterns of printer's ink

(physical traces) and letters (non-physical signs), and between sentences (sequences of non-physical signs) and meaning (propositions or proposition-like entities), are not being made. The result of ignoring those distinctions is the text as the ordinary conception presents it—and that is a text you can certainly read, for it is a physical thing, even though the text also has abstract constituents of various kinds (words, meaning).

The standard contemporary philosophical conception of a text is something else again. In a sense at least, the philosophical conception is derived from the ordinary conception. The aim of philosophical conceptual analysis is to make concepts clearer by analysing them. The idea is that a somewhat muddled concept, or a concept of which we have a somewhat muddled idea, can be exposed to philosophical scrutiny and reflection, and that the scrutiny and reflection can finally produce a tolerably perspicuous and logical version of the concept. Let us follow a philosopher explaining the philosophical conception of a text and providing a motivation for it: David Davies, in his book *Aesthetics and Literature* (2007).

Davies is speaking of literary texts (literary *works*, in his terminology, which is also the terminology normally used in analytical aesthetics), not of texts in general, but that makes no real difference to me: my concept of a text includes both literary and non-literary texts, and it should be obvious that a novel printed in 300 copies exists in precisely the same manner as my latest book, whatever that manner is. Davies has just discussed some intricate arguments for the idea that paintings are physical objects, and he now proceeds to dealing with literary texts.

> Fortunately, we don't need to dwell further on these questions in order to see that, were it to be suggested that *literary* works are physical objects, no such subtle stratagems would avail. For it is obvious upon even the briefest reflection that there is no physical object which has anything like the identity and persistence conditions of a literary work. Any individual copy of a literary work—including the original manuscript—can perish without threatening the work's existence. Nor can we identify a literary work with the entire collection of physical copies of a book. For one thing, this would mean that the work was still coming into existence as long as new copies were being printed. And we surely want to insist that, even if all physical copies of a literary work were destroyed, and no new copies were ever printed, the work itself could still persist as long as it was preserved in the memory of either its author or those who have read it. For similar reasons, when we spoke ... of the artistic

> vehicle of a literary work as a text, we cannot have meant by that a par-
> ticular physical object or collection of physical objects. But what, then,
> did we mean? (2007, 18-19)

Davies then points to a specific abstract entity as being identical to a lite-
rary work: a "text-type as used as an artistic vehicle in a particular genera-
tive context" (ibid., 30)—that is, approximately, a given sentence or se-
quence of sentences as actually used by a person on some specific occasion.
Davies' analysis is a representative example of how the philosophical con-
ception of a text comes about. The physical part of the ordinary version of
the text is being amputated, as it were, and that is why the text, in the phi-
losopher's version, cannot literally be read.[10]

It needs to be said that *Aesthetics and Literature* is an introductory text-
book, but I am still confident in presenting David Davies's general way of
thinking as being representative. The ontology of art, which includes the
problem of what kind of entity a (literary) text is, is one of Davies's special-
ties. The idea that texts are abstract entities is the majority view in philos-
ophy, and Davies's diagnosis of what kind of abstract entity a text is,
quoted above, comes very close to those of other leading aestheticians like,
for instance, Jerrold Levinson and Robert Stecker.[11] The methodological
problems with Davies's analysis, to which I attend in the next section, are
equally typical.

3. Some methodological issues

The most interesting question, faced with the philosophical conception
of the text, is of course how good and reasonable that conception is. Before
coming to that, however, I would like to bring up some methodological is-
sues while David Davies's argument is still fresh in our minds.

---

[10]   It is easy to show that the idea of a text, literary or non-literary, does not reduce to
the idea of a physical object, and that result can be achieved in several ways. For exam-
ple, Mag Uidhir argues that repeatable artworks (like musical and literary works) must
be abstract objects since repeatable artworks "cannot be coherently or viably construed
as concrete things" (2013b, 8, note 4).

[11]   See Levinson (1996, 146), (a musical work is "a structure-as-indicated-by-P-at-t-
in-musicohistorical-context-C") and Stecker (2003, 88) (literary works are "structures-
in-use"). See also Levinson (2013).

Davies's analysis has no empirical backing. As far as I understand, his analysis of the ordinary conception is based on his own understanding of the concept of a text in his capacity of a competent user of English, and on profound familiarity with the philosophical discussion about that concept in his philosophical tradition, and on philosophical reflection. But he does not seem to have made any attempt to collect examples of how the ordinary concept of a text is actually employed, and for what purposes, in order to get a firmer grasp of its logic. If Davies had actually sought to gather and analyse that kind of evidence, he might well have seen that the ordinary conception of a text is logically incoherent (but has other redeeming features, more important than logical coherence in very many contexts).

As it is now, Davies is probably not aware of the ambiguities built into the ordinary conception of the text, and that creates its own problems. The easiest way of making the point I am driving at may be to say that, as a consequence, Davies is probably also unaware of the very substantial differences between the concept he is analysing (the ordinary conception of the text) and the concept that comes out of his analysis (his philosophical conception of the text). Indeed, there is no sign that he keeps the two conceptions apart.

Davies is right in pointing out that a text, such as texts are portrayed by ordinary language, cannot be reduced to a physical object or a collection of physical objects. But nor can one reduce the text, in its ordinary version, to an abstract object, as Davies does. (After all, as long as we are speaking in everyday, common-sense terms, texts can certainly be read.) To be able to present texts as logically impeccable entities, one will have to depart from the ordinary conception of a text. If one wishes to think of a text as one specific logically impeccable object, one will simply have to get rid of some of the ordinary concept's contradictory elements. The text will have to be made either physical or non-physical, either one or many. Davies chooses to make it one and to make it abstract, but he does not see that that is a decision on his part rather than an analysis of the ordinary concept of a text. There is a confusion at the bottom of the philosophical conception of the text: the philosophers do not quite perceive that the ordinary conception and their own conception are two different things.

Philosophers like to disqualify such alternatives as the cluster conception by portraying them as absurd denials of the incontrovertible fact that texts exist—that is, of the supposed fact that, for every text, there is some distinct object that is the text itself. But I do not deny that my latest book

exists, which would be ludicrous. The criticism builds on a mistake. The trump card in the philosophical argument is the demonstration that the cluster conception and similar notions do not tally with the ordinary conception of the text, according to which the text is a specific, single object. But the ordinary conception cannot be the criterion of correctness, for the ordinary conception is illogical. Texts, such as the ordinary conception portrays them, certainly cannot exist in reality (but only in language and thought), for they are contradictions in terms. Indeed, the philosophical conception, too, would have to go if the ordinary conception were to be taken as the criterion of correctness, for the philosophical conception does not measure up to the ordinary conception either. According to the philosophical conception, texts cannot be read—an absurdity, if one adopts the standpoint of the ordinary conception.

In brief, I believe that the philosophical discussion of what a text is exhibits several weaknesses. There is too uncritical an endorsement of our received concept of a text, too little interest in empirically grounded investigation of the structure and function of the concept, and too little attention to what is sometimes called "the paradox of analysis"—the fact, in principle well known, that philosophical analysis inevitably changes the concept under analysis to a larger or lesser extent.[12]

4. Problems with the philosophical conception of a text

The dominating philosophical conception of a text can be said to be that of a complex of signs as used on a specific occasion by a certain sender. As we saw, the philosophical conception diverges strongly from the ordi-

---

[12]   Robert Stecker observes that his own characterization of musical and literary works as "abstract structural types" introduces a concept that is "a theoretical one not in ordinary use", something which I find perceptive and candid. Stecker nevertheless denies that his concept is "revisionary", since he thinks that it "captures a maximal set of shared intuitions about such works" (2003, 94, note 12). It is tempting to observe that the intuition that literary works can be *read* seems a highly significant intuition not captured by Stecker's concept. But the real problem with reasoning such as Stecker's lies elsewhere: the value of our "intuitions" about literary works/texts is taken for granted, and no attempt is made to understand and evaluate the structure and function of the ordinary concept of a text. That would require, as a basis, an empirical study of how the concept is actually used.

nary conception of a text, but is the philosophical conception a good one, considered in its own right?

The first question to ask is: a good conception for what purpose? We should not answer: a good conception for taking the place of the text such as ordinary language portrays the text, for it should be clear by now that the ordinary conception is intellectually untenable despite its practical usefulness, and that its practical usefulness depends on its lack of logical consistence. The concept is economical to use precisely because it disregards a number of fundamental distinctions, and nothing can occupy its place successfully without being equally contradictory. What we really need—not in unpretentious contexts, but when theoretical precision is required—is a better conceptualization of linguistic communication than the everyday model that revolves around the sending and receiving of texts such as ordinary language portrays them.

In my view, the concept of a complex of signs as used on a specific occasion by a certain sender does have a place in such a conceptualization. David Davies and his fellow philosophers are entirely right in insisting that some of what listeners and readers establish in a genuine situation of real-world linguistic communication is such a contextualized complex of signs. A sentence or a sequence of sentences can, in itself, often be given innumerable different concrete interpretations—the linguist Roy Harris has pointed, very appositely, to the one-word sentence "Yes" (1981, 200)—but the listener or reader in a genuine situation does not treat the complex of signs as a linguistic formula in abstracto but as something uttered by a specific person and supplied with a specific communicative point. I have no objection to that analysis, and when offering the cluster conception's picture of a text and speaking of a complex of signs as forming part of that picture, I had in mind a complex of signs as used by a given individual on a given occasion. I do thus not find any fault with the concept per se. When explaining how linguistic communication works, it is difficult to get by without the idea of a complex of signs as actually used by a sender on a specific occasion.

I can see no reason, however, *to identify the complex of signs with the text*. In ordinary language, when speaking informally, we have the ordinary conception of the text at our disposal, and that conception serves us well. And in purely theoretical contexts, when attempting to achieve the best possible understanding of linguistic communication, we do not need the concept of a text. The philosopher's ambition to have a well-analysed conception of a text is a consequence of the mistaken elevation of the ordinary conception

to a norm. Since, according to the ordinary conception, texts exist, it is felt that one needs to be able to point to an object that is the text itself, and the contextualized complex of signs can be thought to be a promising candidate. In reality, however, the complex of signs does not fill that bill: the complex of signs does not answer at all to the ordinary conception of the text. Unlike the text, the complex of signs cannot be read—nor, strictly speaking, does the complex of signs have a meaning. (True, the complex of signs has a meaning in a manner of speaking since there is meaning associated with the complex of signs, but the complex of signs itself cannot be said to actually have a meaning.)

Also, if one compares the philosophical conception with the cluster conception it becomes obvious that the philosophical conception only covers part of what is involved in the ordinary talk of texts. The physical exemplars are being left aside, and also the meaning. The philosophical conception picks out one element in the cluster in an ultimately arbitrary fashion. The contextualized complex of signs can certainly be made the starting-point of a definition of the whole cluster: the relevant exemplars can be defined as the exemplars exhibiting the relevant contextualized complex of signs, and the relevant meaning can be defined as the meaning associated with that contextualized complex of signs. But one could just as well choose a physical exemplar as one's starting-point, defining the other relevant exemplars as the exemplars exhibiting the same contextualized complex of signs and the relevant meaning as the meaning associated with that contextualized complex of signs. The contextualized complex of signs does not occupy a privileged position within the cluster.

It is perhaps easy to believe that the contextualized complex of signs must, after all, be the systematically central factor in the cluster, the element that, in the final instance, really defines and determines the others. But that is not so. If one looks for a systematically central factor, one should look to a physical object. Let me round off the discussion of the philosophical conception with a concrete example.

In 1862 or 1863 Emily Dickinson wrote an untitled poem beginning "I heard a fly buzz when I died". An original manuscript is preserved: a single copy in Dickinson's own hand, now in the Amherst College Archives and Special Collections.[13] If something can be said to define the poem, it is

---

[13] See Dickinson (1981, 591). Cf. Emily Dickinson Museum – The Manuscripts: http://www.emilydickinsonmuseum.org/emily_manuscripts (12 June 2014).

that original manuscript. Certainly not the contextualized complex of signs that the manuscript represents, for the idea of what that complex of signs should be taken to be has changed considerably over time. Dickinson's early editors can be said to have treated her somewhat condescendingly, "improving" her texts by removing some of their unconventional formal traits, something which also affected "I heard a fly buzz when I died". That practice is now a thing of the past, but editors still waver about how best to define the complexes of signs associated with her poems. In particular, Dickinson frequently uses an idiosyncratic sign, a kind of punctuation mark often described as an elongated dot. For a long period, such marks were consistently rendered as n-dashes by Dickinson's editors, but in the latest scholarly edition, R.W. Franklin's 1998 edition of Dickinson's collected poems, the marks are presented as a kind of free-standing hyphens.[14] The contextualized complex of signs associated with Dickinson's "I heard a fly buzz when I died" is thus a contested entity hard to pin down, but not so the slowly yellowing piece of paper with traces from her pen that is her original manuscript. That manuscript can be said to define her poem. To put it simply: the manuscript is one of the exemplars, and the other exemplars are the copies of (copies of …) the manuscript. The contextualized complex of signs is the complex of signs represented by that manuscript and hence by every adequate copy (whatever we take that complex of signs to be), and the meaning of Dickinson's poem is the meaning of the complex of signs as produced by her in 1862 or 1863. (Much more could be said about textual meaning—I do not want to convey the impression that I take textual meaning to be a simple or even a singular entity—but this is not the place or the time.)

5. A trivial paradox and its deeper interest

Texts can certainly be read. The paradox around the readability of texts only arises when one attempts to press the ordinary concept of a text to perform heavy theoretical duty for which it was never conceived.

The paradox or seeming paradox is a symptom of dual allegiances in the philosopher: adherence to a traditional framework of views and approaches

---

[14] See Dickinson, "591", in Dickinson (1998, 587-588). The volume editor R.W. Franklin also registers the textual history of the poem (ibid., 588).

and, at the same time, aspiration to theoretical consistency. The paradox is a sign of the impossibility of simply combining those two attitudes, and therein consists, for me, its deeper interest. We should learn from the paradox that the concept of a text belongs in ordinary language, while the more fine-grained conceptualization of linguistic communication requires a different outlook and partly different notions.

References

DAMROSCH, D. (2009): *How to Read World Literature*. Malden, Mass.: Wiley-Blackwell.

DAVIES, D. (2007): *Aesthetics and Literature*. London and New York: Continuum.

DAVIES, S. (2005): Ontology of Art. In: Levinson, J. (ed.): *The Oxford Handbook of Aesthetics*. Oxford: Oxford University Press, 155-180.

DICKINSON, E. (1981): *The Manuscript Books of Emily Dickinson*. Vol. 1. Ed. R.W. Franklin. Cambridge, Mass. and London: The Belknap Press of Harvard University Press.

DICKINSON, E. (1998): *The Poems of Emily Dickinson: Variorum Edition*. Vol. 2. Ed. R.W. Franklin. Cambridge, Mass. and London: The Belknap Press of Harvard University Press.

HARRIS, R. (1981): *The Language Myth*. London: Duckworth.

LEVINSON, J. (1996): *The Pleasures of Aesthetics*. Ithaca, N.Y.: Cornell University Press.

LEVINSON, J. (2013): Indication, Abstraction, and Individuation. In: Mag Uidhir (ed.) (2013a), 49–61.

MAG UIDHIR, C. (ed.) (2013a): *Art and Abstract Objects*. Oxford: Oxford University Press.

