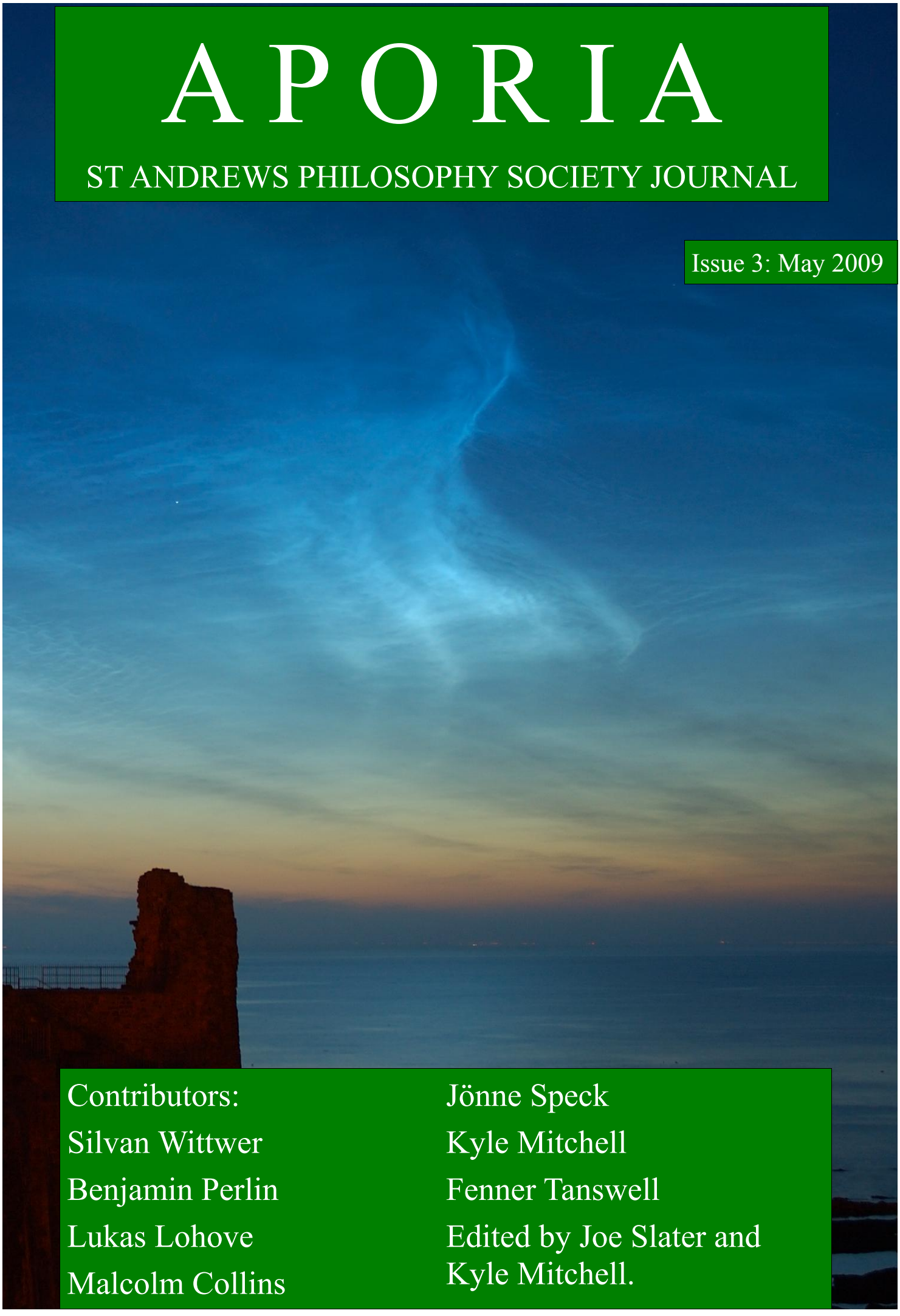
APORIA

ST ANDREWS PHILOSOPHY SOCIETY JOURNAL

Issue 3: May 2009



Contributors:

Silvan Wittwer

Benjamin Perlin

Lukas Lohove

Malcolm Collins

Jönne Speck

Kyle Mitchell

Fenner Tanswell

Edited by Joe Slater and Kyle Mitchell.

Welcome

Welcome to the third issue of *Aporia*, the St Andrews University Philosophy Society journal! We are thrilled to present this second issue of the academic year 08/09, and delighted with the high quality of submissions received. The papers in this issue cover a wide range of issues (despite all being contained in just one… how deliciously absurd!), from epistemology to ethics and from the material conditional to conceptual analysis. We believe there is something for everyone here, and hope you will enjoy reading through our latest pride and joy. At this point we'd like to express our undiminished admiration, amazement and gratitude to our hard-working journal editing team, all the submitting authors, our tireless committee, our immensely generous department and, last not least, you, the readers of this latest issue of *Aporia*.

As we did last issue, we present some warm-up questions, just to get you in a philosophical mood. (As last time, we welcome pretty postcards with answers, be they subtly clever or obviously true):

1) We all know that in exams the only food and drink allowed is a bottle of water. Would a bottle of specially imported Twin-Earth XYZ be allowed?
2) Are your hands closer to you than your feet?
3) What, if anything is wrong with the following argument?

> Vanilla ice-cream is tastier than steak ice-cream.
> Therefore vanilla is tastier than steak.

There: now you're ready to get stuck in! See you on the other side,

Jael and Duncan,

President and Vice-President of the Society.

Credits

Special thanks to Simon Prosser for contributing the cover photograph.
Design and Layout – Joe Slater.

University of St Andrews Philosophy Society Committee:

Jael Kriener – President
Duncan Reynolds – Vice President
Lukas Lohove – Treasurer
Benjamin Hofmann – Secretary
Fenner Tanswell – External Speakers Coordinator
Sarah Lohmann – Debates Coordinator
Joe Slater – Journal Editor
Kyle Mitchell – Journal Editor
Malcolm Collins – Publicity Officer

Also, thanks to all contributors and all who have supported Aporia.

Contents

Epistemic Contextualism and Error Theory

Silvan Wittwer

1. Introduction

In this essay, I argue that Schiffer's error-theoretical objection against epistemic contextualism (EC) does not hold, that 'know(s)' is context-sensitive and that there is a potential error theory for epistemic contextualism.

The argument unfolds in two parts: after some introductory remarks (sections 2&3), I first critically assess the recent discussion of Schiffer's error-theoretical objection (section 4) and show that it rests on a confusion that can be avoided by making a previously unstated distinction (section 5). In the second part, I showcase a model for the context-sensitivity of 'know(s)' (section 6) and sketch out a pragmatic approach to the error theory required by EC (section 7).

2. Epistemic Contextualism: everyday cases and skeptical puzzles

EC is the semantic thesis that 'know(s)' is a context-sensitive term.[1] Thus, the content or the truth-conditions of knowledge attributions ('S knows that p', S being some subject, p some proposition) vary across contexts. What is true in a low standard context (e.g. pub), may be false in a high standard context (e.g. philosopher's conference). The epistemic standards are determined by the context. The context is conceived as the context of utterance and refers to features of the attributor's psychology.[2]

EC is primarily motivated by everyday cases that involve so-called 'shifty data'.[3] Let's consider an example:

A: I know that is a zebra.
B: But can you rule out its being a cleverly painted mule?
A: I guess I can't rule that out.
B: So you admit that you don't know that's a zebra, and so you were wrong earlier?
A: Oh, c'mon. I didn't say I know it's a zebra.[4]

The last sentence strikes us as blatantly false because it contradicts sentence one. EC prevents A from saying something contradictory by claiming that 'know(s)' is context-sensitive and that there is a context shift in the course of the conversation. The standards in play at the beginning of the dialogue are not the same as those in the end. Since the standards rise, what is true in the beginning may be false in the end. A is not to blame: she is simply *ignorant* of the context-sensitivity of 'know(s)'. There is no contradiction, since the first and the last sentence have different contents. She needn't retract her initial statement.

Furthermore, EC claims that skeptical arguments resemble the example above.[5] Hence, the puzzles they generate can be analyzed and resolved analogically.[6]

---

1 I will use the terms 'context-sensitivity' and 'indexicality' interchangeably. I know that this is controversial, cf. MacFarlane 2007

2 I refer to a generic account of EC that does not distinguish between content and truth context-sensitivity. Moreover, Rysiew 2007 writes that 'context' may additionally refer to the 'conversational-practical situation'. The literature I used for this essay does not do that. So I won't conceive 'context' in that fashion.

3 DeRose 1999: 194

4 Blome-Tillmann 2008: 33

5 EC has focused on arguments for external world skepticism. However, Neta 2003: 398 points out that probably all skeptical templates prey on the context-sensitivity of 'know(s)'.

6 See DeRose 1999 for a detailed account.

3. Schiffer's error-theoretical objection

EC's solution to the skeptical puzzle does not come cheap, though. To establish the positive claim that 'know(s)' is a context-sensitive term, EC has to subscribe to a negative claim. The negative claim consists in an error theory that explains why competent speakers systematically fail to recognize the context-sensitivity of 'know(s)' and get puzzled by the skeptic's argument.

Basically, EC's error theory postulates that speakers get 'bamboozled by [their] own words'.[7] Put differently, speakers are afflicted by some sort of 'semantic blindness'[8] to the context-sensitivity of 'know(s)'.

EC's error theory seems hardly satisfactory, though. As Stephen Schiffer (1996) objects, we do not fail to detect the context-sensitivity of ordinary context-sensitive terms, say indexicals like demonstratives (e.g. 'that').[9] Thus, context-sensitivity may be assumed to be a *transparent* semantic feature. So, EC's claim that there is *hidden* context-sensitivity contradicts this linguistic data.

However, Schiffer's error-theoretical objection is in need of refinement. For it remains unclear against what the objection is directed. Two different kinds of speaker's ignorance might be targeted: either speakers being ignorant of the *content* of their utterances or them being ignorant of what their *communicative intentions* are.[10]

4. How to get it wrong: the problems of inaccessible content and mistaken intention

The specification reduces Schiffer's objection to two problems: the problem of inaccessible content (that speakers cannot know what propositions they express) and the problem of mistaken intention (that speakers are mistaken about their own communicative intentions).