MAG UIDHIR, C. (2013b): Introduction: Art, Metaphysics, and the Paradox of Standards. In: Mag Uidhir (ed.) (2013a), 1-26.

PETTERSSON, A. (1981): *Verkbegrepppet: En litteraturteoretisk undersökning*. Oslo: Novus.

PETTERSSON, A. (2012): *The Concept of Literary Application: Readers' Analogies from Text to Life*. Basingstoke and New York: Palgrave.

REDDY, M. (1979/1993): The Conduit Metaphor: A Case of Frame Conflict in Our Language about Language. In: Ortony, A. (ed.): *Metaphor and Thought*. 2nd ed. Cambridge: Cambridge University Press, 164-201.

RUDNER, R. (1950): The Ontological Status of the Esthetic Object. *Philosophy and Phenomenological Research* 10, 380-388.

STECKER, R. (2003): *Interpretation and Construction: Art, Speech and the Law*. Malden, Mass.: Blackwell Publishing.

Contribution to Fictional Epistemology

GÖRAN ROSSHOLM

Department of Literature and History of Ideas. Stockholm University
SE – 10691 Stockholm. Sweden
goran.rossholm@littvet.su.se

ABSTRACT: In his article "Who is Who in the Fictional World", Petr Koťátko argues that fictional worlds are, in general, complete, that is the logical law of the excluded middle holds in fictional worlds, though he admits that there are exceptions—he mentions Samuel Beckett's trilogy as an example. The present article agrees with both these conclusions, and it continues Koťátko's discussion by suggesting an explanation why so many scholars have claimed that fictional worlds in general are incomplete, and by presenting different kinds of exceptions from Koťátko's basic position, and, finally, by sketching alternative interpretations of these examples of incomplete fictional worlds.

KEYWORDS: Completeness – fictional worlds – incompleteness.

Is the number of Sherlock Holmes' cousins odd or even? Or is it neither odd nor even? Are fictive worlds in distinction to the actual world construed in such a way that some statements are neither true nor false? Basically, this discussion concerns the ontology and epistemology of narrative fiction, what exists (in a manner of speech) in a certain fictive world, and how we reach knowledge (figuratively speaking) about its contents. At the end of this essay I will turn to the question of how the conclusions drawn relate to what I will call "narrative immediacy"—the quoted term will be explained in that context.

*   *   *

Roman Ingarden, in his *Das Literarische Kunstwerk* from 1931, first formulates the presumed fact that fictive worlds in distinction to the real

world are incomplete. His example is a table in a hypothetical novel. The material of the table is not mentioned, and, consequently, it does not consist of any particular material. Ingarden calls such incompletenesses "Leerstellen". He contrasts them to the conditions that hold for the real world: "In a real object such Leerstellen are not possible. At most the material is unknown." Thus, according to Ingarden, incompleteness in fiction is ontological, and in the real world it is "at most" epistemic.[1]

After Ingarden, the discussion about fictive incompleteness has been stimulated and refined by the influence of possible world semantics on literary theory. Within this theoretical frame, Lubomír Doležel (1998) has claimed that fictive worlds are incomplete, while Marie-Laure Ryan (1991) has argued against the general validity of the incompleteness-thesis from the same point of departure.[2] David Lewis' article "Truth in Fiction" from 1978 has played a crucial role for both these literary scholars. A simplified description of Lewis' position with respect to incompleteness runs as follows: there are (at least) two principles relevant for the question of what is true in a particular fictional world, one rooted in the actual world, and one rooted in the "collective belief worlds of the community" of the work in question. Thus, in order to conduct an investigation concerning what material Ingarden's table was made of, we may either ask ourselves what the world would be like if the book were told as a true story, not as a work of fiction, or, in the second case, we would ask ourselves what would be true if the book were true against the background of common convictions in the author's community, when this background is taken as true as well. In the second case we would certainly have Leerstellen, as some questions are always left undecided by the body of shared beliefs in any community, and when these questions coincide with gaps in the book, we have instances of incompleteness. The first approach also admits for Leerstellen: if the hypothetical novel were true, it is as likely that the mentioned table would be made of oak as of birch (and so on). The question concerning what it is made of is left unanswered.

In logical terms, this leads to an anomaly: the fictive world of Ingarden's hypothetical novel contains neither "t is an oak table" nor "t is not an oak table" as truth, that is, the logical law of the excluded middle is violated.

---

[1]   See Ingarden (1960, 264-265); the quoted translation is my own.

[2]   For an exposé of the application of possible world semantics in narrative and fiction theory, see Ryan (2005).

Kendall Walton calls his pair of principles, roughly corresponding to Lewis' pair, "the Reality Principle" and "the Mutual Belief Principle" (Walton 1990, 141-161). Like Lewis, Walton stresses the fact that the result generated by one of them may contradict the result generated by the other, and, again like Lewis, both principles accord with the thesis of fictive incompleteness. Walton illustrates how the two principles work by presenting an example, with two different readings, of one and the same hypothetical work. One reader, Loretta, reads a fictional narrative about a character, Andy, who behaves in certain anti-social ways. After consulting modern medical expertise, she concludes that Andy suffers from an inherited neurological disorder. Mabel reads the same book but establishes her epistemic background by consulting historical archives. She concludes that Andy is possessed by the devil, an explanation well in accord with beliefs held in the author's community. Loretta applies the Reality Principle and Mable the Mutual Belief Principle. One lesson of this is that the phrase "what's true in the fiction" may cover more than one truth. Moreover, since neither what is true nor what was commonly believed to be true at the time and place of the writing of this novel will dictate the answer to all the questions regarding the material of all pieces of furniture mentioned in the book, there will be Leerstellen in it. Walton's example may also be used to illustrate the fact that the metaphorical terms "truth", "conviction", and "knowledge" always applies in relation to a certain interpretation or reading—both Mabel's reading and Loretta's reading generate "knowledge", in spite of the fact that they are in conflict.

<p style="text-align:center">*　*　*</p>

Like Ryan, Petr Koťátko critically discusses the incompleteness thesis. His example is Balzac's novel *Splendeurs et misères des courtisanes.* His example runs as follows:

> "Was Mme d'Espard's gall bladder [at a certain time] in good condition?" (Koťátko 2010, 97)

Since nothing is stated or presupposed about the condition of her gall bladder, and since neither the Reality Principle nor the Mutual Belief Principle settle the case, both assumptions are excluded from this fictional world in accordance with the incompleteness thesis. It is not the case that Mme d'Espard's gall bladder is in good condition, and it is not the case that

her gall bladder is not in good condition. Koťátko concludes: "A novel whose world would be inhabited by such bizarre creatures ought to aspire to some genre rather different from 'scenes from Parisian life'". (ibid., 97) I agree; the very idea that fictional worlds are inhabited by incomplete persons is absurd. Our knowledge about them is incomplete, but they are not incomplete themselves. Incompleteness is epistemic, not ontological; in Koťátko's words:

> It is [then] right to say that our *construction* of the character is incomplete—but the incomplete construction of a character is something very different from the construction of an incomplete character. (ibid., 99)

In his article, Koťátko refers to realistic novels of the nineteenth century—novels by Balzac, Stendahl, Tolstoy. Does his defence of completeness comprise all fictional narratives, or are they valid only for realistic fiction?

A final quote about Balzac's novel seems to indicate the latter, more limited, interpretation: "To give up this assumption [that is, the assumption of completeness] means to start reading the *Splendeurs* as one of the novels of Beckett's Trilogy". (ibid., 99) Thus, he points at the possibility to interpret the fictional worlds of Samuel Beckett as ontologically incomplete.

I will come back to this, but I will first address a question not discussed by Koťátko: why do many scholars treat fictional and factual discourses so differently with respect to incompleteness? As said, Ingarden calls the incompleteness of fiction "Leerstellen" and contrasts them to the conditions that hold for the real world: "In a real object such Leerstellen are not possible. At most the material is unknown". Why not just say the same about fictional and factual accounts that mention a table without mentioning the material it is made of: we do not know the material, and that is all. Koťátko does not discuss this question.

My first approximate and preliminary suggestion as an answer to this is: the reader of a fictional work receives "the whole truth and nothing but the truth". We may know that a certain factual statement is true or false for various reasons. We may see with our own eyes that it is true (or false), or we may infer from a reliable source, other than the text we read, that it is true (or false), but we believe that a certain fictional statement is true in the fiction simply because the text says so. If we want to reject or revise the information we receive from a factual text we are free go to other, more reliable, sources. The understanding of a fictional narrative coincides with, or

even is the same as, the recognition of fictional truth. The understanding of a factual report is not the same as the recognition of its truth. The reader of fiction is in the highly privileged state of being informed about "the whole truth and nothing but the truth," and this state is impossible to achieve in real life. Furthermore, this intimate relation between understanding and truth has a consequence with respect to the possibility of completion. When we have read Ingarden's hypothetical novel to the end, there is in principle no way to find out what material the table is made of. When we read the same sentence in a factual report, there may be difficulty in finding out what the table is made of—and it might even be practically impossible—but it is not in principle impossible. A reliable document may turn up which says that it is made of oak.

Or, put in another way: the understanding of a fictional narrative, and what is true in fiction, are two categories which come so close to each other that the border between epistemology and ontology seems to break down; the reader, and the theorist, become victims of the illusion that the two categories coincide.

*   *   *

As said, Koťátko hints at the idea that certain modernist literature may be ontologically incomplete. Is this right? He refers to Samuel Beckett's Trilogy, but I believe that Franz Kafka's *The Process* may equally well serve as an example. According to most commentators, the protagonist Josef K. is arrested in the first chapter. But how does this happen? Only one utterance made by a man—obviously, or should we say, seemingly an employee of the Court—provides the appearance of an arresting procedure: "Sie sind ja verhaftet". Nevertheless, such a formulation rather presupposes that K. already is arrested—to perform the speech act of arrest, one should say "Sie sind verhaftet" and nothing more. Moreover, after this scene, that is after having been arrested—if he is—he is free to go. Observations of this sort could, I assume, be used to support an interpretation according to which it is neither true nor false that he is arrested. In semantic terms, the word "arrest" is vague in the world of Kafka's novel, and I guess that the interpretation may be elaborated to include more juridical phenomena. Another close interpretative possibility can be exemplified by the final point of the juridical process, the death of Josef K. A common reading of the killing of K. is that he is executed after having been found guilty by the court, but Lubomír Doležel has pointed to several circumstances that indicate that we

do not witness an execution but a slaughter with no explanation (Doležel 1998, 195-196). A possible interpretative strategy could be to select convincing textual evidence for the thesis that K. is not executed, and equally convincing evidence for the thesis that he is executed. A similar strategy might be applied to question whether he is arrested or not in the first chapter. That is, the examples of incompleteness in *The Process* could easily be reinterpreted as examples of contradictions.

This case differs in several ways from Mme d'Espard's gall bladder and Ingarden's table. In the two latter cases the reader does not pay any attention to the incompleteness. The reader is not aware of the gaps unless the theorist forces her to see them, and there is nothing in the text that may give rise to an impression of a contradictory state-of-affairs, and, finally, the so-called gaps are insignificant to a global interpretation. In Kafka's novel, the Leerstellen are staring the reader in the face, they may also be interpreted as contradictions, and they are highly significant, but the crucial point is this: is the arrest/non-arrest in the first chapter, in distinction to the previous cases, the table and the gall bladder, an instance of ontologic incompleteness?

Koťátko points to a circumstance that has a bearing here. He says that even if we—that is, the readers—never get any information about the condition of Mme d'Espard's gall bladder, we have counterfactual access to relevant facts in the case. If a skilled medical doctor, with a modern medical education and with modern medical equipment, should examine Mme d'Espard's gall bladder he would be able to tell whether it was healthy or not. Balzac's world is like ours in that respect. I would like put it this way: it is presupposed in the novel that Mme d'Espard's gall bladder is *examinable* in this sense, and to be examinable means that it has a property, a property that in turn implies that her gall bladder is in a certain condition. If we try to apply a similar line of reasoning about the arrest—or non-arrest—in the first chapter of *The Process*, we will fail. No counterfactual procedure available (that is, imaginable) will decide the question about the arrest/non-arrest. To try and find such a test would be as much in vain as finding a test for deciding of every person whether or not he or she is bald. It might be argued that this amounts to saying that this incompleteness is ontological, but it may also be called an instance of semantic vagueness. Whatever label we prefer, cases like these differ markedly from Koťátko's example with Mme d'Espard's gall bladder.

Before leaving this topic I will present one more kind of incompleteness. Comical narratives require that we dispose our attention in certain ways in order not to weaken the comical point. In Astrid Lindgren's three books about Pippi Longstocking, the comical stories and the practical jokes produced by the protagonist Pippi are intentionally unclear. The reader usually has no access to Pippi's inner life, in particular whether she acts with the intention of being funny or not, and sometimes whether she acts with any intention at all. It might be argued that we should not pay any attention to these Leerstellen, indeed, we should not notice that they are Leerstellen because this should just diminish the comical effect. In these cases, the missing information is relevant in one sense, but not in another. The information is relevant for understanding Pippi's motives, that is, they are relevant for our understanding of the causal structure of the story, but they are not relevant for intended effect. On the contrary, insertions of this missing intentional information would reduce the humorous effect.[3]

Thus, the conclusion with respect to the impression of incompleteness is that two factors contribute to explain our inclination to talk about fictive incompleteness: closeness between understanding and truth, on the one hand, and narrative irrelevance in terms of story-understanding, or rhetorical effect, on the other. In addition to this, we have incompleteness as vagueness, cases which may be labelled incomplete in a more substantial sense.

* * *

Back to the epistemic basis of fiction:

As said, the formulation of the basic principle of fictive knowledge—that is, the book as "the whole truth and nothing but the truth"—is only an approximation. Firstly, we have to skip the words "the whole truth". In many cases it is clear that information belonging to the world of the work is hidden from the reader. Information about Mme d'Espard's gall bladder is unavailable but still counterfactually accessible in the sense I have indicated above. Further, in contrast to this example, the missing facts may be relevant to a more complete understanding of the story. Julian Barnes' novel *The Sense of an Ending* can be used as an illustration. The reader follows the narrator's reconstruction of his own life, a revision caused by a letter he has received, but not all questions raised by this letter are answered in the

---

[3]   This paragraph is a condensed version of the essay Rossholm (2010b).

book. The narrator is informed, by the letter, that the mother of his ex-girlfriend has died and that she has left 500 pounds and the diary of his friend Adrian, who had committed suicide several years earlier, to the narrator. However, he only receives one page of the diary. This page together with some events sheds new light on the narrator's life, but still, some questions remain: why 500 pounds? What more is in the diary? These questions are relevant for a more complete understanding of the story, but we never get the answers. That is, we certainly do not get the "whole truth". This example points to one more important circumstance. When we recognize what belongs and what does not belong to the world of the work, we cannot confine ourselves to what is said in the text, plus what is inferred from what is said and some relevant background. We must also ask ourselves what information is omitted from the presentation of the story.

One more type of narrative phenomenon may seem to conflict with my thesis even after deleting the words "The whole truth", namely narratives told by unreliable narrators. Unreliable narrators are often *not* telling "Nothing but the truth". Sometimes the term unreliable is used to refer to punctual and transparent unreliability, that is, the narrator sometimes says things that are incorrect. The reader recognizes this and also understands what is behind the words of the narrator, that is to say, the reader sees through what is said. The Swedish novel *The Dwarf* by Pär Lagerkvist is an example. Narratives of this kind do not cause any trouble—the information received by the reader is all true. Nevertheless, there is more radical unreliability. A narrative may as a whole be taken in two distinct ways, and we, the readers, are never told which is the right one. Henry James *The Turn of the Screw* is often mentioned as such a work, and, of course, the alternatives may be more that two (*American Psycho* is ambiguous in several dimensions: what is true and what is imagined, who is X and who is Y?). One more example is the stories of Baron Münchhausen, as told by himself. The reader understands that all he says is false; he is just making things up, and there is nothing behind it. He certainly does not us tell us "nothing but the truth". On the contrary, he tells nothing but lies.[4]

However, the thesis about the truthfulness of the fictive narrative is not a thesis about the truthfulness of the narrator. The thesis says that so-and-so is true because *the book* says so. The radically unreliable narratives at least tell us that they are false. The emergence of the unreliable narrator in

---

[4] Cf. Lewis' discussion of a similar case (Lewis 1978/1983, 279-280).

modern prose is, I admit, an important epistemic change in the history of literature, but it does not constitute any challenge of the accuracy of the principle "nothing but the truth". Nevertheless, there are other reasons to question the unexceptional certainty of fictive information. As mentioned in the beginning of this essay, we understand what we read against backgrounds of several kinds, most importantly what we ourselves take to be true—factually true, not fictionally true—and what we take to be generally accepted beliefs in the community of the author. Kendall Walton's example about Loretta and Mabel reading the same book, about a man with an inherited neurological disorder in Loretta's reading, and about a man possessed by the devil in Mabel's reading, might be changed a little to illustrate how our reading might be affected by exterior information. For example, when we read we only notice how the character acts, that is, Andy's anti-social behaviour. Later, after we have studied history or medicine, we could remember what we have read and conclude that the character is possessed by the devil, or that the character suffers from an inherited neurological disorder. In this case, we *add* something to our reading. We could also read as true that the character suffers from an inherited neurological disorder against the background of our medical knowledge, and later learn from studying history that, from the point of view of the community of the author, the character should be interpreted as possessed by the devil; moreover, we could also assume, in addition to this, that we have come to the conclusion that this latter interpretation is more valid than the anachronistic neurological-disorder-reading. Thus, we occasionally *revise* and *reject* our previous readings.

* * *

The conclusion that fiction reading is not immune to revisions does not mean that fictional reception copies factual reception. A historical account of a man named Andy identically spelled to the fictional novel about Andy may simply be false—his behaviour may be fabricated by the author, and even the protagonist himself may by invented. The novel could not be so blatantly untrue. Even if the principle "The whole truth and nothing but the truth" does not ultimately hold, it is still true that the gulf between understanding and conviction is narrowed in the reading of narrative fiction in comparison to factual reading. This difference is important concerning what was mentioned in the introduction, that is to say, the connection between fictional reading and direct, not mediated, information.

Before I go into that, a terminological digression: By the term "immediacy" I mean the quality of not being mediated but being experienced directly; not being represented in words, pictures or in any other semiotic medium, but being seen, heard, etc. directly (see Rossholm 2004; 2010a; 2012a; 2012b). Of course, narratives *are* mediated, by words or something else, but the term "narrative immediacy" is to be taken metaphorically. A central narratological thesis, which I believe is true, is that narrative immediacy characterizes narrative reading in general. We, the readers of narratives, tend to experience ourselves as directly experiencing what is represented by the text. Of course, narrative immediacy varies in intensity from genre to genre, text to text, passage to passage, and from reader to reader. Nevertheless, I still believe that narrative immediacy is pertinent for all kinds of narratives, factual and fictional.

However, this thesis is not presupposed in the present context. Most narrative theorists agree that some narratives are composed in ways that make the readers experience themselves as being dragged into the world of the narrative, as if directly in contact with narrative events. If you only want to subscribe to this modest thesis of narrative immediacy, the following will present the question: does the character of fictive knowledge presented in this paper in any way contribute to the experience of immediacy?

Fictive truths have a more stable ground (paradoxical as it may seem) than the corresponding factual truths; most fictive statements we accept without the shadow of a doubt. We do not doubt that Sancho Panza is sitting on a donkey, but we are less sure when reading a statement that Roman Ingarden was sitting on chair made of oak. In reality we often want evidence independent of the text, but in *Don Quixote* the text is proof enough. The kind of information this resembles is direct information, that is to say, what we are told by our eyes and ears. Most of the time we rely on what we see and hear, and when mediated information such as a text conflicts with our direct knowledge, we trust our senses. Still, it happens that we revise and reject our previous perceptual convictions, and when we do it is usually because our background beliefs have changed, a process similar to our revision of fictive beliefs.

However, in addition to this, there is one more dimension of directness. Per Krogh Hansen (2007) discusses readings of Nabokov's novel *Lolita* in an article about unreliable narrators. According to standard readings today, the narrator Humbert is unreliable, in particular when he describes the girl Dolores as a "nymphette", because his picture of her is dominated by his

own paedophile projections. Krogh Hansen demonstrates how several Danish literary scholars, ten or more years ago, agreed with Humbert, that is, they did not see Humbert as unreliable, at least not in this respect. We have, as in the imagined novel about Andy, two conflicting readings, and this situation resembles another situation with two conflicting versions of direct perceptions. For example, two persons meet a third, face to face, and they get very different impressions of the person—one as reliable and another as unreliable. Immediacy has moved to another level, and the reader can adjust his optics back and forth so that the narrating process and the narrated process come, in turns, into the focus of directness. To settle the question as to what degree such interplay can take the form of a simultaneous double perception requires empirical ingenuity. If the term "simultaneous" is taken in a more approximate sense, it seems to me obvious that we can hear—metaphorically speaking—the voice of Baron Münchhausen at the same time as we perceive him—metaphorically—raising himself and his horse out of a swamp by grabbing his own hair and lifting them up. What this final example also demonstrates is that what I call immediacy or directness is not only confined to what is true in fiction—we all know that the Baron is telling a pack of lies.

References

DOLEŽEL, L. (1998): *Heterocosmica. Fiction and Possible Worlds*. Baltimore, London: John Hopkins University Press.

INGARDEN, R. (1931/1960): *Das literarische Kunstwerk*. 2., verbesserte und erweiterte Auflage. Tübingen: Max Niemeyer Verlag.

KOŤÁTKO, P. (2010): Who is Who in the Fictional World. In: Koťátko, P. – Pokorný, M. – Sabatés, M. (eds.): *Fictionality – Possibility – Reality*. Bratislava: aleph, 89-100.

KROGH HANSEN, P. (2007): Reconsidering the Unreliable Narrator. *Semiotica* 165, 227-247.

LEWIS, D. (1978/1983): Truth in Fiction. *Philosophical Papers*. Vol. 1. Oxford – New York: Oxford University Press.

ROSSHOLM, G. (2004): Now's the Time. The Fiction Reader's Temporal Perspective. In: Rossholm, G. (ed.): *Fiction and Perspective*. Bern – New York: Peter Lang.

ROSSHOLM, G. (2010a): Fictionality and Information. In: Koťátko, P. – Pokorný, M. – Sabatés, M. (eds.): *Fictionality – Possibility – Reality*. Bratislava: aleph, 19-31.

ROSSHOLM, G. (2010b): Pippi och den intentionella tomhetens komik. In: Cullhed, A. – Johansson, C. – Westin, B. (eds.): *Otryckt och omtryckt. Pippis intentioner, Strind-*

bergs perspektiv, Becketts kombinationer och Hans nådes tider. Stockholm/Stehag: Symposion.

Rossholm, G. (2012a): Mimesis as Directness. In: Currie, G. – Koťátko, P. – Pokorný, M. (eds.): *Mimesis: Metaphysics, Cognition, Pragmatics*. London: College Publications, 14-39.

Rossholm, G. (2012b): Narrative as Story Representation. In: Johansson, C. – Rossholm, G. (eds.): *Disputable Core Concepts in Narrative Theory*. Bern – New York: Peter Lang.

Ryan, M.-L. (1991): *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Bloomington: Indiana University Press.

Ryan, M.-L. (2005): Possible-Worlds Theory. In: Herman, D. – Jahn, M. – Ryan, M.-L. (eds.): *Routledge Encyclopedia of Narrative Theory*. New York – London: Routledge, 446-450.

Walton, K. (1990): *Mimesis as Make-Believe. On the Foundations of Representational Arts*. Cambridge, Mass. – London: Harvard University Press.

Is Fictional Reference Rigid?[1]

MANUEL GARCÍA-CARPINTERO

Departament de Lògica, Història i Filosofia de la Ciència. Universitat de Barcelona
Montalegre, 6-8. 08001 Barcelona. Spain
m.garciacarpintero@ub.edu

ABSTRACT: Are ordinary proper names *rigid* when they occur in fictional discourse? In previous work (García-Carpintero 2010a) I have argued that they are not, at least when we consider the core case of *textual* fictional discourse, and also the case, dependent on it, of *paratextual* fictional discourse. From a perspective on such discourse which—on my reading of Koťátko (2004) and (2013)—I understand to be not far away from mine, Petr Koťátko (2010) argues that they are not: "proper names remain *rigid designators* in Kripke's sense ... even if transferred from 'everyday' communication to literary texts" (*op. cit.*, 94). After explaining how I understand the relevant concepts in the first section, I'll critically take up Petr's arguments in the second.

KEYWORDS: Fiction – fictional reference – reference – rigid designators.

1. Fictional reference and its varieties

We should distinguish three types of discourse related to fiction, which pose their own specific problems. Consider these sentences:

(1)   Mr. Leopold Bloom ate with relish the inner organs of beasts and fowls.

---

(2)   (According to *Ulysses*) Mr. Leopold Bloom ate with relish the inner organs of beasts and fowls.

(3)   Leopold Bloom is a fictional character.

Take firstly an utterance of (1) by Joyce, as part of the longer utterance of the full discourse which, with a measure of idealization, we can think constitutes his creation of *Ulysses*. It is distinctive of such uses, which I will be calling *textual*,[2] that they are not intuitively truth-evaluable. The other two uses differ in that they appear to be truth-evaluable. There is, first, the use of sentences such as (1) that we make when we are stating the content of a fiction. I will call these uses *paratextual*; according to Lewis (1978) and other philosophers, they are simply elliptic for intuitively equivalent ascriptions of propositional content such as (2). Finally, I will call uses of sentences such as (3) *metatextual*; they are similarly intuitively truth-evaluable but not directly content-reporting, in that they are not (or at least not obviously) equivalent to propositional content ascriptions like (2).

Philosophers take very different views on the way singular terms in these utterances work. Kripke (1980) argued that a proper account of metatextual uses requires interpreting names such as 'Leopold Bloom' in them as referring to fictional entities. The most influential fully developed argument for such realism about fictional entities is van Inwagen's (1977) Quinean appeal to non-eliminable quantification over, and reference to, such entities in *prima facie* serious, truth-evaluable discourse, such as utterances of (3) and related metatextual uses in contexts of literary criticism. Such *ficta* could then be taken to be Meinongian non-existent entities, concrete non-actual *possibilia*, or (as both Kripke and van Inwagen prefer) abstract existent entities of various sorts, fully-fledged Platonic *abstrakta* or rather created artifacts, as in Salmon (1998), Thomasson (1999) and Schiffer (2003). Fictional entities of any of these sorts could also be invoked to account for either of the other uses, textual and paratextual, but this requires extra work; for neither of those entities can be straightforwardly taken to be the sort of thing capable of eating birds' inner organs.

The intuitive obviousness of negative existentials involving fictional names counts against non-Meinongian realist views, a point that Everett (2007; 2013, ch. 7) forcefully presses. Everett (2005; 2013, ch. 8) provides

---

2   I take this and the other two related labels from Bonomi (2008).

an interesting elaboration on equally well-known indeterminacy concerns about fictional realism, echoing Quine's (1948, 23) indictment:

> the possible fat man in that doorway; and, again, the possible bald man in that doorway. Are they the same possible man, or two possible men? How do we decide? How many possible men are there in that doorway? Are there more possible thin ones than fat ones? How many of them are alike?

While focusing on metatextual uses leads naturally to think of the referential expressions in (1)-(3) as in fact referring to certain entities, and hence to some form of realism, focusing instead on textual uses leads to an altogether different, anti-realist picture. When the creator of a fiction uses declarative sentences such as (1) (as, indeed, when she uses sentences of other types, imperatives, etc), we do not feel tempted to think of her as really performing the speech acts one typically performs with them in default contexts. In such cases, the sentences are used in some form of *pretense*, the way we take the actions that actors perform on stage: they do not need to be drinking whisky, they are merely pretending to do so.

Now, if the apparent assertions are merely pretend, perhaps we could say the same of the apparent acts of reference; and in this way perhaps an avenue is opened to account for such uses without the need to posit actual referents for fictional singular terms. Walton (1990) has provided a very elaborate and deservedly influential account of textual uses along such lines, which he then extends to deal with both paratextual and metatextual uses; Everett (2013) provides a compelling elaboration of the program. As above with the realist picture, the extension is not so straightforward, for there does not appear to be any pretense in assertions of (3). A possible option would be to combine fictional realism for (3) with a pretense-theoretic account of authors' uses of sentences like (1), but, in addition to the profligacy involved, (2) occupies a problematic middle ground for this ecumenical rapprochement, and, as Everett (2013, 163-178) emphasizes, there are many mixed cases such as 'Anna admires Holmes'.

In previous work (García-Carpintero 2010b) I have defended a form of anti-realism for metatextual discourse, a version of Yablo's (2001) *figuralist* brand of fictionalism. Its classification as a sort of *anti-realism*, however, is a delicate matter. The proposal is, basically, that in metatextual uses the syntactic features that Quine calls a language's *referential apparatus* (a complex set including occurring in argument-position, openness to existential

generalization and substitutivity of identicals, etc) is used in a loose, hypostatizing figurative way. However, the metaphors in question are pretty much "dead"; they are, say, like the use of prepositions with a basic spatial sense ('on', 'out') for abstract relations. Thus, in contrast to the case in true pretense-theoretic fictionalist proposals, on this view utterances in metatextual discourse are straightforward assertions with straightforward truth-conditions.[3]

Nonetheless, the figuralist proposal is fictionalist in spirit. We use 'lion' in a loose way to refer to lion-statues ('I'll meet you beside the lion'), etc. Now, even if the figures of speech we rely on in such uses are not at all creative, and should be counted as literal, deploying the lexical meaning of the expressions involved, it would be quite unwarranted to worry about the ontological status of the sort of property we literally mean by 'being a lion', its degree of naturalness, and so on and so forth. The facts of those uses, even if they should be counted as literal, "semantic", not "pragmatic", do not stand in the way of a view that, strictly speaking, 'being a lion' designates a natural (biological) kind, while in such uses it is applied to things that, in a variety of respects, *count as* such. It is with this dismissive attitude that the defender of the figurative view of metatextual discourse looks at debates between fictional realists whether characters are non-existent *concreta*, *possibilia*, or rather created or Platonic *abstracta*.[4]

Be this as it may, what matters for present purposes is that, even if it is a form of realism, the figuralist view of metatextual discourse does not offer by itself an account of referential expressions in textual and paratextual uses. In the terms of this proposal, an illuminating view should take the form of an account of the figurative content of utterances such as (2), ultimately of the nature of the discourse that (2) reports—of whatever utterances such as (1) express. As in general is the case with such contents, we

---

[3]   To circumvent the problems that this causes, in more recent work Yablo (2014) has developed an alternative framework to his earlier figurativism to present the fictionalism he favors. I like this work, and, given the role that presuppositions play on the view I present below, it would be straightforward to articulate the form of fictionalism about fictional characters I support by means of it; but I have stuck to the figurativist framework for reasons of familiarity and consequent ease of exposition.