A good deal of the contributions to the debate have focused on resolving one or both of these problems in order to refute or invigorate Schiffer's objection. In this section, I critically assess three contributions and show how they fail to resolve the problems. Their failure indicates that they somehow misconceive Schiffer's error-theoretical objection. The misconception will be specified in section five.

Thomas Hofweber (1999) sets out to devise a model for the context-sensitivity of 'know(s)'. Thereby, he adopts a rough propositional approach to sentences or utterances that features unarticulated constituents. Unarticulated constituents are conceived as functional parts of the proposition that do not appear explicitly at the sentential level. In fact, they commonly occur in cases of *implicit relativity*. Adjectives like 'tall' exhibit implicit relativity, since they have an unarticulated constituent which refers to a comparison class. Although implicit, the reference is cognitively accessible to the speaker.[11]

A second type of unarticulated constituents lacks this property, though. Hofweber calls it *hidden relativity* and considers it to account for the hidden context-sensitivity of 'know(s)' advocated by EC.

Hofweber offers the following example for hidden relativity: we often utter sentences like 'my car moves at 25 mph', treating physical motion as an absolute property. Thanks to recent discoveries in physics, however, we know that motion is a relative property. The motion of an object can be measured only in relation to some framework of reference. Thus, the sentence features an unarticulated constituent we are unaware of.[12]

7 Schiffer 1996: 329
8 Metaphor coined by Hawthorne 2004: 107
9 Schiffer 1996: 326f.
10 Rysiew 2001: 483
11 Hofweber 1999: 4
12 Hofweber 1999: 10f.

Unfortunately for EC, there is dissimilarity between this instance of hidden relativity and the 'hidden relativity' allegedly at work in knowledge attributions. In the case of motion, the unarticulated constituent is *invariant*, the framework of reference being some commonsensical understanding of motion as an absolute property. Although speakers cannot strictly speaking access the content of their utterances, it does not matter, since the unarticulated constituent unknown to almost everyone is also the same for everyone.[13] Were hidden relativity an adequate model for the context-sensitivity of 'know(s)', the situation would differ: since the unarticulated constituent is some *variant* feature of the context, the content changes across contexts. When it comes to accessing contents, speakers fail altogether. Consequently, they become unaware of sameness, difference and incompatibility of contents.[14] But obviously, that contradicts linguistic data and renders the hidden relativity approach to the context-sensitivity of 'know(s)' inadequate. Hofweber gets stuck with the problem of inaccessible content.

Unlike Hofweber, Patrick Rysiew (2001) does not particularly care about the problem of inaccessible content. He grants EC the inaccessibility of content, but emphasizes the implications: since content and context are closely tied on EC's account, denying accessibility of content implies that speakers are systematically mistaken about their communicative intentions as well. It simply proves impossible to hold track of the context if one loses the content.[15] Even if EC found a solution to the problem of inaccessible content, it would not prevent EC from falling prey to the problem of mistaken intentions.

Ram Neta (2003) faces the challenge set up by Rysiew. He tries to give a solution to the problem of mistaken intention by biting the bullet and admitting that we can be partially mistaken about our communicative intentions. Moreover, this does not harm our communication capacity, as Rysiew suggests.

To make sense of Neta's argument, we should return to Hofweber. As we have seen, there is a significant difference between Hofweber's hidden relativity and the hidden context-sensitivity of 'know(s)'. One might wonder, however, whether Hofweber's hidden relativity case really is that unproblematic. After all, it contains inaccessible contents. And according to Schiffer, it is a general truth about language that the content of an utterance has to be backed up by speaker's communicative intentions.[16] But how can you back up a content you cannot access? It seems to end in mistaken intentions, regardless of the unarticulated constituents being invariant.

Hofweber's response to this problem is the application of his propositional model of unarticulated constituents to mental states, such as communicative intentions. Since the mental unarticulated constituent is invariant in genuine instances of hidden relativity, no further problems whatsoever arise. Or so he argues.[17]

Basically, Neta gives a Hofweberian theory of unarticulated mental constituents for *contextual features*. He argues that there is indeed evidence for some unarticulated constituent on the mental level.[18] But unlike in Hofweber's application, the unarticulated constituent is a variant, contextual feature, namely some communicative intention, since context for EC is the attributor's psychology. Put differently, there are communicative intentions we can be mistaken about, but that does not harm EC's case for the context-sensitivity of 'know(s)'! Let's have a closer look at how Neta establishes the first claim – and why we need not bother having a closer look at the second.

---

13 Hofweber's example is far from uncontroversial. Let's grant it for the sake of the argument.

14 Hofweber 1999: 16

15 Rysiew 2001: 485

16 Hofweber 1999: 8f.

17 Hofweber 1999: 14. It is irrelevant whether Hofweber's application really works. I need it only to establish the claim (introduced below) that Neta pursues the same line of argument.

18 Neta 2003: 404f.

Firstly, Neta claims that there are communicative intentions we are unaware of and, hence, can be mistaken about. To illustrate his point, he considers a situation of disagreement that allegedly exhibits the confusion found in skeptical puzzles:

"Two people who think they are in the same room but are in fact in different rooms [and] are talking to each other over an intercom [will] mean something different by 'this room' when one claims 'Frank is not in this room' and the other insists 'Frank is in this room – I can see him!' "[19]

According to Neta, the two people are at the same time mistaken and not mistaken about their respective communicative intentions. On the one hand, they mistakenly believe their communicative intentions to be directed at an incompatible content, although the content cannot be truly incompatible. Incompatibility presupposes sameness of content, which is not given in this case, because the demonstrative 'this' gets assigned a different contextual value for each speaker. On the other hand, they are not mistaken about their communicative intentions, since '...each knows *something* about her own communicative intentions, but she doesn't know *the whole truth* about her own communicative intentions. Specifically, she doesn't know what inferential relation her own intended content bears to the other's intended content.'[20] This specific ignorance results from the ignorance about the non-semantic fact that both speakers are not in the same room.[21] Conversely, it is the ignorance about the non-semantic fact that results in a *partial* ignorance about one's own communicative intentions. Thus, we can be mistaken about communicative intentions.

Secondly, Neta claims that partial ignorance does not harm EC's case for the context-sensitivity of 'know(s)'. Thereby, he devises an argument for minimal conversational rationality: we do not need total access to our communicative intentions in order to participate rationally in conversation.

However, we need not evaluate this second claim, since Neta's first claim fails to be consistent. In fact, I think it is essentially flawed when it treats the self-ignorance featured in the Frank-case as related to the semantic blindness afflicting speakers in skeptical cases. Here is why:[22]

In the Frank-case it is an ignored non-semantic feature of the context that leaves the two interlocutors puzzled (and explains their confusion to us). It is *not* the partial ignorance of communicative intentions in the first place. Rather, the ignorance of the non-semantic fact induces the partial confusion about the communicative intentions. Analogically, in the skeptical case, Neta could not postulate partially mistaken communicative intentions (induced by the context-sensitivity of 'know(s)' and the ignorance of the context alone) and go on to launch an argument for minimal conversational rationality. He could not do it without introducing some non-semantic fact first.

On closer examination, Neta commits a fallacy of equivocation: in the Frank-case, the 'context' (we are partially ignorant of) encompasses a non-semantic or non-psychological fact (that the two persons are located in different rooms), whereas the 'context' in the skeptical case is supposed to be a much narrower notion, merely including the attributor's psychology.

Since Neta cannot apply the solution worked out for the Frank-case to the skeptical case, his argument breaks down between the claim of partially mistaken intention and the argument for minimal conversational rationality. His failure renders the whole hidden relativity approach implausible at last.

19 Neta 2003: 400. The example was originally devised by DeRose 1992. Rysiew 2001 comments on it to expound the problem of mistaken intention.

20 Neta 2003: 405

21 Neta 2003: 406

22 For brevity's sake, I cannot discuss Neta's explanation of skeptical puzzles. This is not needed anyway: what is at stake is Neta's *application* of his solution to the Frank-case to skeptical puzzles.

5. How to get it right: the distinction between indexicality and intelligibility

The line of argument pursued by Hofweber, Rysiew and Neta fails because it misconceives Schiffer's error-theoretical objection. More precisely, it conflates the distinction between the indexicality of 'know(s)' and the intelligibility of this particular indexicality. Accordingly, two separate philosophical endeavors were run together: the quest for an adequate model of context-sensitivity for 'know(s)', and the pursuit of an explanation as to why the context-sensitivity of 'know(s)' remains unintelligible to us in skeptical cases.

I think that the conflation roots in Hofweber's notion of hidden relativity. For it unsuccessfully tries to explain the context-sensitivity of 'know(s)' by emphasizing its cognitive inaccessibility. Even Neta's much more sophisticated argument is pervaded by this idea that unintelligibility should somehow account for context-sensitivity.

Fortunately, the issues are not that closely tied. EC's positive and negative claim can be treated separately. Adjusting one does not mean to lose the other. A separate treatment might even be required in order to comprehensively explain knowledge attributions.

Therefore, on my reading, Schiffer's error-theoretical objection raises two questions which can be answered independently.

(i.) Which semantic model does best explain the context-sensitivity of 'know(s)'?
(ii.) Which pragmatic model governs the intelligibility of this particular context-sensitivity?

In the second part, I attempt to answer these questions by presenting Michael Blome-Tillmann's analysis of the indexicality of 'know(s)' and by putting forward some reflections on the pragmatics of knowledge attributions.

6. A model for the context-sensitivity of 'know(s)'

Michael Blome-Tillmann (2008) puts forward a model for the context-sensitivity of 'know(s)' that might be taken as a convincing answer to the first question. He claims that 'know(s)' is a linguistically exceptional term, for it features a special combination of semantic, syntactic and pragmatic properties. More precisely, 'know(s)' proves to be indexical and factive, non-gradable and functioning as the epistemic norm of assertion.[23] The unique nature of 'know(s)' would also partly explain the difficulties we face in detecting its context-sensitivity.

For reasons of brevity, I will just present his argument for the indexicality of 'know(s)' which coincides (not coincidentally) with the refutation of the error-theoretical objection. Two further objections that establish the properties of non-gradability as well as factivity and normativity of assertion, respectively, cannot be addressed here. Moreover, the argument for the indexicality of 'know(s)' will not be assessed critically.

The aim of this section is to showcase one specific feature of a recent indexicalist approach which can accommodate most of the criticism directed at EC so far.

Blome-Tillmann's argument for the indexicality of 'know(s)' runs as follows: on closer examination, the indexicality of 'know(s)' is no more obscure than the indexicality of gradable adjectives like 'flat'. Both of them may violate the 'transparency requirement' Schiffer holds for ordinary indexicals. Imagine the following dialogue:[24]

23 Blome-Tillmann 2008: 52
24 Blome-Tillmann 2008: 36

A: That meadow is flat.
B: But have you considered there are some molehills in it?
A: I guess I haven't.
B: So you admit that meadow isn't flat, and so you were wrong earlier?
A: Oh, c'mon! I didn't say that the meadow is flat.

Intuitively, the last sentence seems false, because we fail to spot the indexicality of 'flat' right away. But our initial confusion can be straightened out by applying so-called 'degree modifiers' like 'completely'. Compare: [25]

A: That meadow is flat.
B: But have you considered there are some molehills in it?
A: I guess I haven't.
B: So you admit that meadow isn't flat, and so you were wrong earlier?
A: Oh, c'mon! I didn't say that the meadow is *completely* flat.

Now we realize that A's first and last sentences are not really contradictory. And, as it turns out, the same can be done for 'know(s)'. Recall the zebra-case above:

A: I know that is a zebra.
B: But can you rule out its being a cleverly painted mule?
A: I guess I can't rule that out.
B: So you admit that you don't know that's a zebra, and so you were wrong earlier?
A: Oh, c'mon. I didn't say I know it *with absolute certainty*.[26]

Presumably, we do not usually talk like this, but that may have reasons other than the context-sensitivity of 'know(s)'.[27] What matters is that the modifier phrase applied reminds us of the two epistemic standards at stake. And that the content of 'know(s)' varies accordingly.

After having considered such and similar cases, Blome-Tillmann derives the following manual for EC's handling of error-theoretical objections: First, one needs to construe parallel problem cases for gradable adjectives. Second, one smoothens those examples containing apparent contradictions by inserting modifier expressions.[28]

Sure, this semantic model for the context-sensitivity of 'know(s)' is just one side of the coin. The finding that 'know(s)' is a unique expression with certain linguistic properties does not yet fully explain our systematic failure to recognize its context-sensitivity. But it gives us a hint: since the context-sensitivity of 'know(s)' is semantic, we are not afflicted by *semantic* blindness. Rather, the lack of intelligibility must enter on the *pragmatic* level of knowledge attributions. Therefore, let's have a closer look at the pragmatics of knowledge attributions.[29]

25 Blome-Tillmann 2008: 39
26 Blome-Tillmann 2008: 39f.
27 Presumably, the reasons are 'know(s)''s being factive and the epistemic norm of assertion. Cf. Blome-Tillmann: 48 f.
28 Blome-Tillmann 2008: 41. Obviously, Blome-Tillmann holds that gradable adjectives are context-sensitive. That's not universally agreed.
29 Semantic blindness is not the only term we should ban from our vocabulary. A 'particular model of context-sensitivity for 'know(s)' seems a candidate as well. After all, 'know(s)' is simply indexical. Its special linguistic behavior is due to its combination with other semantic, syntactic and pragmatic features.

7. Intelligibility and context

A generic approach to the pragmatics of knowledge attributions could look like this: basically, two types of context are operative in knowledge attributions. There is not only the attributor's context, but also the shared 'conversational score' between the attributor and (an)other interlocutor(s). The 'conversational score' could be modeled roughly in Lewisian terms: it manages all the information relevant to a conversation and makes it available to the participants.[30] Misunderstandings occur if we no longer share the same score.

Accordingly, 'know(s)' gets its contextual values assigned in two different stages. The epistemic standards may be determined by the attributor's psychology alone. Additionally, however, there is a parameter on the 'conversational scoreboard' that determines whether the context-sensitivity of 'know(s)' is intelligible to the speakers or not. Let's call it the intelligibility parameter.

Presumably, the intelligibility parameter does not only exist for the context-sensitivity of 'know(s)'. After all, 'know(s)' features the same context-sensitivity as any other, ordinary context-sensitive term. If this is true, their difference in intelligibility can be explained as follows:

In the case of an ordinary indexical like a personal pronoun, the mechanism is comparably simple because of two reasons: the context-sensitive expression at stake does not feature exceptional linguistic properties, and, if intelligibility is a parameter on the scoreboard, there will be an according rule of accommodation. A rule of accommodation serves the purpose of keeping a conversation alive by adjusting apparently incorrect linguistic behavior.[31] So does the rule of accommodation for intelligibility: if the conversationalist ignores the context-sensitivity of a term, the rule assigns the contextual value that is most suitable for the current course of the conversation.

In the case of 'know(s)', things get slightly more complicated. First of all, we are dealing with a linguistically exceptional expression: 'know(s)' does not only feature the semantic property of being indexical, as indicated above. Rather, there are other features bound to interfere with the 'conversational scoreboard'. Normally, I suppose, the 'conversational scoreboard' can handle the variety of parameters pretty well. And the rules of accommodation take care of the rest.

In the skeptical case, however, there seems to be too heavy 'pragmatic traffic' on the scoreboard. As a consequence, we lose track of the intelligibility parameter. As a result, we become ignorant of the context-sensitivity of 'know(s)' and end up being puzzled by the skeptical case.[32]

Obviously, I try to make sense of Neta's inconsistency. Not our communicative intentions, but the 'conversational score' provides us with the non-semantic and non-psychological contextual feature ignored in the skeptical case. When we fail to detect the context-sensitivity of 'know(s)', we are simply ignorant of the intelligibility parameter.

Sure, the model put forward is but a sketch. Nonetheless, I believe it to have the potential for explaining our 'pragmatic blindness' in respect to knowledge attributions. And even if solutions were not to be found in elaborating on my account, it highlights two areas on which proponents of EC should focus their philosophical efforts: First, EC needs to revise its notion of 'context', since it has proven to be too narrow. Second, EC needs to flesh out the pragmatics of knowledge attribution. Sure, EC is an essentially semantic thesis. But, as we have seen, it cannot refute criticism without making sense of some fundamentally pragmatic concepts.

30 Lewis 1979: 344ff.
31 Lewis 1979: 346f.
32 In detail, this process could be modeled after Lewis' explanation of 'relative modality'. Cf. Lewis 1979: 354f. Unfortunately, for reasons of brevity, I cannot discuss that here.

8. Conclusion

In my essay, I have argued that Schiffer's error-theoretical objection rests on a confusion that can be successfully disentangled by introducing the distinction between indexicality and intelligibility. Consequently, I suggested that one treats the problems the distinction frames separately by devising a semantic model for the context-sensitivity of 'know(s)' and a pragmatic model for the intelligibility of its context-sensitivity. I showcased Blome-Tillmann's semantic model and advocated a pragmatic model that operates on two different notions of context, a psychological and a conversational one.

9. Bibliography

Blome-Tillmann, M. (2008): 'The indexicality of 'knowledge'', *Philosophical* Studies, 138 (1): 29-53.

DeRose, K. (1992): 'Contextualism and Knowledge Attributions', *Philosophy and Phenomenological Research*, 52(4): 913-929.

DeRose, K. (1999): 'Contextualism: An Explanation and Defense', in *The Blackwell Guide to Epistemology*, J. Greco and E. Sosa, eds., Malden MA, pp. 185-203.

Hawthorne, J. (2004): *Knowledge and Lotteries*, New York and Oxford: Oxford University Press.

Hofweber, T. (1999): 'Contextualism and the Meaning-Intention Problem', in *Cognition, Agency and Rationality*, K. Korta, E. Sosa, and X. Arrazola eds., Dortrecht, Boston, and London: Kluwer, pp. 93-104 or URL = <http://www.unc.edu/~hofweber/papers/con.pdf> (4.12.2008)

Lewis, D. (1979): 'Scorekeeping in a Language Game', *Journal of Philosophical Logic*, 8: 339-359.

MacFarlane, J. (2007): 'Non-indexical contextualism', DOI 10.1007/s11229-007-9286-2, URL = <http://www.springerlink.com/content/e072383726380533/fulltext.pdf> (4.12.2008)

Neta, R. (2003): 'Skepticism, Contextualism, and Semantic Self-Knowledge', *Philosophy and Phenomenological Research*, 67(2): 397-411.

Rysiew, P. (2001):'The Context-Sensitivity of Knowledge Attributions', *Noûs*, 35(4): 477-514.

Rysiew, P. (2007): 'Epistemic Contextualism', *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/entries/contextualism-epistemology/> (4.12.2008)

Schiffer, S. (1996): 'Contextualist Solutions to Skepticism', *Proceedings of the Aristotelian Society*, 96: 317-333.

Material Implication and Indicative Conditionals

Benjamin Perlin

Introduction and Definitions

It has often been asked whether the truth-function known as material implication correctly accounts for conditionals in the indicative mood. After defining material implication and indicative conditionals (hereafter just "conditionals"), I will discuss why I believe the former does not always account for the latter. Defences for a material interpretation of conditionals by H. P. Grice and Frank Jackson will then be given.

A function is analogous to a machine which outputs something when something is input. The inputs and outputs of truth-functions are truth values: "true" or "false". The symbol for material implication ('⊃') is thus formally defined: if the sentence before it (the antecedent) is true and the sentence after it (the consequent) is false, then the material implication is false; otherwise it is true.

Conditionals are a complex sentence form; they are made up of sentences and can be either true or false (but not both). If *A* and *B* are any sentences, then "If *A*, then *B*" is the conditional form. The previous sentence is also a conditional (*A* and *B* can be complex sentences, like "The flag is raised and somebody is dead.") As with material implication, *A* is the antecedent and *B* is the consequent.

Conditionals with synthetic antecedents and consequents will be considered, rather than conditionals with analytic antecedents or consequents. The subject in a synthetic sentence – like "the flag" in the sentence "The flag is raised" – does not somehow contain the predicate (here "is raised"). Contrast this with the analytic sentence "The white swan is white." Since this cannot be false, we cannot speak of "If the white swan is white, then the white swan is white" having a false antecedent or consequent, which is crucial.