[4]   The dismissive Wittgensteinian attitude applies only to the *ontological* worries. There are very interesting difficult theoretical issues, in linguistics and in philosophy, in understanding the lexical processes involved in these cases. My previous work provides some references.

should not count on anything like a "translation recipe", a general procedure for articulating it in any particular case; perhaps this can only be given on a case-by-case basis. What is clear is that we will not find reference to any posit of realist accounts in an acceptable characterization of figurative contents of this kind (and we had better not). Hence, we need an independent account both of the meaning of expressions such as those intuitively lacking a referent, like 'Leopold Bloom', and of those intuitively having a referent, like 'Dublin', as they occur in textual discourses of the sort illustrated by (1). I will briefly sketch the view of textual discourse I have defended elsewhere.

I support pretense-theoretic accounts of textual discourse of the kind advanced by Walton (1990) and developed by Everett (2013) and Friend (2011; 2014). My main difference with them lies in that I reject Millian accounts of singular terms, proper names in particular. To understand this, I need to outline first the account of singular thought I prefer. I take it that there are two sets of intuitive data to be accounted for by theories of singular contents. There is, firstly, the Quinean distinction between (general, "*de dicto*") belief-states concerning particular spies with contents that would be uninteresting to intelligence agencies (such as believing that the shortest spy spies) vs. those (singular, "*de re*") that do not. Secondly, there is the intuitive data that Kripke and others marshalled against traditional descriptivist accounts of the thoughts expressed by utterances including literally used proper names or indexicals, and referentially used descriptions; in particular, the "intuitions of rigidity," that when we consider possible states of affairs compatible with the truth of a given utterance we keep fixed the denotation of the referential expression in the actual state of affairs, if any.[5]

The account I support for those intuitions has (appealing to Kaplan's 1989 distinction, cf. García-Carpintero 2012) semantic and metasemantic aspects.[6] On the semantic side, we should distinguish the contribution of

---

[5] "My main remark ... is that we have a direct intuition of the rigidity of names, exhibited in our understanding of the truth conditions of particular sentences" (Kripke 1980, 14; cf. 6, 62).

[6] The word 'semantic' is here understood in a wide sense. In this sense, referential uses of descriptions have a "semantics" according to which they are "directly referential", and in fact rigid; but this is compatible with the view that these uses are non-semantic in a narrow sense (on which semantics aims at providing a compositional account of the meanings of natural language expressions), a generalized conversational implicature.

the referential expression to the content of the main speech act being made—the assertion or judgment, given that we are considering a default utterance of a declarative sentence for methodologically useful concreteness—from its contribution to an ancillary presupposition.[7] The contribution to the asserted content is the object itself, if any, as in Millian views; the contribution to the presupposition is an assumed "important predicate", a prima facie identifying property that could be offered as an answer to questions as to who or what the referent is.[8] On the metasemantic side, I take reference to be an ancillary speech act. I will assume a normative account of speech acts,[9] along the lines of Williamson's (1996) well-known account of assertion. According to Williamson, assertion is constituted by a simple knowledge rule, KR; I propose the ancillary reference rule RR, which is just an application of Russell's Acquaintance Principle:

(KR) One must ((assert p) only if one knows p)
(NR) One must ((refer to o) only if one knows who or what o is)

The semantic side of the account has affinities with other suggestions in the literature, and thus for present purposes it can be further sketched in relation to them. Thus, consider S's literal utterance of a declarative sentence such as 'he is hungry'. In truth-conditional terms, the proposal delivers familiar hedged truth-conditions (Sainsbury 2005, 54-59): *on the assumption that x is the demonstrated male that S's use of 'he' refers to, S's utterance is true iff x is hungry*. The view differs from truth-conditional accounts such as Sainsbury's in providing a more complex semantics, with two different contents: (focusing just on the contribution of a referential expression) the singular asserted truth-condition and the conventionally impli-

---

[7] We are here considering the semantics (in the wide sense) of uses of sentences, not the semantics of sentence-types. I think that the account to be outlined can be transferred from linguistic to mental acts with small variations; a useful starting point is to consider the judgments made by default by "uttering" declarative sentences in *foro interno*, instead of assertions.

[8] I borrow the notion of "important predicate" from Boër and Lycan's (1986) well-known account of knowing who, on which this is a matter of knowing identifying properties relevant for contextually specific purposes.

[9] This is one of the many fundamental agreements with Koťátko's views, even though he has advanced a different normative view of speech acts, taking the notion of normative *commitment* as primitive (cf. Koťátko 1998).

cated descriptive presupposition. In that respect, it is closer to views such as Perry (2001), with the singular asserted proposition being Perry's *referential* or *official* content, and the conventionally implicated proposition close to Perry's *reflexive* content.

In the case of a use of a proper name N, the descriptive presupposition conventionally implicated is that the use of N refers to the object saliently being called N in that use. "Being called N" in a use leads us to a contextually identified *naming practice*. I mostly agree with what Sainsbury (2005, 106-124) has to say about naming practices and their identity, in particular with his view that a naming practice has at most one referent, but may lack one. These conventionally implicated descriptive contents are usually pragmatically enriched with further assumed properties of the referent derived from context, from perception, memory, or anaphoric relations to previous discourse: on the assumption that x is the demonstrated male that S's use of 'he' refers to, *the one looking so-and-so from that perspective*, S's utterance is true iff x is hungry.

The view has a further important affinity to Sainsbury's, in that it admits "reference without referents"; utterances of 'Vulcan is bigger than Mercury' are fully meaningful expressions signifying (with respect to every possible world) "gappy" untrue singular propositions (Braun 1993; 2005). Referring expressions purport to have referents, but they do not need to have them; on the present view, this is not just a teleological feature but a normative one: they should have them for their uses to meet their constitutive norms. The account thus requires a free logic—which in my view (unlike Sainsbury's 2005 or Braun's 2005) should be a supervaluationist non-bivalent one, but we do not need to go into that here. Sainsbury's is of course a *modest* truth-conditional proposal, but I think that the relative full-bloodedness that comes from acknowledging a further level of presupposed contents is required, both to properly account for the datum of rigidity and also to help us understand how utterances signifying necessarily untrue gappy contents can nonetheless be fully meaningful.

According to Sainsbury (2005, 76-81), the explanation for the datum of rigidity is given by the "essence of reference", which on his view is captured by Evans's principle (P):

(P)　　If S is an atomic sentence combining the n-place concept-expression R with singular terms $t_1 \ldots t_n$, then S is true iff ⟨the referent of $t_1 \ldots$ the referent of t_n⟩ satisfies R.

However, I do not think this principle can adequately carry the explanatory burden. Sainsbury has to stipulate the restriction to atomic sentences so as to exclude sentences formed with definite descriptions; and he must interpret the metalinguistic descriptions used in the principle as themselves referring expression, or alternatively rigidify them by inserting 'actual', for the principle to deliver the intuitively correct rigid truth-conditions. Hence, in both respects, in order to obtain the intuitively desired results from the principle, we must already apply it in accordance with the distinction we are trying to account for. On the present view, the explanatory burden is rather carried by the metasemantic epistemic norm of reference. Some literal speech acts that language allows their users to make express singular thoughts: thoughts involving different objects (in the same "positions") would be different thoughts; properly understanding these thoughts requires knowing those objects.

The norm of reference by means of which I try to capture Russell's Acquaintance Principle requires much more elucidation than I can provide here. It requires first of all motivation, which (following Russell) I am assuming comes from general requirements on understanding. Its normative character should be further clarified, to emphasize that the norm is still in place in cases in which it is not fulfilled, the clearest among them being 'Vulcan'-like cases involving failure of intended reference. The context-dependence that its application shares with all knowledge-ascriptions should also be addressed, in the framework of a general account of such apparent context-dependence: Is it really the case that ascriptions of knowing-which/who are interest-relative, as Böer and Lycan (1986) claimed? Last, but of course not least, the (absolute, or context-relative) epistemic requirement that it imposes should be explored and clarified. In particular, I (2008) have defended that Leverrier's uses of 'Neptune', or similar uses of 'Jack the Ripper' or 'Unabomber' in worn-out examples, *meet* the requirement, in that there is a sufficiently substantial causal-evidential relation with the referent (a perceptual relation with other objects related to it, together with good evidence concerning the relevant relations). This should be further elaborated in a general epistemic framework, addressing among other matters the recent criticism of accounts like this by Hawthorne and Manley (2012, ch. 3).

A fuller exploration of these issues is out of the question here. Let us just keep in mind the two main features of the view I have outlined: an epistemic requirement on singular thought (as opposed to a merely psycho-

logical one in terms, say, of dependence on "cognitively significant mental files"); and a semantics with descriptivist features, even if they remain merely at a presuppositional level in the case of the assertoric utterances made by default when uttering declarative sentences that we have considered so far. What account does this provide for the use of names (and other *de iure* rigid expressions) in textual discourse?

As suggested above, our intuitions support *non-committal* accounts of the semantics of textual and paratextual uses of sentences such as (1), not committed to the existence of fictional entities. The most popular contemporary account in sync with those intuitions is Walton's (1990) deservedly influential theory, or variations on it such as Currie's (1990) or Lamarque and Olsen's (1994). On the version that I myself (2007; 2013) have advanced, Joyce's utterance is not an assertion, but a different speech act of pretending or make-believe, which should be understood in terms of norms stating contents that proper appreciators of Joyce's tale ought to *imagine*. As Walton rightly emphasized, these are not just pretense views; for it is not enough to say that the fiction-maker pretends to assert—or to order, ask, etc.—to fully characterize what he does, among other things because there are cases of pretense which in no way constitute fiction-making. The fiction-maker is engaged in a fully contentful intentional behavior.

This poses a problem for Walton's own proposal, or the related one by Evans (1982), which I articulated in previous work. Evans adopts Walton's appeal to practices of make-believe, but Walton follows Evans in assuming a very strong non-descriptivist version of referentialism for (most uses of) singular terms, according to which sentences including empty singular terms lack content, whether or not they are embedded in intensional contexts. Singular thoughts are object-dependent not just in that a difference in objects is thereby a difference in thoughts, as I assume is required to account for rigidity, but in that the thought cannot exist unless the object does. Now, even if Joyce's act is not an assertion but rather an invitation to his readers' imagination, the purported imaginings should nonetheless have contents; and non-descriptivists must tell us what, on their view, the contribution of names such as 'Mr. Leopold Bloom' to such contents is. Their referentialism makes it difficult for these accounts to characterize the semantic content of the contents of the acts of make-believe that on these views fictions mandate.

Let us see in contrast how an account of singular thought with the two features I highlighted at the end of the preceding section can deal with these

problems; and let us begin with a case involving indexicals, such as the short story I imagined earlier including the sentence 'he thought this just before dying'. The narrow semantics of utterances of sentences like this which I outlined above has it that they are assertions of singular contents about the most salient male when the token of 'he' is produced, if any, i.e., if the speaker meets the norm of reference relative to the purported referent. If this is not the case (say, in the most obvious case, because there is no such male), the utterance still semantically (in the narrow sense) constitutes the meaningful, though unsuccessful, assertion of a singular content.

Now, I do not think we can stop here if we want to provide an accurate account of what the fiction-maker is doing, i.e., of the semantics (in a wide sense) of textual uses. This is essentially because of a point that Walton makes in the text quoted above: "not all acts of pretence of *this* sort are of [the relevant] kind". Both when Conan Doyle writes 'Holmes is a clever detective' and when Chandler writes 'Marlowe is a clever detective', the sentences they use express the very same gappy singular proposition; in both cases the utterer pretends *to assert de re of someone that he* is a clever detective. But I do not think we want to say that the contents of *their* speech acts (the contents we, as sensible appreciators, are supposed to imagine) are those "gappy" singular contents shared by the two utterances. Fiction-producers merely pretend to assert these gappy propositions; what they want fiction-consumers thereby to imagine are not those rather uninteresting contents, but other related descriptive ones instead.

My account of sentences like 'he thought this just before dying' or 'Marlowe is a clever detective' is "two-dimensional," in that it gives us descriptive contents additional to the singular contents. These descriptive contents are still singular, in that they are about the tokens of the referential expressions or, in general, the contexts in which they are produced; in our cases, that the salient male when the token of 'he' was produced thought such and such before dying, and that the object called 'Marlowe' is a clever detective.[10] These are the contents I contend we are supposed to imagine. In textual uses, as Walton suggests, the fiction-makers are primar-

---

[10]   See my (2006) for elaboration. Two-dimensional contents like those I am envisaging where firstly introduced in Stalnaker (1978); Stalnaker provides at the end of the paper an application to 'Sherlock Holmes does not exist'. A similar idea is contained in Perry's (2001) multi-propositional proposal; Perry's reflexive propositions are described in a framework of structured contents closer to the one I am assuming than Stalnaker's possible worlds metatheory.

ily indicating by exemplification the kind of speech acts that should be imagined; we avoid Walton's and Evans' difficulties because the speech acts that the sentences they use signify by default—the ones intended to serve as illustration for those that the readers are supposed to imagine, even if unsuccessful if taken at face value—are fully meaningful.

On this view, the semantics of textual uses is descriptive, and hence singular terms *both empty and non-empty* in them are not rigid.[11] The same applies to paratextual uses, which I take to pragmatically implicate explicit ascriptions of content to fictions, along the lines of (2). To deal with the semantics of the latter, we would need to draw on neo-Fregean accounts of propositional attitudes, such as "hidden-indexical" or "interpreted logical form" views, taking advantage of the descriptive features already present in non-embedded uses.[12] Stacie Friend (2011; 2014) has objected that a descriptivist view concerning the content of imaginings cannot properly capture a certain "object-directness" intuition that we have regarding the content of fictions; her interesting points complement the arguments by Koťátko I will discuss in the next section, some of which are similar to the ones she had given before (Friend 2000). They both support the main claim I want to discuss here, which Kroon (1994)—in a persuasive critical discussion of views of this kind, providing arguments complementary to mine—states thus:

> (R)   Occurrences in fictional contexts of real proper names like 'London', 'Baker Street', 'Napoleon', and so on, are purely referential and take their usual reference.

---

[11]   In this respect, the proposal is close to Currie's (1990) and Lamarque and Olsen's (1994).

[12]   See Crimmins and Perry (1989) for the former, and Pietroski (1996) for the latter. Of course, the proposal should not have it that a paratextual use of 'he thought this just before dying' ends up ascribing to the relevant fiction a content concerning the token of 'he' used by the ascriber, or (1) one concerning a 'Bloom' naming-practice leading to the token used in its context, in the latter case because the existence of no such practice needs be assumed, in the former because such token is irrelevant to the content of the fiction. The ascribed content concerns a token of 'he' used by the fictional narrator that my account presupposes, or a naming-practice fictionally depended upon in the context of that narrator's acts. On everybody's account, ascriptions of propositional attitudes are heavily context-dependent, independently of the issue of whether or not descriptive material ends up constituting the truth-conditions of the report, as I assume.