Material Implication does not Necessarily Express Conditionals

Does material implication correctly account for, say, "If the flag is raised, then somebody is dead"? The question is whether the sentence is false when "The flag is raised" is true and "Somebody is dead" is false, but true otherwise.

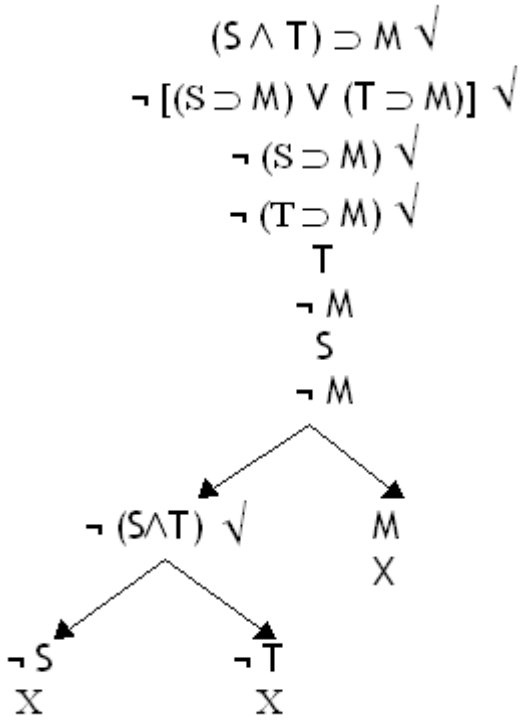
First of all, a speaker of the sentence is not necessarily saying anything about "The flag is raised" being false or anything consequent on its falsity. They are not doing so explicitly in any case. The assertion may just be that a dead person is a necessary and sufficient condition for a raised flag. The sentence is true if both the antecedent and consequent are true; the sentence is false if the antecedent is true and the consequent is false. That is all.

Secondly, conditionals can be used within a non-formal language for different purposes. They do not always operate under the same truth conditions. There are circumstances in which the truth conditions of the sentence "If the flag is raised, then somebody is dead" are more numerous than the above: a person may say it within the context of a military base, implying strongly that if the flag is not raised then nobody is dead. If this occurs, then the sentence (its suggestion strictly speaking) is true. It is, however, difficult to imagine a case where the sentence is true when the antecedent is false and the consequent is true.

There is a popular counterexample to the material account of conditionals by William S. Cooper. Suppose there is a motor hooked up to two switches (S and T) and that the only information we are given is expressed by the sentence "If S and T are presently thrown, then the motor starts."

This sentence is formalized, on the material interpretation, as *(S ∧ T) ⊃ M*. Throwing both switches is a sufficient condition for the motor starting, but it is unknown whether the motor starts if either switch is thrown independently. The sentence "It is the case for one or other of the switches that if that switch is thrown (independent of whether the other is) that the motor will start" (formalized on the material interpretation as *[(S ⊃ M) v (T ⊃ M)]*) can be false.

But in normal classical logic, the latter cannot be false if the former is true:



Defences of the Material Account of Conditionals

The traditional view nevertheless posits conditionals as accounted for by material implication. One argument for this view (**A1**) relies on the implicational relationship between disjunctions (complex propositions of the form "either A or B") and conditionals:

Assumptions	**Formulae**	**Justification**
1	(1) E	Assumption
2	(2) D	Assumption
1,2	(3) L	1,2
1,2	(4) ¬ R v S	3 NC formalization
1,2	(5) R⊃S	4 Implication

("D" = "if the flag is raised, then somebody has died"; "E" = "propositions of the form 'if A, then B' are equivalent to propositions of the form 'either not A or B'"; "L" = "either the flag is not raised or somebody has died"; "R" = "the flag is raised"; "S" = "somebody has died.")

Grice and Jackson: The Counterexamples Cannot be Asserted

H. P. Grice accepts the material interpretation of conditionals. He therefore considers statements such as the following to be paradoxes: no proposition can imply (as the antecedent of a conditional) an arbitrary consequent by being falsified; yet 'P ⊃ Q' cannot be false if '¬ P' is true. His response to purported counterexamples is to introduce a distinction between two properties of propositions: appropriateness for conversation and truth. Neither implies the other.

Whether or not a proposition should be asserted is determined by certain maxims. The maxims of quality and quantity particularly ensure the cooperation of language users. The maxim of quality is a requirement for propositions to be true and justified. The maxim of quantity requires the contribution of the speaker to be sufficiently informative but not more informative than is necessary (Grice is uncertain about the latter point). Suggestions follow from conversation when the maxims are assumed.

If the material interpretation of conditionals is correct, then "If S and T are thrown, then the motor starts" (P ⊃ Q) [1] is false only when the antecedent is true and the consequent is false. Grice points out that the conditionals in the purported counterexamples are consistent with this; they demonstrate rather that the conditionals should not be asserted. If I understand him correctly, he assumes that what should not be asserted cannot be formalized; if P ⊃ Q cannot be formalized, neither can $\neg P \mid =_{NC} P \supset Q$.

Why should these conditionals not be asserted? The falsity of the antecedent or the truth of the consequent occurs in these scenarios. If the sentence "S and T are not thrown simultaneously" (¬ P) conveys as much information as P ⊃ Q, then it meets the maxim of quantity when $P \supset Q$ does not. $P \supset Q$ asserts more than is necessary. The same holds for the sentence "the motor is starting" (Q) in place of $\supset P$.

To refute the counterexamples, Grice relies on a suggestion which follows from conversational maxims; Jackson relies on a conventional suggestion about all propositions of the form $P \supset Q$. Conditionals have a specific purpose in Jackson's account. If a speaker asserts "if A, then B", then she is demonstrating that she accepts the necessary truth of B given A (*modus ponens)*. However, such a demonstration cannot occur in the counterexamples.

Suppose a speaker believes the proposition "S and T are not being thrown" (¬P), for example. If she is not informed about Q – "the motor is running" – then a statement of $\neg P$ is stronger than an assertion of $\neg P \vee Q$, which conveys more information than is necessary. It would nevertheless be appropriate to assert the latter as long as-and this is the crucial point-she believes $\neg P$. If P is found to be true, the disjunction would not be stated or would be withdrawn; she would not move on to infer Q by a negation of $\neg P$ in the disjunction and *modus tollendo ponens*.

Since $\neg P \vee Q$ is equivalent to $P \supset Q$, knowledge of P would also make $P \supset Q$ not highly assertible. The conditional "If S and T are being thrown, then the motor is starting" could not be operated on by *modus ponens*.

This distinguishes the conditionals in the counterexamples from those which are "robust" enough to be believed when their antecedents are true. They are not asserted merely because their antecedents are believed to be false, as in the above example. Take the principle that any proposition is either true or false (but not both): $T \vee F$[2]. This is highly assertible, according to Jackson, even when it is learned which disjunct is correct; $\neg T \supset F$ is highly assertible for any proposition and *modus ponens* can (a priori) operate on "If a proposition is not true, then it is false."

1 This is a valid formalization, though different from the one above, and serves the present purpose better.

2 Strictly, the formula should be written as $(T \vee F) \wedge \neg (T \wedge F)$; but the former conjunct in these cases is often written alone as a matter of convention.

Bibliography

Cooper, W. S. (1968) "The Propositional Logic of Ordinary Discourse", *Inquiry* 11: 295-320

Grice, H. P. "Logic and Conversation." *The Philosophy of Language*. Ed. A.P. Martinich. 5th ed. New York and Oxford: Oxford University Press, 2008. 171-181.

Hausman, Alan, Howard Kahane, and Paul Tidman. *Logic and Philosophy: A Modern Introduction*. 10th ed. California: Thomson Wadsworth, 2007.

Read, Stephen. "Semantics for Classical Logic". Class notes. University of St Andrews: unpublished, 2008.

Freedom and its Capacity to Shape Morality

Lukas Lohove

How Kant's understanding of freedom leads to being obliged to act morally

In our everyday life we face a multitude of moral questions. Often these are not posed explicitly but, still, there are many delicate choices to be made: for example, whether or not we ought to be truthful to a friend knowing this will make her unhappy or whether we ought to scan all shopping goods at the self-service counter in the supermarket although we know that nobody would notice our leaving one out. Most of us have a clear opinion on what we think is right and what we think is wrong, but what is the ground for that? Questions as those posed in the examples above call for principles which guide us to right actions. What kind of principle for morality could there be? And if there is one, do we have the freedom to choose to act in accordance to it or is there an obligation which confines freedom in this sense? How can we be obliged to act morally? How does this relate to our freedom?

In his influential and widely read book *'The Groundwork of the Metaphysics of Morals',* Immanuel Kant (1724-1804) attempts to answer these questions. He tries to establish the supreme principle of morality[1]. For Kant morality is the normative guideline of conduct that all rational agents should follow.

Three concepts are of prime importance for his argument: the *will*, *autonomy* and *freedom*. Crudely stated, the *will* is what causes our action, thus makes us act; *autonomy* means being governed by self-imposed laws and *freedom* has various meanings as we shall see when we move on. The two concepts, autonomy and freedom, are *a priori* propositions, which mean they cannot be justified by appealing to experience. For justification, they require a so-called 'synthetic'[2] argument, that is, an argument linking the two distinct concepts by using a third term.

A synthetic argument is opposed to an analytic argument which merely works through analysing a concept by what the term entails, for example from the word "ice" it can be derived that "ice is solid" by analysing the concept of "ice". An example for a synthetic argument is the sentence "ice is floating", which cannot be derived from the word "ice" but backed up by referring to experience.[3] In Kant's case the concepts – autonomy and freedom – cannot be connected through experience since they are a priori, as mentioned above, but the type of argument required is also a synthetic one since the concepts are distinct.

In order to pursue the intended justification of (i) the autonomy of the will and (ii) the moral demand all imperfectly rational beings experience, Kant introduces the concept of *freedom.* From the concept of freedom he derives morality. Moreover, freedom leads him to the required third term. The moral demand takes the form of the c*ategorical imperative (CI),* that is, Kant's widely known principle of morality.

"Act only on that maxim through which you could at the same time will that it should become a universal law".

What exactly he means by these concepts and how he relates them we shall see shortly. Before we begin to examine his argument it is worth noting that Kant's conception of freedom is different from what is called 'neutral freedom'[4], that is, the freedom of choice whether to act morally or not. Kant's conception, as we shall see, does not leave us with this choice.

1 In this article I will refer to the page numbers of H. J. Paton's *The Moral Law*; I will refer to Kant's text by writing 'Kant' and to Paton's commentary by writing 'Paton': Kant p. 61
2 Kant p.62
3 Ross
4 Timmermann p. 164

The negative conception of freedom

In the third chapter of his book Kant begins with the claim that all rational beings have a will, that is, a 'kind of causality'[5] since it causes actions[6]. If I stand in front of a tree and reach out for an apple then this interaction, given that it was rational, was initially caused by my will. A will, according to Kant, is free in that it is able to work independently from sensuous influences, such as inclinations or desires, and merely springs from reason. The determination by 'alien' causes in general is labelled *natural necessity*[7]. Everything in the natural or sensible world is subject to the natural laws, that is, the laws of cause and effect. For example, if one moving billard ball – the cause – hits another it induces the movement of the second – the effect. Everything that we can *experience* is determined by natural necessity. This is the first conception of freedom: freedom is the non-determination of the will by sensuous or 'alien' causes. It is a negative conception since it states only that the will is not externally determined. From this conception it follows that freedom is opposed to natural necessity.

The second and positive conception of freedom

From this point Kant infers that, even though the will is free from natural laws, it must still be subject to laws since the will is a 'kind of causality' and causality always requires laws. Since if X causes Y, there must be a connection governed by laws between X and Y[8]. Even if X is my will which causes me to pursue action Y, there must be a law governing this relationship. The laws determining the will must, however, be different from the natural laws because otherwise the will - being free from natural necessity - would be self-contradictory. The need for these laws leads Kant to the second conception of freedom which is positive. If the will were determined by an 'alien' cause, it would be a will under *heteronomy*, which is Kant's technical term for the state of being caused by something other than itself. Since the will is not determined by any 'alien' cause but still has to be under laws, it has to be a will under *autonomy*, that is, governed by self-imposed laws. Freedom therefore implies autonomy.

The next step in Kant's argument is contentious and not well-supported. According to Kant, autonomy implies acting only on those principles that can at the same time be willed to become a universal law, which is one formulation of the CI.

It is now clear that the will should be under a set of laws, but why should these laws take the shape of the CI? Kant reasons that saying a will is under self-imposed rules means the same as saying that a 'will […] is in all actions a law to itself'[9], which he classifies as a modified formulation of the CI. Thus, by presupposing and simply analysing this conception of freedom he derives morality in the form of its supreme principle. Morality is in this sense inherent in the concept of freedom. Nevertheless, morality is still a synthetic proposition that is in need of a third term, which can be found by using the concept of freedom.

Freedom as property of all rational beings

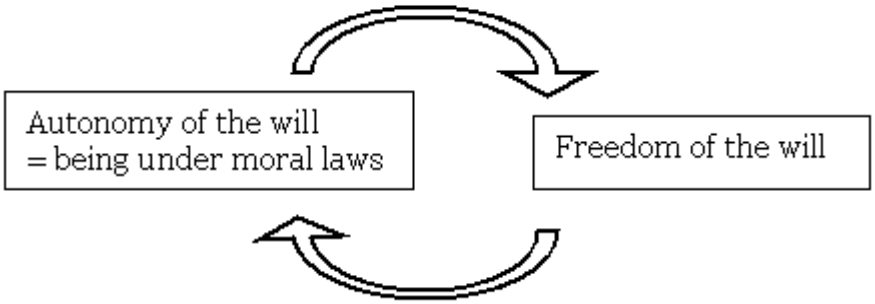
If morality is to apply to every rational being, freedom must be compellingly ascribed to all rational beings. This is impossible by sensuous experience because freedom is an *a priori* concept, which means it is what Kant calls an *Idea*[10]: a concept that cannot be proven by empirical means because it

5 Kant p. 127
6 Kant p. 127
7 Kant p. 125
8 SEP – Kant's moral philosophy
9 Kant p. 128
10 Paton p. 41

does not occur in the natural world. However, freedom can be conceived as the 'property' of every rational agent. In fact, no rational agent could be conceived as being capable of his own thoughts and decisions if he was not free from external determination. Thus, in practical terms, every rational agent with a rational will can only act on the 'Idea of freedom'[11], that is, assuming non-determination and the will as a 'first cause'.

The vicious circle

This leads to a problem: having argued that the freedom of the will implies autonomy - which means self-regulation - and thus being under moral laws, Kant argues in turn that freedom has to be presupposed because otherwise autonomy would not be possible. This appears to be a vicious circle since one concept cannot be used to justify the other if they are reciprocals.



The two standpoints

This is solved by Kant though appealing to his metaphysics. All things can be viewed from two standpoints. By merely 'observing' something – using one's senses -, a thing is perceived as a mere *appearance*[12]. Kant calls this the *sensible world* where the laws of nature apply. In this world one billiard ball hits the other which causes the second to move. By contrast, there is a world that is 'something more'[13] beyond one's senses. This world can only be conceived by reason and here the laws of morality apply. In this world a thing is not a mere *appearance* but a *thing in itself*, that is, it contains a part that cannot be experienced by our senses. This world Kant calls the *intelligible world*.

A rational being that is imperfect in the sense that it is influenced by both reason and sensuous inclinations, necessarily has to conceive itself as a member of both worlds. Consequently, it is subject to two different kinds of laws. As far as one is under sensuous influence one conceives oneself as part of the sensible world, therefore being subject to the laws of nature. However, as far as one is rational one conceives oneself as part of the intelligible world - being free from 'alien' determination - and is thus bound to conceive one's causality under the 'Idea of freedom'[14], which is directly linked to autonomy, which in turn means being under moral laws.

In conclusion, rationality entails the ability to distinguish between the two standpoints. Acknowledgement of membership of the intelligible world shapes one's conception of one's causality as being free. Consequently, provided that one agrees with his metaphysics, Kant avoids being trapped in the vicious circle mentioned above.

11 Kant p. 130
12 Kant p. 133
13 Kant p. 146
14 Kant p. 135

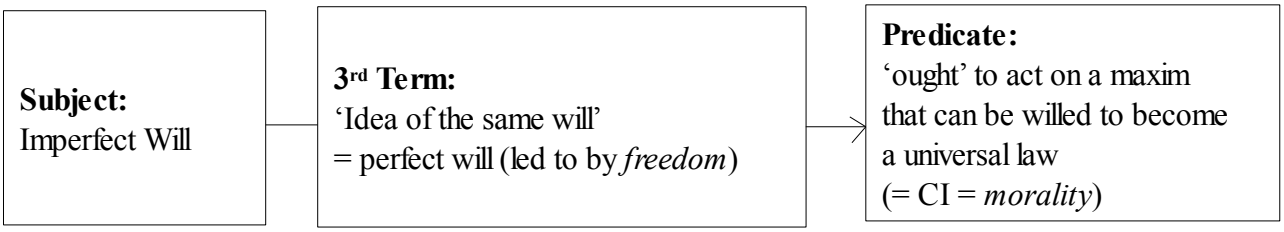
Why is the categorical imperative binding? - The third conception of freedom

The binding character of morality, that is, the CI, is possible because 'the intelligible world contains the ground of the sensible world and also of its laws'[15]. It should be noted that Kant is regrettably vague in his explanation as to why this assumption is true. However, he seems to reason that the will of a rational being 'ought'[16] to conform to the principle of autonomy, although this being is, from a different standpoint, also part of the sensible world. This shows the third conception of freedom: the capacity to subordinate all sensuous influences to reason. This capacity implies a necessity of the free will to conform to moral laws.

How does this necessity come about? The necessity or 'ought' statement mentioned above is an a priori proposition which can only be conceived but not be proven empirically. For a logical connection between a subject and a predicate in a synthetic argument a third term is required that establishes the link. Thus, in order to link the imperfect will of a rational being to the moral obligation to act in accordance with the CI a third term is needed of which they are both part.

A rational being that is free in the third sense conceives of itself as part of the intelligible world and thus has a conception of its own will as a solely intelligible will. What is meant by a solely intelligible will? To explore this in more detail, let us think of a person that is perfectly rational and is not influenced by any desires or inclinations. This person would naturally act in accordance with the laws of morality and the will of that person would be perfect. But humans are under the influence of the sensuous world and thus they are only imperfectly rational and possess only an imperfect will. However, they are able to conceive of their will as being perfect since they are part of the intelligible world. Kant dubs the conception of a perfect will the 'Idea of the will'[17]. This solely intelligible will – being beyond any sensuous influences – serves as the third term which Kant was seeking. It is a supreme condition of the will which we were directed to by the third conception of freedom.

Since, firstly, a rational being that is free in the third sense is capable of subordinating all sensuous influences to reason and, secondly, is able to conceive of its own will as being perfect or solely intelligible, the binding character of the the moral law becomes evident: subjectively, an imperfectly rational being perceives the law of morality thus as a categorical imperative, that is, as an 'ought' statement without exceptions, and the actions that conform to these laws as duties.



Conclusion

Having examined the question of how Kant relates freedom to morality, we have seen that, according to Kant, freedom – as non-determination by external sources – is a necessary presupposition of all rational beings. This leads to the positive conception of freedom as reciprocal of the principle of autonomy. As such, a will under freedom is one and the same as a will under the CI since a self-governed will is subject to its own laws and these laws can be identified as the CI. However, I find this argumentative connection between autonomy and CI questionable. Moreover,

15 Kant p. 136
16 Kant p. 137
17 Paton p. 43

freedom – as a conception of the capacity to subordinate inclinations to reason – is the reason as to why moral laws are binding. His metaphysics - the two standpoints – play a pivotal role for this moral authority of the CI: every rational being that is – necessarily so because it is rational - capable of conceiving itself as member of both the intelligible and the sensible world, will conceive the moral law as what a 'pure will' would aspire and perceives it as an imperative that it 'ought' to act upon. However, I object to Kant's metaphysics since such a rigid distinction between reason and emotions it hardly existent in any human being and therefore implausible. This objection thus questions his argument since it is then trapped in the vicious circle. Despite this objection, Kant uses the *Idea* of freedom to justify both (i) the existence and (ii) the authority to act in accordance with morality. Being free in this sense implies the obligation to act morally.