Let me outline the main considerations against (R) and for a descriptivist view of the sort I gave in earlier work (García-Carpintero 2010a), to be further developed in the next section. The suggested view adopts from referentialists such as Braun (1993; 2005) the idea that assertions of 'Vulcan is smaller than Mars' have gappy contents. Assertions of atomic sentences with these contents are untrue, false according to the free logics that Braun and Sainsbury opt for, neither true nor false according to the supervaluationist one I prefer. This captures the fact that these assertions are *wrong*, with respect to a dimension of evaluation (truth) essential for the nature of assertions; similar remarks could be made about questions or orders with these contents. However, there is absolutely nothing wrong about the acts of fiction-makers who use empty names; there is, for instance, no appearance of "imaginative resistance" on the part of appreciators of such fictions.[13]

Similarly, by placing features accounting for differences in "cognitive significance" between 'Hesperus is smaller than Mars' and 'Phosphorus is smaller than Mars', or 'today is Tuesday' and 'tomorrow was Tuesday' (with the respective contexts of utterance coordinated so that indexicals and tenses have the same referents) at a different level than that of the asserted content—the "ways of believing" of referentialists such as Salmon and Braun, or my presuppositional level—we capture the intuitive commonalities in "what is said" among utterances made by people otherwise with very different perspectives on what they talk about, explain communicative success (cf. Perry 2001, 5, 19), and, importantly, account for our reflective intuitions about the objectivity of many subject-matters for our representational acts (Schroeter 2008). A good case can be made that these commonalities extend to straightforward assertions of 'Marlowe is a clever detective' and 'Holmes is a clever detective' by confused speakers who have taken fictional stories for factual ones; the manifest differences in cognitive significance between such utterances would be accounted for in the usual ways.

However, nothing of this sort can be said about the contents that fictions intend proper appreciators to imagine. While the mode of thinking through which we think of Venus when we assert 'Hesperus is smaller than Mars' is intuitively and theoretically irrelevant to what we assert, in that many other modes of thinking about it may do as well, the corresponding

---

[13]   Weatherson (2004) offers a good presentation of the puzzles of imaginative resistance.

modes of thinking "about" Marlowe and Holmes provided by the relevant fictions are essential to their contents: no proper appreciation can ignore them; no proper appreciation can do without building the corresponding files, starting with 'object picked out by the relevant 'Marlowe' naming practice', and stacking into it all the information about the character derived from the fiction. (We should not be misled here by the fact, which Walton 1990 emphasizes, that not all propositions constituting the content of fictions are on an equal rank with respect to a proper appreciation; many can be ignored, while still having a good notion of what the fiction is about.) All of this applies equally well to non-empty singular terms occurring in fictions, such as 'Napoleon' in *War and Peace* or 'London' in *1984*.

To sum up, I do not think that there are good reasons to support the claim that either 'Bloom' or 'Dublin' behaves like a rigid designator with respect to the content of utterances such as (1). In the first place, the descriptive content associated with those names (in particular, *person named 'Bloom'/city named 'Dublin' in relation to tokens used in* Ulysses) is *not* intuitively irrelevant with respect to that content; in the second place, it is not intuitively the case that, when we consider counterfactual circumstances to establish whether or not they constitute the contents the fiction ask us to imagine, we just consider how things are with a single Bloom/Dublin.

2. Koťátko's arguments

I now move on to critically examine Koťátko's (2010) considerations for the claim quoted above, "proper names remain *rigid designators* in Kripke's sense ... even if transferred from 'everyday' communication to literary texts" (*op. cit.*, 94); as mentioned, Friend (2000) argues for a similar claim along the lines of (R): "connected names [by which she means "real" names like 'London' and 'Napoleon' used in fictions] refer to their ordinary referents" (*op. cit.*, 186).

Koťátko starts his paper with a critical discussion of views on which "the world of the fiction" (and in fact all possible worlds) are "constructs of human productive abilities". To counter such views, he provides a toy possible worlds model for a small part of Shakespeare's *Julius Caesar*, involving Brutus, Caesar, and some relations between them. I agree that the content of fictions, as Walton (1990) has emphasized, is a sufficiently objective matter, and I of course agree with Petr that possible worlds are in

no way "constructs of imaginative abilities", so I do not have any quarrel with this argument. However, the fact that the relevant possible worlds in his model have been specified by means of directly referring expressions, and that the worlds can be considered a toy model for a part of Shakespeare's fiction, is later invoked (*op. cit.,* 94) in support of the claim I am taking issue with. This will not do. The fact that the worlds can be specified in terms of the relevant objects does not prove that *the fiction* actually specifies them in that way; hence, without further evidence, the argument is question-begging.

We should add that talk of worlds should be handled with care in this context. As is well known, in spite of the efforts by Stalnaker and others, there are very good reasons not to take possible worlds as primitive entities in our semantic endeavors; the particular case of fictions, of course, multiplies such concerns.[14] It is perfectly ok, in my view, to use them as a semantic tool, and not just for instrumental reasons; the presence of modal expressions in our languages, in addition to the need to account for semantically grounded modal notions like logical validity, sufficiently establish that. But possible worlds need not be primitive entities. They might be determined by structured propositions, for instance. In such terms, the present question is whether it is worlds determined by singular propositions, or rather worlds determined by general descriptive propositions, that we should take the textual discourse of fictions to portray.

The main argument that both Friend and Koťátko give for their related claims relies on the fact that a proper appreciation of a story including "connected" names (to use Friend's term) requires invoking information that appreciators associate with the names in virtue of their beliefs about the names' actual referents. Thus, Friend (2000) illustrates her argument with a story of her own about Anthony Everett and John Perry (both present at the conference where her paper was originally given, I gather), including witticisms to be appreciated by people familiar with those characters. She writes:

> this story was designed for a certain audience who would appreciate the allusions. So if we deny that we are supposed to be imagining, of Perry, that he never tells a joke, or of Everett, that he is a secret agent, then it seems as if we lose the point of the story altogether (*op. cit.*, 193).

---

[14]   Sainsbury (2014) expresses a much more wide-ranging skepticism about possible worlds as a tool to understand fictional content.

In a similar vein, Koťátko (*op. cit.*, 95-96) writes:

> As a reader of *The Count of Monte Christo* or of *Lost Illusions* I am confronted with a series of terms like 'The Emperor', 'Napoleon', 'Bonaparte', 'The Corsican' or 'The Usurper': then the *literary* construction of the text invites me to activate my *historical* knowledge and on its basis interpret these terms, as used in given contexts, as coreferential ... The most natural way of doing justice to this observation is to admit the real Napoleon I as the referent of the utterances of the word 'Napoleon' (or 'Bonaparte', 'The Usurper' etc.) in the literary text.

In response, I will rely on an observation that "latitudinarians" with respect to *de re* attitudes from Sosa (1970) to Hawthorne and Manley (2012) have pointed out: to wit, that even Boër and Lycan's (1986, 125-132) lowest "grade of *de re* involvement" allows for a measure of aboutness. Boër and Lycan's grades range from grade 1 (the attitude deploys an "attributively used" non-rigid description) to grade 6 (Russellian unmediated acquaintance with *sense data* or the self), through representation by means of rigidified descriptions (grade 2), representations causally dependent on the referent (grade 3) and representation by means of *de iure* rigid (linguistic or mental) terms (grade 4).[15] Boër and Lycan provide the following example that even grade 1 allows for intuitive aboutness, adapted from Sosa (1970, 894-895):

> Ludwig is a successful arsonist. The police know that an arsonist is at work in the community, but Ludwig is so discreet that they have no clue whatever as to his identity. Nevertheless, they speculate that whoever set the fires is from out of town ... and this conjecture is reported in the newspapers. Ludwig's wife reads her paper and says to him, *They think you are from out of town.* (*op. cit.*, 189)

Examples likes this are easily reproduced; many others can be found in the literature. It does not matter how such reports should be understood; even if they are literally false, what matters for our purposes is that they show that we intuitively accept as felicitous *de re* ascriptions in cases in which the content of the representation in question only involves purely general descriptions.

---

[15] Boër and Lycan (*op. cit.*, 131) suggest that the true dividing line between general and singular attitudes separates grades 3 and lower from grades 4 and higher. I agree with them, cf. García-Carpintero (2014).

Taking this into consideration, let me now go a bit further into the details of how the semantic proposal I have outlined in the previous section handles the use of singular terms, both empty and non-empty, in the production of a fiction. It should be clear that the view I have outlined in the previous section is not reductive. Far from purporting to reduce singular representational states/acts to descriptive general ones, it assumes that there are such primitively singular states. The view is only that general descriptive information helping to fix the individuals they are about is a constitutive feature of the states. This information figures in associated presuppositions. These presuppositions are themselves singular, and not just because they may mention singular token representational states; the intended referents, if there are any, also figure in their contents. Singularity is here understood as a constitutive feature of the representational devices, a "semantic requirement" (Fine 2007) on them—a fact to be embedded in a general theory of such representations, which must be grasped if they are to be fully comprehended. Finally, singular representations thus understood may fail to have an object.

Now, on the Stalnakerian (Stalnaker 1978) picture of presuppositions I partly assume, presupposition and assertion interrelate in dynamic ways. Presuppositions are checked against a "common ground", a set of propositions that are common knowledge among conversational participants. Asserted contents, if accepted, become part of the common ground, and thus legitimize presuppositions later on in the discourse. Consider an utterance of (1) in its assumed context. This is a declarative sentence that would be used by default to make an assertion. The assertion, I have suggested, is merely pretend, which is why we would not complain that it cannot be true or impart knowledge by its including an empty name. However, it behaves with respect to the dynamic of discourse exactly like the corresponding assertion would have, legitimizing presuppositions; thus, the next sentence could have been "it was not just relish that he experienced, it was something closer to devotion"—a cleft construction presupposing that Bloom felt relish—and it would feel entirely felicitous (unlike "it was not just disgust …"). It is in virtue of examples like this that the common ground is not taken to consist of propositions that are strictly speaking common *knowledge*, but merely commonly "accepted" (Stalnaker 2002). I suggest we take such an "acceptance" to be a matter of further pretense: an accepted pretend assertion becomes a pretend presupposition. Fully understanding fictional discourse involves additional pretend presuppositions to the ones created by pretend assertion:

the singular reference-fixing presuppositions that my proposal associates with empty names such as 'Leopold Bloom' are similarly merely pretend presuppositions. It is thus neither here nor there that they cannot be true, nor therefore matters of common knowledge.[16]

Not all presuppositions that a piece of textual discourse assumes are pretend, of course. As has frequently been pointed out, even the more fanciful tales assume facts that truly are (taken to be) common knowledge, in order to determine their contents. Special among them are presuppositions constitutive of the meaning of the terms the tale uses; these cannot be pretend. The singular presuppositions associated with connected names belong in this category of non-pretend presuppositions. They interact with pretend presuppositions to determine the content of the fiction, in ways that have been famously explored by Lewis (1978) in his second analysis of truth-in-fiction, by Walton (1990) for his "principles of generation" for make-believe and by many others under their influence. Does this imply that the content of fictions includes singular propositions conveyed by sentences including connected names, as Friend and Koťátko want?

This does not follow; it does not follow even if—as I think we should—we understand the pretense we are talking about here along the lines that Walton (1990) suggests and many others, including Everett (2013), have developed, i.e., as proposals to imagine. The text constituting a fiction (perhaps derivatively from intentional acts by his actual or implied author, we do not need to go into this) proposes that its readers imaginatively presuppose some contents. But this does not entail that such contents are *eo ipso* contents of the fiction. The text constitutes an act of fiction-making;[17] but not everything that the reader should imagine in order to comprehend the text needs be part of the content made fictional by it. There are imaginative acts required to understand the text that are merely ancillary to the determination of the contents that the text invites proper appreciators to imagine. Let us suppose a story beginning thus: "this is the tale of a little dragon, Urkul, who lived at a time when there were no people around and things still did not have names". The reader is to im-

---

[16] Cf. Sainsbury's (2010, 143-148) related discussion of "truth under a presupposition"—which is not, however, put in the framework of a general theory of reference involving presuppositions.

[17] As I said, I have developed elsewhere my own understanding of this, inspired by that of Currie (1990) and Walton (1990); cf. García-Carpintero (2007; 2013).

agine that x is the dragon picked out by the naming-practice on which that token of 'Urkul' relies;[18] but this is clearly not intended to be part of the content of the fiction. What is to be part of the content of the fiction is determined by relevance-like factors, relative to the nature of the act of fiction-making, including the intentions of the author, related conventions, etc, of which at a sufficiently general level we have a poor grasp.[19] Still, I think that, whatever they are, the previous claim that *that the dragon is called 'Urkul'* is not part of the fictional content is sufficiently safe.[20]

Thus, I agree with Petr that, as readers of the works he mentions, we are supposed to accept (imaginatively presuppose) singular propositions about Napoleon, stating his identity with The Emperor, Bonaparte, The Usurper or The Corsican. He might also be right that the most natural way of doing justice to this is to admit the real Napoleon I as the referent of the utterances of the word 'Napoleon' in the literary text. And yet this does not follow from the observation, as the theoretical proposal I have outlined shows: it is still compatible with the observation that the contribution of 'Napoleon' to the literary text is constituted by purely descriptive features entailed by the imaginatively presupposed singular contents: *the person called 'Napoleon', who is also called 'The Usurper', who is also called 'Bonaparte',* ... and so on and so forth, till we complete the full description that can be abstracted by ramsification (Currie 1990, 150-154) from a full presentation of the fictional content.

We have seen that what we might call the "importation" argument (the fact that we import to the content of fictions information we associate with the referents of the connected names occurring in it) does not establish (R). Searle (1974) has a related argument for what appears to be a version of the same claim, based on the reverse of importation—the other side of the coin: the fact that we tend to infer that the referents of connected

---

[18]   As I said, I assume our meta-theory will make use of a free logic.

[19]   If the text has artistic ambitions (which, together with most writers on the topic, I am not assuming is essential for fiction-making), these factors will surely include the "literary aspirations" that Koťátko (2004; 2013) has theorized about.

[20]   Byrne (1993) accuses Lewis' (1978) account of truth in fiction of "idealism", in that in all the possible worlds constituting the content of the story its declarative utterances are asserted as known facts. The point I made in the main text would deal with that objection, but it is made from a perspective crucially different from Lewis' on account of different objections I have raised; cf. García-Carpintero (2007; 2013).

names do actually have the features ascribed to them in "realist" fictions which decisively contribute to their realism, and criticize fictions betraying such dispositions. As Searle puts it, "if Sherlock Holmes and Watson go from Baker Street to Paddington Station by a route which is geographically impossible, we will know that Conan Doyle blundered" (*op. cit.*, 331). This is a practice we certainly follow, as Salman Rushdie's polemic against the film *Slumdog Millionaire* (*Guardian*, 24/2/2009) witnesses. But, as before, (R) does not follow, for it is compatible with this that the content that proper appreciators are required to imagine is descriptive. In a text I quoted above, Friend says that audiences of her story who did not connect the names in her fiction to the real Perry and the real Everett will miss the fictions' "allusions". This is correct. It is just a particular case of the more general point that fictions can be in the service of truth (Lewis 1978, 178-179); for grasping the relevant allusions involves comparing contents that we can export from the fiction with contrasting beliefs we had about real people, thereby enjoying mild jokes about them. But this is just one more case of exportation, which poses no new problems.

The two considerations in support of (R) we have examined are deflated, in a nutshell, by the point made at the beginning of this section, that is, that even representations whose contents are entirely general-descriptive can be considered to be *about* the actual entities that fit the descriptions, and reported to be *de re* about them. The theoretical framework distinguishing ancillary imagined presuppositions from fictional contents helps to develop this crucial point.

These were defensive moves. The main positive reason in favor of the view is the one I have given at the end of the previous section. As many defenders of direct reference and anti-individualism have argued, those views have a very important intuitive appeal; as Schroeter (2008, 138) nicely puts it, developing points made before by Burge and others, "commitment to anti-individualism is integral to our own epistemic agency". The assertions that we make with the help of singular terms have an objective content, in that they are essentially about their referents; the descriptive fixes we might have on them are just contingent convenient ways of having them in focus. This is reflected in the sort of intuitions of rigidity Kripke (1980) belabors, relative to an assertion of 'Aristotle was fond of dogs':

> there is a single individual and a single property such that, with respect
> to every counterfactual situation, the truth conditions of the proposi-

tion are the possession of the property by that individual, in that situation. (*op. cit.*, 10)

When we gather together the worlds meeting the conditions for applying the fundamental property for the evaluation of the claim, truth, all that matters is how things stand with the referent of 'Aristotle'; it is irrelevant how things stand with whoever was the last great philosopher of Antiquity, if that is the fix on the referent we are assuming.

Consider, however, what to say about the contribution of 'Dublin' when it comes to lining up the worlds of *Ulysses*—those that should be considered in order to meet the constitutive features of the act of fiction-making by its creator, say, to comply with his proposal to imagine and thereby appraise whether it is worth carrying out. I submit that exactly the opposite is now the case. We cannot disregard for that task the descriptive assumptions the text makes about the assumed referent; for how the city is presented in the text is essential to an adequate appreciation. And it is irrelevant how things might have counterfactually been with the actual city, to the extent that that is not something we are meant to infer from the text of the fiction. This is, at its core, what I take to be the main consideration for a descriptive understanding of the content of fictions, also when it comes to the contribution of connected names to them.

I will conclude with a concession. Kot'átko discusses an argument based on the fact that the content of fictions is indeterminate. The argument appeals to the well-known point that "the world of the fiction" appears to be ontologically indeterminate in ways that we do not take the actual world to be (at least, its past), even if we grant that it might be epistemically inscrutable in many ways. For instance, whether Mme d'Espard's gall bladder (at the time she was spinning her plots against Lucien de Rubempré) was in good condition is a fairly typical example "of questions that, unless I have overlooked something, are never answered in the text of *Splendeurs et misères des courtisanes* and yet they do not mark any gaps that are required to be filled in, given the literary functions of the novel" (*op. cit.*, 97). The argument that Kot'átko discusses appears to conclude from this that the entities singular terms in fiction refer to cannot be objects in the actual world, but weird incomplete entities: "Mme d'Espard is ... a creature whose gall bladder is such that in principle it cannot be diagnosed (or as a creature without gall bladder, or as a creature such that it neither possesses nor lacks a gall bladder" (*ibid*.). In response, Kot'átko presents a familiar compelling

rejoinder: the relevant indeterminacy is not so much a matter of "the world of the fiction" being constituted by so weird entities that it cannot possibly overlap with the actual world, but rather a matter of several different worlds constituting it: worlds in which Mme d'Espard's gall bladder is in good condition, and others in which it is not, etc. The reason, once more, has to do with the phenomenon of *importation*: "one can hardly deny that in the world of the *Splendeurs* the following conditional holds: If Mme d'Espard was ever thoroughly examined by the legendary Dr Bianchon (not to mention Dr Desplain!), ... he diagnosed the state of her gall bladder" (*ibid*.); as he rightly concludes, "the incomplete construction of a character is something very different from the construction of an incomplete character" (*op. cit.*, 99).

It should be clear that I have not argued for the view that Koťátko rightly criticizes here. The singular terms in a fiction like the ones he discusses refer in my view (albeit by description) to perfectly ordinary entities; they might in fact be, and frequently are, fellow inhabitants of the actual world. I have only argued for the view that such reference is just definite descriptions' denotation, belonging in Boër and Lycan's first grade of *de re* involvement. However, I should say that Bonomi (2008), whose classification of fictional discourses I have borrowed, and who advances a similar descriptivist view of the semantics of singular terms in textual discourse, in fact presents an argument for a claim that appears to come very close to the one that Koťátko rightly rejects. I do not think it is more successful than the sketchier ones Koťátko discusses.

Bonomi assumes an event semantics, on which utterances including verbs such as 'writing "Dublin"' posit an event. Events have a temporal and spatial extension. Events have proper parts, which have smaller temporal and spatial extensions. Thus, an event of writing 'Dublin' will include as a proper part an event of writing the initial d, whose temporal and spatial extension will be included in those of the former event. Now, Bonomi argues that, when indicated in a fiction, the temporal and spatial extension of the parts will typically be indeterminate, even if that of the event is explicitly indicated, as a temporal interval and a location in the actual world. Because of this, he concludes, "no real extension, in this world, can be attributed" to such events, and "strictly speaking, one cannot say that [they] occur" in the explicitly indicated temporal and spatial locations in the actual world (Bonomi 2008, 3). But I cannot see how that follows, for the reasons that Koťátko points out. As far as I can tell, it only follows that many worlds are

compatible with the fiction; but in any of them each part of the event has a unique location, and I cannot see how it can be excluded that the actual world is one of them. Bonomi acknowledges that the "downward indeterminacy" of temporal and spatial qualification on which his argument relies "is just an aspect of a more general phenomenon of indeterminacy" (*ibid.*), and in fact it is: there is no relevant difference between the indeterminacy of the location of parts of fictional events occurring in actual locations, and the indeterminacy in the state of Mme d'Espard's gall bladder. By the same token, nothing follows from it regarding the possibility of fictions being about (in some, lower or higher, grade of *de re* involvement) actual entities. This is therefore one more point of agreement with Petr's views. I have also learned from trying to articulate my objections to the ones regarding which I feel more critical.

References

BOËR, S.E. – LYCAN, W. (1986): *Knowing Who.* Cambridge, Mass.: MIT Press.

BONOMI, A. (2008): Fictional Contexts. In: Bouquet, P. – Serafini, L. – Thomason, R. (eds.): *Perspectives on Context.* Stanford: CSLI Publications, 213-248.

BRAUN, D. (1993): Empty Names. *Noûs* 29, 449-469.

BRAUN, D. (2005): Empty Names, Fictional Names, Mythical Names. *Noûs* 39, 596-631.

BYRNE, A. (1993): Truth in Fiction: The Story Continued. *Australasian Journal of Philosophy* 71, 24-35.

CRIMMINS, M. – PERRY, J. (1989): The Prince and the Phone Booth: Reporting Puzzling Beliefs. *Journal of Philosophy* 86, 685-711.

CURRIE, G. (1990): *The Nature of Fiction.* Cambridge: Cambridge University Press.

EVANS, G. (1982): *The Varieties of Reference.* Oxford: Clarendon Press.

EVERETT, A. (2005): Against Fictional Realism. *Journal of Philosophy* 102, 624-649.

EVERETT, A. (2007): Pretence, Existence, and Fictional Objects. *Philosophy and Phenomenological Research* 74, 56-80.

EVERETT, A. (2013): *The Nonexistent.* Oxford: Oxford University Press.

FINE, K. (2007): *Semantic Relationalisms.* Oxford: Blackwell.

FRIEND, S. (2000): Real People in Unreal Contexts. In: Everett A. – Hofweber, T. (eds.): *Empty Names, Fiction and the Puzzles of Non-Existence.* Stanford: CSLI, 183-203.

FRIEND, S. (2011): The Great Beetle Debate: A Study in Imagining with Proper Names. *Philosophical Studies* 153, 183-211.

FRIEND, S. (2014): Notions of Nothing. In: García-Carpintero, M. – Martí, G. (eds.): *Empty Representations: Reference and Non-Existence.* Oxford: Oxford University Press, 307-332.

García-Carpintero, M. (2006): Two-Dimensionalism: a Neo-Fregean Interpretation. In: García-Carpintero, M. – Macià, J. (eds.): *Two-Dimensional Semantics*. Oxford: Oxford University Press, 181-204.

García-Carpintero, M. (2007): Fiction-Making as an Illocutionary Act. *Journal of Aesthetics and Art Criticism* 65, 203-216.

García-Carpintero, M. (2008): Singular Thought and the Contingent A Priori. *Revue Internationale de Philosophie*, special issue "Philosophy of Mind" (edited by J. Proust), 62, 79-98.

García-Carpintero, M. (2010a): Fictional Singular Imaginings. In: Jeshion, R. (ed.): *New Essays on Singular Thought*. Oxford: Oxford University Press, 273-299.

García-Carpintero, M. (2010b): Fictional Entities, Theoretical Models and Figurative Truth. In: Frigg, R. – Hunter, M. (eds.): *Beyond Mimesis and Convention – Representation in Art and Science*. Springer, 139-168.

García-Carpintero, M. (2012): Foundational Semantics, II. *Philosophy Compass* 7, No. 96, 410-421.

García-Carpintero, M. (2013): Norms of Fiction-Making. *British Journal of Aesthetics* 53, 339-357, DOI: 10.1093/aesthj/ayt021.

García-Carpintero, M. (2014): Introduction: The Problem of Empty Representations. In: García-Carpintero, M. – Martí, G. (eds.): *Empty Representations: Reference and Non-Existence*. Oxford: Oxford University Press, 1-22.

García-Carpintero, M. (forthcoming): Pretend Reference and Coreference. In: Capone, A. – Lo Piparo, F. – Carapezza, M. (eds.): *Perspectives on the Pragmatics of Indirect Reports*. Perspectives in Pragmatics, Philosophy and Psychology Series. Dordrecht: Springer.

Hawthorne, J. – Manley, D. (2012): *The Reference Book*. Oxford: Oxford University Press.

Kaplan, D. (1989): Afterthoughts. In: Almog, J. – Perry, J. – Wettstein, H. (eds.): *Themes from Kaplan*. Oxford: Oxford University Press, 565-614.

Koťátko, P. (1998): Two Notions of Utterance Meaning. *Proceedings of the Aristotelian Society*, XCVIII, 225-239.

Koťátko, P. (2004): Truth Conditions and Literary Aspirations: Basic Correlations. In: Rossholm, G. (ed.): *Essays on Fiction and Perspective*. Frankfurt am Main: Peter Lang Verlag, 155-170.

Koťátko, P. (2010): Who Is Who in the Fictional Work. In: Koťátko, P. – Pokorný, M. – Sabates, M. (eds.): *Fictionality – Possibility – Reality*. Bratislava: aleph, 89-102.

Koťátko, P. (2013): Text, Work, Interpretation: Some Implications of the Menard Case. In: Koblížek, T. – Koťátko, P. – Pokorný, M. (eds.): *Text + Work: The Menard Case*. Prague: Litteraria Praguensia, 54-67.

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

Kroon, F. (1994): A Problem about Make-Believe. *Philosophical Studies* 75, 201-229.

Lamarque, P. – Olsen, S.H. (1994): *Truth, Fiction and Literature*. Oxford: Clarendon Press.

L EWIS, D. (1978): Truth in Fiction. *American Philosophical Quarterly* 15, 37-46. Reprinted with postscripts in: Lewis, D. *Philosophical Papers. Vol. 1.* Oxford: Oxford University Press, 1983, 261-280.

P ERRY, J. (2001): *Reference and Reflexivity.* Stanford: CSLI Publications.

P IETROSKI, P.M. (1996): Fregean Innocence. *Mind & Language* 11, 338-370.

S AINSBURY, M. (2005): *Reference without Referents.* Oxford: Clarendon Press.

S AINSBURY, M. (2010): *Fiction and Fictionalism.* London: Routledge.

S AINSBURY, M. (2014): Notions of Nothing. In: García-Carpintero, M. – Martí, G. (eds.): *Empty Representations: Reference and Non-Existence.* Oxford: Oxford University Press, 277-289.

S ALMON, N. (1998): Nonexistence. *Noûs* 32, 277-319.

S CHIFFER, S. (2003): *The Things We Mean.* Oxford: Clarendon Press.

S CHROETER, L. (2008): Why Be an Anti-Individualist? *Philosophy and Phenomenological Research* 77, 105-141.

S EARLE, J. (1969): *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

S EARLE, J. (1974): The Logical Status of Fictional Discourse. *New Literary History* 6, 319-332. Reprinted in: Searle, J. *Expression and Meaning.* Cambridge: Cambridge University Press, 1979, 58-75.

S OSA, E. (1970): Propositional Attitudes *De Dicto* and *De Re. The Journal of Philosophy* 67, 883-896.

S TALNAKER, R. (1978): Assertion. In: Cole, P. (ed.): *Syntax and Semantics* 9. New York: Academic Press, 315-332.

S TALNAKER, R. (2002): Common Ground. *Linguistics and Philosophy* 25, 701-721.

T HOMASSON, A. (1999): *Fiction and Metaphysics.* Cambridge: Cambridge University Press.

VAN I NWAGEN, P. (1977): Creatures of Fiction. *American Philosophical Quarterly* 14, 299-308.

W ALTON, K. (1990): *Mimesis and Make-Believe.* Cambridge, Mass.: Harvard University Press.

W EATHERSON, B. (2004): Morality, Fiction, and Possibility. *Philosophers' Imprint* 4.

W ILLIAMSON, T. (1996): Knowing and Asserting. *Philosophical Review* 105, 489-523.

Y ABLO, S. (2001): Go Figure: A Path through Fictionalism. In: French, P.A. – Wettstein, H.K. (eds.): *Midwest Studies in Philosophy.* Vol. xxv. Oxford: Blackwell, 72-102.

Y ABLO, S. (2014): *Aboutness.* Princeton: Princeton University Press.

Inadvertent Creation and Fictional Characters

Zsófia Zvolenszky

Department of Logic. Institute of Philosophy. Eötvös University
Múzeum krt. 4/i. H-1518 Budapest. Hungary
zvolenszky@elte.hu

ABSTRACT: In several papers, Petr Koťátko defends an "ontologically modest account of fictional characters". Consider a position (which I have been defending) that is anything but ontologically restrained: positing fictional characters like Andrei Bolkonsky in *War and Peace* as abstract artifacts. I will argue, first, that such a position turns out to offer a nice fit with Petr Koťátko's proposal about narrative fiction, one that fares better than an alternative pretense-based theory that doesn't posit Bolkonsky as existing in any sense. Second, I will explore a recent challenge by Jeffrey Goodman—which I will call the *inadvertent creation challenge*—that is originally posed for those who hold that fictional characters and mythical objects alike are abstract artifacts. The crux of the challenge is this: if we think that astronomers like Le Verrier, in mistakenly hypothesizing the planet Vulcan, inadvertently create an abstract artifact, then the "inadvertent creation" element turns out to be inescapable yet theoretically unattractive. Third, based on considerations about actually existing concrete objects being featured in fictional works (as Napoleon is in *War and Peace*), I argue that regardless of where one stands on mythical objects, admitting fictional characters as abstract artifacts is enough to give rise to the inadvertent creation challenge; yet this very set of considerations serves to undermine the challenge, indicating that inadvertent creation is not nearly as worrisome after all as Goodman is suggesting. Taking fictional characters (and mythical objects) to be abstract artifacts therefore remains a viable option.