References

Kant, Immanuel. *The Moral Law,* H. J. Paton (trans.) (Oxfordshire: Routledge, 2005)

Ross, David. The *Analytic-Synthetic Dichotomy,* March 26th 1992, <http://enlightenment.supersaturated.com/essays/text/ioe1/09.html>, Accessed on 4th May 2009

Johnson, Robert, "Kant's Moral Philosophy", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2008/entries/kant-moral/>

Timmermann, Jens. *Kant's Groundwork of the metaphysics of morals : a commentary* (Cambridge: Cambridge University Press, 2007)

Recent Developments in Neuroscience and Moral Objectivity

Malcolm Collins

A discussion by a neuroscience student

In this short paper I would like to propose the following: human judgments of morality are not, and can not be objective given unavoidable aspects of human neural anatomy. They can be influenced by brain damage, your genetics, or even switched up and down in intensity at will using methods like Transcranial Magnetic Stimulation (TMS), which I shall later discuss. Therefore any sense of an ability to make objective morality judgments is an illusion. Consider, for example one realist moral philosopher in his room making his decisions in normal circumstances, and another who is trapped by a mad scientist manipulating his mental states without him even realizing. Is there really a difference between these cases? We will first explore the evidence for thinking there may not be, then return to our captive moral philosopher strapped to a TMS device.

While there are multiple ways to both investigate and alter a person's judgment of morality this paper will focus primarily on the 'ultimatum game' as a method of measuring morality judgments and the right dorsolateral prefrontal cortex (DLPFC) as the part of your brain making the call as to what is moral and what is not. The ultimatum game developed by Güth, Werner, Schmittberger, and Schwarze in 1982 is a stylized representation of negotiation often used when exploring game theory and models of economics though more recently it has been used as a measure of fairness judgments both by anthropologists in cross-cultural studies and psychologists/neuroscientists. In it one player is given an amount of money and has to offer a portion of it to a second player. If the second player rejects the amount of money offered then neither player is allowed to keep any of the money but if the second player accepts the proposal, the money is divided along the lines suggested by the first player and kept by the two players. The two players interact anonymously and only once so reciprocation in not an issue. 50/50 splits are almost always accepted but splits of 20% or less are often rejected being deemed as “unfair” *(Oosterbeek et al. 2004; Henrich et al. 2004).* “Humans appear willing to forego material payoffs to punish unfair behavior,” *(Wallace et al. 2007).* As a side note this behavior of judging fairness and then punishing unfair behavior seems unique to humans and is not observable in chimpanzees *(Jenson et al 2007).*

The judgment of fairness of different proportional splits of money can be influenced by external variables beyond the decider's control. Studies in which identical and fraternal twins separated at birth were measured to find the point at which they made the judgment that an offer was unfair have shown that “additive genetic effects account for 42% of the observed variation in (the) responder” *(Jenson et al 2007.)* and “we estimate that >40% of the variation in subjects' rejection behavior is explained by additive genetic effects.” *(Wallace 2007)*

Studies on the effects of hormones on one judgment of fairness found that “High-testosterone men reject low ultimatum game offers” *(Burnham 2007)* and that manipulating a person's serotonin (5-HT) levels will effect their judgments of offers as fair or unfair *(Crokett et al 2008).* More dramatically than the above are the ability of TMS to virtually turn judgments of fairness up or down.

Transcranial Magnetic Stimulation (TMS) is a noninvasive process whereby neurons are excited by weak electric currents created by a device using rapidly changing magnetic fields also known as electromagnetic induction. A technique using TMS called repetitive TMS (rTMS) can actually “turn off” part of the brain for a period of time. Depending on what area of the brain it is being used on, the subject can not even tell that they have been affected, this is the case with

judgments of when to accept a fair offer. Studies have repeatedly shown that using this technique you can alter someone's perception of when it is appropriate to accept an offer *(Wouta et al 2005; Knoch et al 2006)*. More specifically, "After rTMS over the right DLPFC, however, this pattern was changed, with longer reaction times for rejecting unfair offers, and a trend towards more acceptances of unfair offers" *(Wouta et al 2005)*. It is worth noting that at least one study – Knoch et al 2006 – suggests that patients can still judge the offer as being unfair but has less qualms with accepting it.

We readily accept that in certain circumstances our judgments, including moral ones, are influenced and could lead us to make bad judgments and decisions. For instance, a person may commit actions they would usually deem as immoral if coerced, or drunk or in some variety of high stress situation. In these circumstances, however, it seems that there is some alteration in the first-hand experiences of the agent. In the case of TMS, however, an agent can be entirely unaware that they are being effected by certain psychological factors.

The fact that a philosopher's judgments of morality or at the very least how they act on those judgments can be so easily influenced leads to a number of interesting questions about human perspectives of morality. For example consider our philosopher held captive and strapped into a TMS device. Let's say that he is given the option of escape if he presses a button that will kill a random stranger. While the TMS is acting on him he is more willing to make the less morally hard line decision. Later, when the TMS's effects have worn off he judges his action as morally wrong. Is he at fault? It was his own line of logic that lead to him choosing to press the button after all. And if he wasn't at fault, does that mean that the judgment of morality "he made" wasn't actually made by him? If this is the case then it would mean that if you have a neurochemical state that is causing you to make one moral choice over the other and that state is out of your control then you are absolved of your choice, but we are all influenced in the same way by both our genes and hormone levels. So are no moral choices really our own? In this paper I will not address the questions this has brought up, but perhaps it will give you something to think about, particularly next time you make a moral decision...

References

Burnham T (2007). High-testosterone men reject low ultimatum game offers. *Proc Biol Sc*,. 274, 1623, 2327–2330

Crockett M, Clark L, Tabibnia G, Lieberman M, and Robbins T. (2008). Serotonin modulates behavioral reactions to unfairness. *Science*, 27, 320, 5884,1739.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis (2004). *Foundations of Human Sociality*: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies. Oxford University Press.

Jensen K, Call J, and Tomasello M (2007). Chimpanzees Are Rational Maximizers in an Ultimatum Game. *Science*, 318, 5847, 107-109

Knoch D, Pascual-Leone A, Meyer K, Treyer V, Fehr E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, 3, 314, 5800, 829-32.

Oosterbeek, Hessel, Randolph Sloof, and Gijs van de Kuilen (2004). Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. *Experimental Economics*, 7, 171–188

Sanfey G, Rilling K, Aronson A, Nystrom E, and Cohen D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, 13, 300, 5626, 1755-8.

Singer T. (2007) The neuronal basis of empathy and fairness. Novartis Found Symp, 278, 20-30

Wallace B, Cesarini D, Lichtenstein P, and Johannesson M (2007). Heritability of ultimatum game responder behavior. *PNAS*, 104. 40. 15631–15634

Wouta M, Kahnb R, Sanfeyd G and Alemanc A (2005). Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Cognitive Neuroscience and Neuropsychology Neuroreport*, 16, 16 7

A Priori Entailment is not Worth the Costs

Jönne Speck*
29th April 2009

Is metaphysics essentially an investigation from the armchair? An exercise characteristic of armchair philosophy is analysis. The philosopher takes a term '*F*' she is interested in and enquires into the necessary and sufficient conditions for something being an *F*. In a series of articles Frank Jackson argues that such analysis does play an essential role in metaphysics.

This essay evaluates his argument. In the first section, I reconstruct Jackson's inference from what constitutes serious metaphysics to the essential role of analysis. Section 2 presents obvious objections to this argument which cause Jackson to elaborate a two-dimensional descriptivism of natural kind terms.

This, however, leads straight into a dilemma, or so I argue (3). The final section bolsters my refusal of Jackson's argument by identifying a valid and less controversial alternative.

1. Why serious metaphysics would be committed to analysis

Jackson starts from the assumption that metaphysical theorising is only serious if it attempts to explain everything in a restricted basic language [Jackson, 1994, 25]. Therefore, a serious metaphysical theory *T* would be equivalent to a global supervenience claim

superT: Any world whose description in terms of *T* is identical to *T*'s description of the actual world is a duplicate simpliciter of the actual world.

superT holds if and only if at any world where *T* is true, any true sentence not in *T*'s basic language is also true. Assuming, for example, that *T* contains

1. *NaCl* contains 1.8% iodine ,

superT implies that

2. Salt contains 1.8% iodine.

is true at all worlds where *T* is true. In other words, *superT* presupposes that *T* entails (2). Hence, *T*, being equivalent to *superT*, has to account for the fact described by (2) to become one of its theorems (26). In Jackson's terms, 'entry by entailment' is the only solution for this 'placement problem'.

The entailment at issue, however, must transcend metaphysical entailment according to which *p* entails *q* iff at any world where *p* is true, *q* is true. This cannot vindicate *T* because *superT* already implies that at any world where (1) is true, (2) is also true. Therefore, citing metaphysical entailment in support of *T* would beg the question. Additionally, metaphysical entailment would fail to elucidate how one arrives from (1) at (2). To cross this explanatory gap, the metaphysicist better add some reasoning. Surely, assuming

3. Salt is *NaCl*.

the step from (1) to (2) is a plain Leibniz substitution. Nonetheless, Jackson denies that this straightforward deduction solves placement problems.

*University of St Andrews and Ludwig-Maximilians-Universität München,
jonne.speck@gmail.com

Deviating from Kripke's original distinction between epistemological and metaphysical necessity [Kripke, 1980, 35 - 37] Jackson contends that (3) and

4. *NaCl* is *NaCl*.

express the same necessity but differ in how this necessity can be known [Jackson, 1994, 34] [Jackson, 1998a, 77]. Whereas (3) is necessary merely *a posteriori*, (4) is an *a priori* necessity .

Based on this epistemological understanding of necessity Jackson gives a stronger notion of entailment which he thinks is necessary to fill the explanatory gap between (1) and (2): a priori entailment. Although metaphysical entailment $\ulcorner p \Vdash q \urcorner$ guarantees a conditional $\ulcorner p \rightarrow q \urcorner$ to be necessary, this necessity would be merely a posteriori. p a priori entails q, however, not only if p is true in all worlds where q is true, but also $\ulcorner p \rightarrow q \urcorner$ must be an a priori necessity. To solve her placement problem, accordingly, the metaphysicist has to maintain an a priori necessary conditional with (2) in the consequent and (1) in the antecedent. Because (3) is a posteriori, the above deduction does not suffice for this.