KEYWORDS: Abstract artifacts – mythical objects – realism about fictional characters – semantics of fictional discourse.

1. Introduction

Back in 2005, when Petr Koťátko was putting on a conference in Prague to celebrate the 100th anniversary of *On Denoting*, he was half the age

of Russell's seminal paper; this was a fact I didn't know at the time, arriving at Villa Lana, meeting Petr for the first time. Since then, I have become a Prague "regular", returning every other year or so, following Petr's advice on Czech functionalist and cubist architecture and relishing his remarks and work on a broad range of philosophy of language themes. One of these is the semantics and metaphysics of fictional discourse, a topic we have both been thinking about for years.

Indeed, the last time we met was at a Prague aesthetics conference in 2013, where both of us presented on this very topic. In his conference talk, Petr urged, among other things, that we maintain an...

> ...ontologically modest account of fictional characters: for any text of fictional narration the characters spoken about in it are those persons whose existence in the actual world we have to assume (in the *as if* mode) in order to allow the text to fulfill its literary functions. In *War and Peace* we have to assume, in this sense, among others, Andrei Bolkonsky [a fictional character] and Napoleon [a historical figure featured in the novel] as inhabitants of the actual world. (Koťátko 2013, Section 8)

Meanwhile, in my talk (Zvolenszky 2013), I was giving new arguments for an *artifactualist position about fictional characters* (*artifactualism* for short)—made popular by Saul Kripke, Peter van Inwagen and Amie Thomasson—according to which our ontology for fictional discourse should make room for fictional characters as abstract artifacts: Andrei Bolkonsky is an actually existing abstract object whose existence is due to Tolstoy's having written *War and Peace*. My position certainly isn't one philosophers would call ontologically modest.

My aims in this paper are threefold. First, I will argue that on closer inspection, Petr Koťátko's position about the metaphysics of fictional characters is not nearly as distant from mine as it might first seem (Section 2). Indeed, an ontology that includes Bolkonsky as an abstract—in the sense of nonconcrete—actual object offers a nice fit with Petr's proposal about narrative fiction, one that, on balance, fares better than an alternative theory that doesn't posit Bolkonsky as an abstract artifact. Second, I will explore a recent challenge by Jeffrey Goodman (2014)—which I will call the *inadvertent creation challenge*—that is originally posed for those who hold that fictional characters and mythical objects alike are abstract artifacts (Section 3 and 4). The crux of the challenge is this: if we think that astronomers

like Le Verrier, in mistakenly hypothesizing the planet Vulcan, inadvertently create an abstract artifact, then the "inadvertent creation" element turns out to be inescapable yet theoretically unattractive. Third, based on considerations about actually existing concrete objects being featured in fictional works (as Napoleon is in *War and Peace*), I argue that regardless of where one stands on mythical objects, admitting fictional characters as abstract artifacts is enough to give rise to Goodman's challenge; yet this very set of considerations serves to undermine the challenge, indicating that inadvertent creation is not nearly as worrisome after all as Goodman is suggesting (Section 5).

So taking fictional characters (and mythical objects) to be abstract artifacts remains a viable option in the end, pending a certain worry I will raise but leave unanswered. Whether artifactualism about fictional characters is an option Petr Kot'átko would consider adopting is an issue I would love to discuss with him. But first things first: happy birthday, Petr.

2. Why regard fictional characters as abstract artifacts?

Proper names of fictional characters, like the name 'Andrei Bolkonsky' appear in various kinds of discourse:

(1)   "Just then another visitor entered the drawing room: Prince Andrei Bolkonsky.... He was a very handsome young man, of medium height, with firm, clear-cut features."
(2)   Andrei Bolkonsky was a proud man who has come to despise everything fake, shallow, or merely conventional.
(3)   Among the book's fictional characters, the reader's attention is first focused on Prince Andrei Bolkonsky.
(4)   Andrei Bolkonsky doesn't exist.

Here are some fairly uncontroversial observations about (1)–(4). (1) is quoted from Tolstoy's *War and Peace*. By writing (1), Tolstoy didn't aim to assert anything, and did not aim to refer at anything or anyone by means of the proper name; he merely *pretended* to make an assertion and *pretended* to refer to someone. (2) and (3) are sentences about *War and Peace* based on the Encyclopaedia Britannica's Bolkonsky-entry; both are assertions that feature the proper name as purporting to refer to whoever or whatever Andrei Bolkonsky is. While (2) is true within the fictional world of *War*

and Peace, it isn't true simpliciter. By contrast, (3) is true simpliciter. The same goes for (4).

Attaching labels to these four types of discourse will be helpful. In the table below, I list Amie Thomasson's labels which are more widely used than a more recent alternative set of labels that I myself prefer, by Bonomi (2008) (also favored by García-Carpintero 2014). Throughout the paper, I will be using the latter labels (in boldface).

According to Thomasson (2003)	*Thomasson's labels*	*Bonomi's labels*
(1) exemplifies discourse within works of fiction like *War and Peace*.	Fictionalizing discourse	**Textual** use
(2) exemplifies discourse by readers about the content of works of fiction.	Internal discourse	**Paratextual** use
(3) exemplifies discourse by readers and critics about the characters *as* fictional characters, the circumstances of their crea-tion, their historical relation to other literary figures, etc.	External discourse	**Metatextual** use
(4) exemplifies nonexistence claims.	**Nonexistence Claims**	–

Saul Kripke (1973/2013; 1973/2011) and Peter van Inwagen (1977) fa-mously argued that it is sentences like (3), metatextual uses, that motivate positing in our ontology fictional characters as objects that are nonconcrete (not located in space and time), that is to say, they are abstract. According to these philosophers, objects like Bolkonsky don't exist timelessly; rather, their existence is conditional upon natural languages like English featuring true assertions like (3) that purport to quantify over fictional characters. Clearly, on this view, what facilitates metatextual assertions is the existence of works of fiction, whose existence in turn is due to authors' creating those works. In this framework it is natural to think that the abstract ob-ject Andrei Bolkonsky's existence is due to Tolstoy's having written *War and Peace.* Those philosophers are artifactualist who hold that analyzing *some* of the discourse exemplified in (1)–(4) involves fictional characters as

abstract artifacts. Kripke and van Inwagen are both artifactualists then, as is Thomasson (1999); all three of them hold that metatextual uses require us to include fictional characters as abstract artifacts in our ontology.

How might we resist positing fictional characters as abstract artifacts? Kendall Walton (1990) developed an influential pretense-based account according to which "not just *some*, but *all* talk involving fictional names contains an element of pretense" (Thomasson 2003, 208). Call proponents of such views *pretense purist*. For textual uses like (1), the pretense is about the work of fiction, *War and Peace*, being true. For paratextual uses like (2), the same pretense is operative. For a metatextual use like "Bolkonsky is a fictional character", an *ad hoc* game of pretense is at play, a pretense according to which people come in two types: real people and fictional characters.

Thomasson, an artifactualist, provides two arguments against extending Walton's pretense-based proposal to metatextual uses (2003, 208-209).

First, such a proposal involves far more revision than what is minimally needed to resolve an apparent conflict like that involved in saying, under the same breath, (5) and (6):

(5)   Bolkonsky is a prince.
(6)   Bolkonsky is a fictional character.

According to van Inwagen and Kripke, and also Thomasson (1999), (5) (an instance of paratextual use) involves pretense, but (6) doesn't; instead it is a true assertion about an abstract artifact. This treatment of (6) takes the utterance at face value: (6) seems like a true assertion predicating something of Bolkonsky, and that is exactly what artifactualism delivers. By contrast, a pretense purist like Walton analyzes (6) in a way that involves a departure from face value appearance.

Moreover, for the pretense purist, the departure involves analyzing (6) as involving an *ad hoc* game of pretense or make-believe. This brings us to Thomasson's second argument: such analyses are at odds with how we view metatextual uses like (6). Thomasson illustrate this on an example:

> If two police officers discussing a case say "This is such a tough one, we need Sherlock Holmes to help us solve it", they do indeed seem engaged in a pretense that Holmes is a real detective who could be called upon in times of need. But the point of a humorless colleague's remark "There's no such person as Holmes, it's just a fictional character",

seems to be precisely to step *outside* of these forms of pretense and assert the real truth about Holmes. (Thomasson 2003, 209, emphasis in the original)

She then adds the following to drive home the point that appeal to pretense is generally at odds with our intuitions about what goes on in metatextual uses (see especially the underlined passage):

Indeed, a pure pretense theorist must take all literary historians' and critics' apparently serious claims about fictional characters, their origins, history, development, etc., to involve new, *ad hoc*, games of make-believe—whether these are claims that Shakespeare's character Hamlet was modeled on the 13th century character Amleth of Saxo Grammaticus' *Historia Danica*, that the play *Waiting for Godot* has five characters, or that if Arthur Conan Doyle's medical practice had been busier, the character of Sherlock Holmes might have never been created. Yet none of these seem, pre-theoretically, to involve pretense or games of make-believe, and such additional revisions are not necessary to prevent speakers from saying something self-contradictory or blatantly false, nor could the speakers normally be brought to recognize that they were invoking a pretense—so those grounds for attributing pretense to a piece of discourse do not apply here. (Thomasson 2003, 209, underlining added)

Of course, the strength of these two arguments depends on how artifactualists manage to analyze nonexistence claims like (4), how much revision their proposal requires. After all, Thomasson's two arguments hinge on the observation that when it comes to textual and paratextual uses, artifactualists can do just as well as pretense purists; meanwhile, when it comes to metatextual uses, artifactualists have the upper hand. But what if the advantage they gain there gets lost on nonexistence claims like (4)? Thomasson elsewhere (2009, 15) acknowledges this point: "The greatest difficulty for [artifactualist] views arises in handling ... nonexistence claims". She continues by summarizing some of the response strategies artifactualists have proposed:

denials that Sherlock Holmes exists may be read as denials that there is any such person (Thomasson 1999, 112), or any object answering the descriptions in the stories (van Inwagen 2003, 146). Alternatively, these nonexistence claims may be read as noting that past users of the name

mistakenly supposed that the name-use chain led back to a baptism ra-
ther than a work of fiction (van Inwagen 2003, 146-7; cf. Thomasson
2003). If some such solution to the problem of nonexistence claims can
be shown to be plausible and non *ad hoc*, [artifactualist] theories may
offer the best overall way to handle fictional discourse—a way which
does require positing fictional entities. (Thomasson 2009, 15)

It is well to note also that it is far from obvious that for pretense pur-
ists, handling nonexistence claims is a breeze. Thomasson (2003, 210, n6)
seems to have doubts about Walton's prospects in particular:

pretense theorists also owe us an account of how [nonexistence claims]
can be true, given that the name involved fails to refer. Walton's me-
thod is to treat sentences [such as (4)] as first invoking a pretense to re-
fer, and then (with 'doesn't exist') betraying that pretense. (Walton
1990, 422)

Beyond these remarks, I will not compare how well (with how little re-
vision) artifactualists versus pretense purists can handle nonexistence
claims. Instead, I would like to focus on metatextual uses, and what an op-
timal way to analyze them would be in the context of Petr Kot'átko's (2010,
2013) proposals about fictional characters. I will argue that he would do
well to analyze metatextual uses as involving reference to abstract artifacts.
But beforehand, let us consider briefly why, if someone contemplates being
a *realist about fictional characters*—making room for fictional characters in
her ontology—then her best choice is becoming an artifactualist rather
than subscribing to some other -ism. Two motivations are worth noting.

The first consideration for favoring artifactualism over its realist coun-
terparts: apart from artifactualist theories, all realist alternatives (for exam-
ple, Meinongianism, according to which fictional characters don't exist but
are concrete things that have being; and nonactualism, according to which
fictional characters are concrete existents that are nonactual) have trouble
accounting for the truth of metatextual uses like (3) and (6) (Thomasson
2009, 13).

The second consideration for favoring artifactualism over its realist
counterparts: according to Sainsbury (2010, 61–63, 82–85), the real advan-
tage of artifactualism concerns its ability to respond to the so-called *selec-
tion problem*: upon introducing the name 'Andrei Bolkonsky' in his novel,
how does Tolstoy manage to select one rather than another among the

countless candidate objects? According to Meinongianism, there are count-less nonexistent candidates; according to nonactualism, there are countless merely possible, nonactual candidates. Sainsbury (2010, 63) doesn't see "how a Meinongian can offer any sensible account of how an author's or reader's thoughts are supposed to engage with one rather than another nonexistent entity".[1]

We are now in a position to see why Petr's views about fiction and my artifactualism aren't incompatible; his work concerns textual uses only; it is in this context that he calls for an "ontologically modest account of fiction-al characters". His (2010) paper aims to show that "singular terms used in *texts of fiction* may even there fulfill the referential functions they have ac-quired in 'ordinary' communication" (89, emphasis added). Accordingly, Petr's focus is on the phenomenon of actual-world individuals, like Na-poleon being featured in texts of fiction, that is, in textual uses. His (2013) presentation explores what interpreting a text of narrative fiction requires. His thesis /F/ (Section 4, emphasis in the original) is that "The literary functions of a text of narrative fiction require from the interpreter that she approaches, in the *as if* mode, its sentences as records of utterances of an inhabitant of the *actual world*: the narrator, who tells us what happened *in this world*". Again, what is at issue is interpreting texts of narrative fiction, that is to say, textual discourse.

Meanwhile, artifactualists like Kripke, van Inwagen, Thomasson and myself concede that as far as the ontological needs of *textual* discourse are concerned, we could do without recourse to fictional characters as abstract artifacts; it is *metatextual* discourse that makes commitment to such enti-ties inescapable. Further, in the light of considerations in this section about the prospects of handling metafictional uses like (3), Petr could do worse than become an artifactualist (by becoming a pretense purist instead). If he opted for artifactualism, he, like I, would have to confront a challenge about inadvertent creation in the context of mythical artifactualism, a topic to which we'll now turn.

---

[1] Elsewhere (Zvolenszky 2012, Section 2) I argue that a more decisive objection emerges against the Meinongian once we consider the difficulties that the *nonactualist* encounters when it comes to the selection problem and other problems.

3. Why resist mythical artifactualism?

In a recent paper Goodman (2014) poses a challenge for proponents of a view I'll call *mythical artifactualism*, according to which some objects ('mythical objects') that figure in false theories (or 'myths') are abstract artifacts like artifactualists' fictional characters. Commonly cited examples of mythical objects are phlogiston (which in a now-discarded theory was featured as accounting for rusting and burning) and Le Verrier's Vulcan (a hypothetical planet that in a now-discarded theory was featured as causing perturbations in Mercury's orbit) (Goodman 2014, 35).

Van Inwagen's (1977) influential argument for realism about fictional characters has it that metatextual sentences like (3) and (6) ("Bolkonsky is a fictional character") are straightforwardly true, a phenomenon whose accommodation requires including Bolkonsky in our ontology; and the best we can do is make him a nonconcrete (that is, abstract) object.