Now, analysis enters the stage. According to Jackson [Jackson, 1998a, 80 - 82], it provides a priori knowledge of

3'. Salt is the actually salty stuff

Additionally, for the sake of the argument, *T* contains

4'. NaCl is the actually salty stuff

Since (1), (3') and (4') entail (4), analysis renders '(1) ^ (40) → (43)' an a priori necessity. Thus, (3), although being an a posteriori truth, is derived a priori from (1) and (4'). Therefore, given the above deduction and the logical truth of transitivity '(1) ^ (40) → (2)' becomes an a priori necessity itself. Hence, analysis allows the metaphysicist to demonstrate that (1) does indeed a priori entail (2).

In sum, Jackson argues that serious metaphysics brings with it placement problems which cannot be solved but by identifying entailment relations. As this entailment needs to be a priori, and only analysis provides the required a priori knowledge, analysis plays an essential role in metaphysics.

2. Two-dimensional descriptivism of natural kind terms

Jackson's crucial assumption is that analysis provides the required a priori knowledge of (3'). Taking a step back, though, this seems little plausible[Harman, 1994, 43]. The paradigm of a priority, as Harman points out, is our knowledge of logical truths. Whereas this, however, can be achieved by mere deduction, analysis relies on induction in two respects. First, the philosopher generalises from various judgements to the unique intuition about a possible case. Second, she infers from a small number of cases a definition which is supposed to hold generally. This epistemic difference already requires Jackson to specify what he means by 'a priori'. The various objections raised against the analytic-synthetic-distinction also cast doubt on whether the analysis of concepts yields a priori knowledge. Therefore, in order to render '(1) → (2)' a priori Jackson needs to elaborate the traditional conception of 'a priori'.

This is even more so as Jackson's champions analysis of natural kind terms. Thus, knowledge of (3') being a priori demands the subject to know the reference of 'salt' merely by means of her linguistic competence. If so, 'salt' would refer to whatever is the white, powdery stuff which is present in sea-water and is used to flavour and preserve food. More generally, the reference of a natural kind term '*F*' would be determined by which properties a speaker associated with it.

This, however, amounts to a descriptivist theory of reference. Hence, Jackson's assumption holds only if such descriptivism holds. The well known externalist cases studies, however, have swept away traditional descriptivism by showing that the reference of '*F*' is independent of whichever descriptions speakers associate.

Against this obstacle Jackson applies considerable effort, by developing a two-dimensional approach to natural kind terms [Jackson, 1994, 39], [Jackson, 1998a, 46]. Two-dimensional semantics is the approach of disambiguating traditional conceptions of semantic value into two different aspects. With Jackson, the difference is drawn between two types of functions from possible worlds into extensions, *C*- and *A*-intensions. This distinction is based on two different ways to think of possible worlds [Jackson, 1998a, 47]. From the first stance a natural kind term '*F*' is used at the actual world @ to talk about another world *w*, and thus has the same extension at any *w*, namely whatever is an F at @. In this sense, 'salt' refers to *NaCl* even at Twin Earth. From the second point of view, however, '*F*' is used as if *w* would be the actual world. Then, 'salt' refers to *AbCd*, but again at all worlds, considered as counterfactual. The *A*-intension, now, takes this latter stance and maps actual worlds to extensions, whereas the C-intension distinguishes one actual world and gives the according extension for counterfactual worlds. This is crisply represented in tables, as it is done for 'salt' in table 1.

	@	w_1	w_2	...	possible worlds as counterfactual worlds
@	*NaCl*	*NaCl*	*NaCl*	...	← *C*-intension
w_1	*AbCd*	*AbCd*	*AbCd*	...	
w_2	*EfG*	*EfG*	*EfG*	...	
...	...	...	...	...	
worlds as actual					↖ *A*-intension

Table 1: the two-dimensional meaning of 'salt'

Jackson exploits this framework to revive a descriptivist theory of reference for natural kind terms `*F* '. He admits that its *C*-extension is not determined by a description. 'Salt', as used at @ talking about *w*, refers to *NaCl* although *AbCd* is the salty stuff at *w*. Nonetheless, he claims that the *A*-intension of 'salt' corresponds to a rigidified definite description [Jackson, 1994, 39], that is a conjunction of stereotypical features ('salty') a sortal ('stuff'), a uniqueness clause ('the') enhanced by an operator which species the actual world ('actually'): Salt is the actually salty stuff.

Based on this semantics Jackson elaborates the epistemology of analysis. Whereas linguistic competence alone does not suffice to provide knowledge of its *C*-intension, as the externalist cases reveal, mere reflexion about one's implicit conceptual understanding of *F*, he claims, gives knowledge of its *A*-intension. A speaker of English may not know that salt is *NaCl*, or a Twin-Earthling that what he calls 'salt' is *AbCd*. However, or so Jackson presumes, both know that salt is the salty stuff of their respective acquaintance, know that rigidified definite description which makes up the *A*-intension. Since this knowledge is acquired as soon as the English respectively Twin-English word 'salt' is understood, it does not depend on which world is the actual. For Jackson, this circumstance is sufficient for knowledge of a term's *A*-intension to be a priori [Jackson, 1998a, 50]. Thus, analysis would indeed yield a priori knowledge, could solve placement problems and therefore play an essential role for serious metaphysics. This result, however, stands or falls on the presumption that for any natural kind term, any speaker has a priori knowledge of a rigidified definite description which makes up the term's *A*-intension. In the next section I show how contentious this assumptions is.

3. Two-dimensional descriptivism is controversial

To establish analysis as the only cure against placement problems, Jackson has developed a two-dimensional descriptivist semantics of natural kind terms. Traditional descriptivism was defeated by counterexamples. The fatal weakness of Jackson's argument is that against his two-dimensional descriptivism, too, counterexamples can be construed. They show that any constituent of the definite description, be it the uniqueness clause, the sortal or one of the stereotypical features, is revisable in view of empirical findings. Any adjustment to save a priori knowledge of *A*-intensions, it is demonstrated, either weakens the description to triviality or, as Laura Schroeter puts it, credits '[...] us with a more accurate understanding of the reference of our concepts than we seem to have' [Schroeter, 2004, 432].

First, that natural kind terms are not a priori linked to stereotypical properties is suggested by cases found in [Block & Stalnaker, 1999, 432]. Laurence and Margolis [Laurence & Margolis, 2003, 261-263] explicitly tie up with Putnam's argument against a descriptivism of kind terms [Putnam, 1970, 187-190]. They point out that to command the term 'salt' does not presuppose knowledge of salt being liquid, clear or having anyone of the properties commonly associated with it as these are contingent facts about our world. In some passages [Jackson, 1994, 39],[Jackson, 1998b, 241], Jackson anticipates this objection and allows deviant cases as long as enough of the stereotypes are fulfilled; however, he fails to specify and justify the limit. Presumably, he would have to allow extreme deviations. In fact, as Block and Stalnaker point out, nothing guarantees that any stereotype is fulfilled at all. Effectively, he is compelled to trivialise speakers' knowledge of how salt is like.

Second, Schroeter [Schroeter, 2004, 439] points out that Aristotle thought of salt as one of the four basic configurations of prime matter. Today, chemical inquest has revealed that salt is a chemical kind, accordingly speakers associate a different sortal. Apparently, speakers do not have an infallible understanding and therefore no a priori knowledge of the sortals which are part of the rigidified definite description. A possible response on behalf of Jackson denies that such idiosyncratic metaphysical opinions of the speaker constitute his understanding of the term but more basic and universally shared '[...] principles of theory choice [...]' (438). Nevertheless, any specification of this vague suggestion is refuted by further cases. More important, though, this response is flawed by origin, as it questions the value of knowing *A*-intensions.

Finally, the uniqueness clause of the description is challenged by counterexamples, too. Since salt well might be a mixture of *NaCl* and *AbCd,* associating a definite description, rigidified or no, with the term 'salt' is fallible [Block & Stalnaker, 1999, 18]. In response, Jackson could switch to a partial definition of 'salt' where two different but each again definite rigidified descriptions make up the respective *A*-intension [Block & Stalnaker, 1999, 21], [Schroeter, 2003, 4]. This, however, attributes overly strong cognitive capacities to speakers, as they are supposed to disambiguate infallibly the diverse meanings of natural kind terms. Alternatively, it might be suggested that the rigidified description merely captures the functional role independently of what salt consists of at the single worlds. Again, though, this move trivialises description.

To sum up these inquests, two-dimensional descriptivism falls prey to a dilemma when confronted with externalist counterexamples: Either it generalises the descriptions which speakers are supposed to know a priori such that this knowledge becomes trivial, or it enhances them to capture any far-fetched cases such that knowledge of them exceeds what can reasonably assumed to be human cognitive capacities. Considered by itself, the assumption that speakers have an a priori knowledge of *A*-intensions thus becomes implausible.

Jackson accordingly embeds his epistemological assumptions into a more general picture of language and communication. He sketches it in different ways [Jackson, 1994, 34][Jackson, 1998b, 202][Jackson, 2009, 391, 423f.], but essentially it amounts to three claims [Jackson, 2004, 266f]. First, he understands languages as sets of items by means of which a transmitter conveys

information to receivers. This requires both to associate with these linguistic elements possible ways things are. Second, among these associations needs to be a class of such which remain stable across hypothetical cases of application. Third, these associations are built into the meaning of the terms such that being a competent speaker suffices to know them. By an argument to the best explanation Jackson identifies these stable and a priori knowable associations with his *A*-intensions.

Admittedly, this picture provides Jackson's assumption of a priori knowledge of *A*-intensions with some plausibility. Nevertheless, this achievement turns out to be merely apparent, because Jackson's view on language and communication is by no means less controversial than the claim to a priority itself.