Mythical artifactualists (like Braun 2005; Kripke 1973/2013; and Salmon 1998) can readily appeal to a parallel argument based on (7) or (8):

(7)   Vulcan is a mythical planet.
(8)   Vulcan is a hypothetical planet.

Given that both (7) and (8) are straightforwardly true, a phenomenon whose accommodation requires including Vulcan in our ontology, the best we can do is make it a nonconcrete (that is, abstract) object. So the argument goes for mythical artifactualism.

Goodman (2014, 36) points out that the above argument doesn't yet establish that Tolstoy's fiction-writing brought the abstract object Bolkonsky into existence or that Le Verrier's theorizing brought the abstract object Vulcan into existence. After all, the *intentions* of Tolstoy and Le Verrier were markedly different: the first (according to artifactualism) aimed at creating Bolkonsky and didn't aim at describing reality, while the second aimed at describing reality and a mind-independently existing celestial body (but failed) and didn't aim at describing an abstract object. Why believe in inadvertent creation then, as mythical artifactualists do? Here is why: despite the differences is in some of Le Verrier's and Tolstoy's intentions,

> each had intentions and performed activities that were sufficient to bring about an abstract object; Le Verrier *inadvertently* brought about Vulcan given his particular sort of intentions and activities, while [Tols-

toy] in some sense aimed to create [Bolkonsky]. (Goodman 2014, 30, emphasis in the original)

Goodman quotes Braun (2005, 615) making this point:

> The activities that occur during mistaken theorizing, such as le Verrier's, are importantly similar to those that occur during storytelling. In both, names are used and predicative sentences containing them are formulated. Reasoning and other mental processes occur. Texts that are seemingly susceptible to evaluation for truth are produced. Thus, if story-tellers' activities create fictional characters, then mistaken theorizers' activities create abstract objects of a similar sort. So I grant that Le Verrier's mistaken theorizing creates an abstract artifact.

We are now in a position to appreciate Goodman's (2014, 37-38) *argument against mythical artifactualism* (he acknowledges Phillips 2001 as his key inspiration):

Premise 1: If Vulcan is a created abstractum (like Bolkonsky), then Vulcan is created by Le Verrier in every possible world where Le Verrier performs relevantly similar activities to those he actually performed.

Premise 2: There is a possible world where Le Verrier performs relevantly similar activities to those he actually performed and yet fails to create Vulcan.

Conclusion (by *modus tollens*): Vulcan is not a created abstractum (like Bolkonsky).

The argument is valid; its conclusion is true if its premises are.

To support Premise 1, Goodman invites us to consider intention-identical possible worlds in which the intentional, authorial activities of Tolstoy and Le Verrier are the same as in the actual world. Plausibly, if Tolstoy's actual authorial activities suffice to create Bolkonsky in the actual world, they also suffice to create Bolkonsky in intention-identical worlds that are nonactual. And since the mythical artifactualist maintains an analogy between fictional objects and mythical ones, she would find Premise 1 similarly plausible, Goodman argues.

To support Premise 2, Goodman singles out among the intention-identical possible worlds those that contain both of these individuals: a flesh-and-blood person qualitatively identical to *War and Peace*'s Bolkonsky on the one hand, and a planet fitting the description of the hypothetical pla-

net Vulcan on the other hand. Call these X-worlds. With respect to X-worlds, our intuitions about object creation come apart in the fictional and mythical cases; here is why. Kripke's (1972/1980) arguments in the "Addenda" to *Naming and Necessity* are considered overwhelmingly convincing: if Tolstoy's intention in using the name 'Bolkonsky' was not to write about a real person but a fictional one, then he didn't accidentally, inadvertently write about a concrete, flesh-and-blood individual who happens to be a dead ringer for the Bolkonsky of the novel.[2] This way, the Tolstoy of X-worlds still creates a fictional character, Bolkonsky, if the actual Tolstoy does. But an intuitive assessment of X-worlds with respect to the status of an abstract artifact Vulcan are markedly different: the Le Verrier of an X-world, in formulating (what in the X-world is) a true scientific theory, manages to name the concrete planet that is a dead ringer for Vulcan; it seems outlandish to think that the X-world-inhabiting Le Verrier has created anything abstract. Such an X-world suffices to make Premise 2 true.

If we accept Premise 1 and Premise 2, then in the light of the arguments against mythical artifactualism, we have reason to give up on that theory, even if we maintain our sympathies with artifactualism about fictional characters, a view that does not at this point seem affected by Goodman's argument. In what follows, I will show that appearances are misleading: artifactualism about fictional characters is affected by Goodman's argument after all (Section 5), but this very fact casts doubt on a worry (discussed in Section 4) that Goodman formulated against the following move: denying Premise 1.

4. Why worry about the inadvertent creation challenge?

Goodman (2014, 39) considers and then rejects one strategy for resisting his anti-mythical-artifactualism argument: denying Premise 1. And the reason for his rejection is the inadvertent creation challenge. Let's explore the denying-Premise-1 strategy and the challenge.

The denying-Premise-1 strategy hinges on the observation that there is something special to mythical object creation even if we think there is a tight analogy between Le Verrier's creating Vulcan and Tolstoy's creating

---

[2]   I gave a detailed analysis of this argument of Kripke's as well as related ones in Section 2 of Zvolenszky (2012).

Bolkonsky. After all, when it comes to mythical objects (but not fictional ones)...

> ...[i]t is more than the author's intentions and social/historical context that counts in the abstractum-creation process; the mind-independent physical world must have its say as well. Whether or not one succeeds in creating a mythological object depends on whether the world obliges by providing the relevant entity (in which case, no creation occurs) or fails to oblige (in which case, creation occurs). (Goodman 2014, 39)

According to the denying-Premise-1 strategy, there is a difference between these two processes:

- *the fictional form of creation* "that is dependent solely on authorial intentions and historical/social contexts", and
- *the mythical form of creation* that is "dependent on these factors plus the non-cooperation of the world to provide the relevant entity". (Goodman 2014, 39)

But the denying-Premise-1 strategy, Goodman argues, creates a problem: the *inadvertent creation challenge*, underlined in the passage below.

> [The denying-Premise-1] move simply serves to shine the spotlight on what is so theoretically unattractive about the [second, mythical form of creation] sort of process: it requires that the creation of mythical objects be *inadvertent*. Unlike the situation with [Tolstoy] and his aims, Le Verrier wished to be the discoverer of a planet; according to mythical [artifactualism], he wound up creating Vulcan instead. Now, while it's common to find cases of inadvertent discovery, it's at least unusual to find cases of inadvertent creation. Creation normally involves having a goal that one aims to achieve. ... it is odd to think that there are objects that are produced via a process of scientific theorizing that utterly divorces their production from the desires of the theorist to *not* be creative. **If there were a view that would provide the same theoretical benefits ... as mythical [artifactualism] yet avoid this theoretical oddness, it would be clearly preferable** (Goodman, 2014, 39; italics in the original, underlining and boldfacing have been added)

I agree with Goodman that the denying-Premise-1 strategy brings on the phenomenon of inadvertent creation. But I disagree with the boldfaced excerpt on two counts:

(A)   The inadvertent creation *phenomenon* (as I prefer to call it) is not specific to mythical artifactualism; we have to contend with it even if we are artifactualists about fictional characters who don't take a stand on the ontological status of mythical objects.

(B)   The very ubiquity of the inadvertent creation phenomenon calls into question just how worrisome it is. In fact, it is not at all clear that avoiding the phenomenon is theoretically preferable to not avoiding it.

In the remainder of this paper, I will give reasons for (A) and (B).

5. Why resist worrying about the inadvertent creation phenomenon?

We'll see that Napoleon-related details highlight the plausibility of (A) and (B). At the beginning of this paper, I quoted Petr Koťátko discussing a case in which a text of narrative fiction mentions a fictional character like Bolkonsky and also a historical figure like Napoleon. Petr goes on to say that in the latter case, interpreters of the text are not only called upon to assume, within the scope of the *as if* operator, that Napoleon is an inhabitant of the actual world, but are also called upon to assume...

> ... the existence of [Napoleon] outside the scope of the *as if* operator and ... this assumption as well as my ability to exploit my (rather limited) knowledge about Napoleon's career may belong to the capacities required from the reader by the literary functions of the text. (Koťátko 2013, Section 8)

In earlier work, Petr subscribed to the plausible view (also held by Thomasson 1999) that names like 'Napoleon' in their textual uses (as in *War and Peace*) be interpreted as referring to the historical figure (Koťátko 2010, 96). I suggest we take this view on board and see what it reveals with respect to (A).

Imagine the following scenario T: while writing *War and Peace,* Tolstoy was under the mistaken impression that Bolkonsky, like Napoleon, was a real person. Introducing the name 'Andrei Bolkonsky' he intended to refer to a historical figure he thought existed. For the artifactualist about fictional characters, what follows from the fact that (in the imagined scenario)

Tolstoy was wrong? Quite independently of what the artifactualist thinks about the status of objects of myth, it is overwhelmingly plausible to think that in the imagined scenario, Tolstoy created Bolkonsky as an abstract artifact, and did so inadvertently. And the reason why he did so is because of the non-cooperation of the world to provide the relevant entity.

This way, if we think about it, the kind of creation process that Goodman considers odd because of the inadvertent creation detail is one that is plausibly very common among those who create works of fiction: to give a sense of how ordinary such a scenario is, imagine a little boy, Sam, who is convinced that the Disney castle (the Magic Kingdom) depicted in a poster that hangs on his wall houses a mouse who steps out of the castle at bedtime. Sam tells a story (a work of fiction) about the mouse, intending to refer to a real mouse that he thinks steps out of the poster at night and embarks on some fictional adventures during which it encounters fictional characters. In fact, the mouse doesn't exist. It's a creature of Sam's imagination, we might say. But an artifactualist about fictional characters who remains noncommittal about the status of creatures of the imagination[3] would find it extremely plausible that the mouse is an abstract artifact— a fictional character—who is Sam's creation; crucially, the mouse was *inadvertently* created by Sam. If we were to study his intentions with respect to developing his story, we would find that with respect to the mouse, his desire is not to be creative: he aims to tell a story that features what he thinks is a real, concrete mouse the same way Tolstoy, in writing *War and Peace*, aimed to tell a story that featured someone he thought was and who in fact was a historical figure called Napoleon.

Scenario T and the Sam example demonstrate that:

(a)   the inadvertent creation phenomenon is not specific to mythical artifactualism;
(b)   the phenomenon is already present if we assume artifactualism about fictional characters;
(c)   moreover, the phenomenon is rather commonplace, due to mundane instances of error on the part of the creator of the work of fiction.

---

[3]   On creatures of the imagination and some of the issues that parallel those about creatures of myth, see Salmon (1998) and Caplan (2004).

(b) already shows that the strategy Goodman is converging on—in slogan form: "let's avoid the inadvertent creation challenge by steering clear of mythical artifactualism"—cannot work. After all, artifactualism about fictional characters by itself invites the challenge already (inasmuch as it is a challenge). (a) and (b) together demonstrate (A). And (c), about the phenomenon being commonplace, questions just how worrisome it is to be confronted by the inadvertent creation phenomenon, providing reason to accept (B).

I realize that my observations (a)–(c) can be readily turned upside down and construed as a new set of reasons for resisting artifactualism about fictional characters—a view that I favor and that I recommended for Petr Kot'átko in Section 2 above—and for adopting an alternative account like the pretense purist theory instead. Responding to this turning of the tables is a task that I hope one of us—Petr or I—will tackle in the near future. I have an idea about how to go about it, but will leave it as a surprise for another (not too distant) birthday of Petr's. But quite independently of such an argument, the Sam example (and similar examples) featuring inadvertent creation in the context of fiction telling already indicate that probably, a widespread conception echoed by Goodman (2014, 39)—namely, "creation-by-fiction-telling... is dependent solely on authorial intentions and historical/social contexts"—is a myth.[4]

References

BONOMI, A. (2008): Fictional Contexts. In: Bouquest, P. – Serafini, L. – Thomason, R. (eds.): *Perspectives on Context.* Stanford: CSLI Publications, 213-248.

BRAUN, D. (2005): Empty Names, Fictional Names, Mythical Names. *Noûs* 39, 596-631.

CAPLAN, B. (2004): Creatures of Fiction, Myth, and Imagination. *American Philosophical Quarterly* 41, 331-337.

GARCÍA-CARPINTERO, M. (2014): Introduction: The Problem of Empty Representations. In: García-Carpintero, M. – Martí, G. (eds.): *Empty Representations: Reference and Non-Existence.* Oxford: Oxford University Press, 1-24.

Goodman, J. (2014): Creatures of Fiction, Objects of Myth. *Analysis* 74, 35-40.

Koťátko, P. (2010): Who is Who in the Fictional World. In: Koťátko, P. – Pokorný, M. – Sabatés, M. (eds.): *Fictionality – Possibility – Reality*. Bratislava: aleph, 89-100.

Koťátko, P. (2013): Fiction, Reality and Radical Narration (in Czech). *Organon F* 20, 72-96.

Kripke, S. (1972/1980): Addenda to "Naming and Necessity". In: *Naming and Necessity*. Cambridge, Massachusetts: Harvard University Press, 156-164.

Kripke, S. (1973/2011): Vacuous Names and Fictional Entities. In: *Philosophical Troubles: Collected Papers, Volume 1*. New York: Oxford University Press, 52-74.

Kripke, S. (1973/2013): *Reference and Existence: The John Locke Lectures for 1973*. New York: Oxford University Press.

Phillips, M. (2001): What a Solution to the Problem of Empty Names Cannot Be. Presented at the 2001 Pacific Division meeting of the American Philosophical Association.

Sainsbury, M. (2010): *Fiction and Fictionalism*. London: Routledge.

Salmon, N. (1998): Nonexistence. *Noûs* 32, 277-319.

Thomasson, A.L. (1999): *Fiction and Metaphysics*. Cambridge: Cambridge University Press.

Thomasson, A.L. (2003): Speaking of Fictional Characters. *Dialectica* 57, 205-223.

Thomasson, A.L. (2009): Fictional Entities. In: Kim, J. – Sosa, E. – Rosenkrantz, G. (eds.): *A Companion to Metaphysics*, 2nd edition. Oxford: Blackwell, 10-18.

van Inwagen, P. (1977): Creatures of Fiction. *American Philosophical Quarterly* 14, 299-308.

van Inwagen, P. (2003): Existence, Ontological Commitment, and Fictional Entities. In: Loux, M. – Zimmerman, D. (eds.): *The Oxford Handbook of Metaphysics*. Oxford: Oxford University Press.

Walton, K.L. (1990): *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.

Zvolenszky, Z. (2012): Against Sainsbury's Irrealism about Fictional Characters: Harry Potter as an Abstract Artifact. *Hungarian Philosophical Review (Magyar Filozófiai Szemle)*, No. 4, special issue: Realism (ed. by Bács, G., Kovács, D.M., Tőzsér, J.), 83-109.

Zvolenszky, Z. (2013). Abstract Artifact Theory about Fictional Characters Defended —Why Sainsbury's Category-Mistake Objection is Mistaken. In: Dorsch, F. – Ratiu, D.-E. (eds.): *Proceedings of the European Society for Aesthetics* 5, 597-612.