Various arguments have been raised in the literature, but for reasons of brevity I focus on a already known opponent of Jackson's case. Schroeter casts doubt on argument from language by proposing an alternative model of communication. It dispenses with any core set of resilient assumptions about the extension of '*F*', instead, the folk theory about Fs is continuously developed, changed and adjusted [Schroeter, 2006, 572]. Still, it can account for the difference between change of belief and change in meaning as well as for synonymy, since meaning is solidified by holistic, rationalising interpretations speakers undertake of their own linguistic practise. This 'jazz model' of communication [Schroeter & Bigelow, 2009, 102] is not only more economical than Jackson's account as it reduces linguistic competence to general heuristic abilities, it also captures better the psychological reality. And it does so without any commitment to speakers associating with a term a resilient class of properties which could be identified with *A*-intensions. Hence, Jackson's account is not the best explanation of language and communication and therefore fails to bolster his descriptivism.

4. Serious metaphysics without analysis

In the foregoing section, I have sketched the various obstacles Jackson's argument faces. The number of objections raised indicates how controversial his presumptions are and accordingly how weakly his overall argument is founded. To establish that only analysis allows the metaphysician to solve her placement problems, he commits himself not merely to a two-dimensional semantics for natural kind terms but to a full-blooded descriptivist theory of reference. As these indeed are strong, though contentious theories, he gets hold of a powerful philosophical machinery. It is therefore hardly surprising that based on these assumptions he arrives at the envisaged conclusion that speakers have a priori knowledge about the reference of natural kind terms.

Simultaneously, however, he inherits all the problems of these positions. Thus, his argument becomes considerably vulnerable, at more than one point. If only some of the objections above hold, Jackson's argument fails to show that analysis is essential for serious metaphysics. In any case, however, one has to admit that Jackson's argument rests on highly controversial claims.

It might be replied that hardly any philosophical argument is free of objections and that contentious assumptions do not yet disqualify the overall project. Instead, its value ought to be measured not so much according to its costs but according to the theoretical benefits it promises. The issue Jackson has started from, the need of serious metaphysics to identify entailment relations between theoretical and commonplace truths, is indubitably both relevant and urgent. Therefore, it might be argued, Jackson's argument can still be maintained as a valuable contribution and has to be considered seriously. This defence of Jackson's argument only holds, however, if no alternative solution for placement problems stands to reason.

In a series of articles [Kirk, 1996], [Kirk, 2001], [Kirk, 2006a] Robert Kirk addresses in an initially congenial way the commitments of serious metaphysics. As Jackson, Kirk argues that the global supervenience thesis which a serious metaphysical theory is equivalent to, compels the theorist to identify entailment relations between theoretical truths such as (1) and commonplace truths like (2). Since metaphysical entailment would not suffice (page 1), she needs to establish strict implication instead, such that the conditional '(1) → (2)' becomes a necessary truth.

Kirk and Jackson disagree, however, about which necessity is required. Whereas Jackson contrasts a posteriori with a priori necessity and thus gives analysis a prominent place (page 1), Kirk champions logical necessity [Kirk, 2006b, 529]. For the exemplary theoretical truth (1) to strictly imply the commonplace (2), the conditional '(1) → (2)' must be necessary thus that its negation '(1) ^ :(2)' is inconsistent [Kirk, 1996, 244], [Kirk, 2006b, 544] [Kirk, 2006b, 527]. In the remainder of this paper, I shall elaborate Kirk's approach and sketch an argument why this different conception of necessity offers an alternative route for entry by entailment which does not require a priori knowledge and therefore is not committed to any descriptivism.

To forestall misunderstandings, strict implication (hereafter: SI) covers a priori entailment if there is such. The inconsistency might be of a kind that analysis indeed yields a priori knowledge of the conditional. The crucial difference, however, is that SI can dispense with it. Jackson suggests that if merely metaphysical necessities do not suffice a priori knowable sentences make up the only strengthening available. I deny this dichotomy. There are sentences, and '(1) → (2)' is one of them, which are logically necessary but still not knowable a priori. This is so, because conceivability is not necessary for consistency. For any circle's circumference *c* and diameter *d*, $\ulcorner x = \frac{c}{d} \rightarrow x \in \mathbb{Q} \urcorner$ is necessarily true. Nonetheless, $\ulcorner x = \frac{c}{d} \wedge x \notin \mathbb{Q} \urcorner$ is conceivable, as speakers might grasp the concept of the ratio of a circle's circumference to its diameter and still not know that π is not a rational number [Kirk, 2006a, 533]. Generally, there are strict implications such that it is impossible for speakers to proceed a priori from knowledge of the antecedent to knowledge of the consequent. Therefore SI does not presuppose a priori entailment.

One might deny SI being a valuable alternative since the strict, but a posteriori implication '(1) → (2)' would fail to explain the step from (1) to (2) (page 1). I counter that the consistency necessary for '(1) → (2)' being a strict implication corresponds to a proof, that is a complete unit of explicit reasoning leading from (1) to (2). In fact, this may well be the very same reasoning as given above as an example for a priori entailment. SI therefore has the same explanatory potential as Jackson's a priori entailment, the only difference is that SI does not require (3') to be a priori knowable. Still, doubts might be raised based on the concern that such an account would presuppose and be committed to a certain calculus. However, the consistency at issue need not be proved in a formal system. Informal reasoning suffices to justify strict implication.

It might still be objected that SI eventually collapses into a posteriori entailment. If *p* implicates *q* strictly, it would be argued, such that the conditional $\ulcorner p \rightarrow q \urcorner$ is provable $(\ulcorner \vdash p \rightarrow q \urcorner)$ and completeness holds for *T* then $\ulcorner p \rightarrow q \urcorner$ is also true in all models $(\ulcorner \vDash p \rightarrow q \urcorner)$,which would mean nothing more than being true at all worlds. As this, however, is already given by *superT* (page 1), SI would beg the question and the metaphysicist would be where she started from.

This line of thought, however, goes wrong since it confuses models with possible worlds and therefore model-theoretic with metaphysical necessity. At best, a possible world may count as the domain of a model, which, though, still contains in addition its interpretation function which maps non-logical expressions into the domain. Therefore, if $\models p$, then *p* is true merely in virtue of its logical form, independent of its meaning. Truth in possible worlds, on the contrary, applies to interpreted sentences, such that if $\Vdash p$, then *p* is true because of what it says is the case at any possible world. Accordingly, $\Vdash p$ is not sufficient for $\models p$, as *p*'s truth may depend on its meaning. Hence, model-theoretic necessity is by far a stronger notion than metaphysical necessity, and SI does not beg the question.

In conclusion, SI is not committed to the two-dimensional descriptivism Jackson has developed in support of his a priori entailment. Accordingly, the various objections raised above do not apply. Nonetheless, SI gives a sufficient answer to the placement problems of serious metaphysics. In view of its serious and diverse difficulties and the availability of an alternative I conclude that Jackson fails to show why analysis should play an essential role in metaphysics.

References

[Block & Stalnaker, 1999] Block, N. & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *The Philosophical Review*, 108(1), 146. ArticleType: primary_article / Full publication date: Jan., 1999 / Copyright Â c 1999 Cornell University.

[Harman, 1994] Harman, G. (1994). Doubts about conceptual analysis. In M. Michael (Ed.), *Philosophy in Mind The Place of Philosophy in the Study of Mind* (pp. 43-48).

[Jackson, 1994] Jackson, F. (1994). Armchair metaphysics. In M. Michael & J. O'Leary-Hawthorne (Eds.), *Philosophy in Mind The Place of Philosophy in the Study of Mind* (pp. 23-42). Kluwer.

[Jackson, 1998a] Jackson, F. (1998a). *From Metaphysics to Ethics*. Oxford University Press.

[Jackson, 1998b] Jackson, F. (1998b). Reference and description revisited. *Philosophical Perspectives*, 12.

[Jackson, 2004] Jackson, F. (2004). Why we need a-intensions. Philosophical Studies, 118(1-2), 257-277.

[Jackson, 2009] Jackson, F. (2009). Replies to my critics. In I. Ravenscroft (Ed.), *Minds, Ethics, and Conditionals - Themes from the Philosophy of Frank Jackson*. Clarendon Press.

[Kirk, 1996] Kirk, R. E. (1996). Strict implication, supervenience, and physicalism. *Australasian Journal of Philosophy*, 74(2), 244-57.

[Kirk, 2001] Kirk, R. E. (2001). Nonreductive physicalism and strict implication. *Australasian Journal of Philosophy*, 79(4), 544-552.

[Kirk, 2006a] Kirk, R. E. (2006a). Physicalism and strict implication. *Synthese*, 151(3), 523-536.

[Kirk, 2006b] Kirk, R. E. (2006b). Physicalism and strict implication. *Synthese*, 151, 523-536.

[Kripke, 1980] Kripke, S. A. (1980). *Naming and Necessity.* Cambridge, Massacusetts: Harvard University Press.

[Laurence & Margolis, 2003] Laurence, S. & Margolis, E. (2003). Concepts and conceptual analysis. *Philosophy and Phenomenological Research*, 67(2), 253-282.

[Putnam, 1970] Putnam, H. (1970). IS SEMANTICS POSSIBLE?*. *Metaphilosophy*, 1(3), 187-201.

[Schroeter, 2003] Schroeter, L. (2003). Gruesome diagonals. *Philosophers' Imprint*, 3(3), 1-23.

[Schroeter, 2004] Schroeter, L. (2004). The limits of conceptual analysis. *Pacific Philosophical Quarterly*, 85(4), 425-453.

[Schroeter, 2006] Schroeter, L. (2006). Against _a priori_ reductions. *Philosophical Quarterly*, 56(225), 562-586.

[Schroeter & Bigelow, 2009] Schroeter, L. & Bigelow, P. (2009). Jackson's classical model of meaning. In I. Ravenscroft (Ed.), Minds, Ethics, and Conditionals: *Themes from the Philosophy of Frank Jackson* (pp. 85-109). Clarendon Press.

Saving Armchair Metaphysics from A Posteriori Problems

Kyle Mitchell

Introduction

In this paper I will claim that conceptual analysis can plausibly be held to play an essential role in "serious metaphysics" in spite of skeptical arguments concerning our epistemic access to A-intensions. Before arguing for this claim, I will present Frank Jackson's conception of "serious metaphysics" and show why Jackson thinks that doing conceptual analysis is a necessary part of doing "serious metaphysics". Furthermore, I will canvass Jackson's distinction between A-intensions and C-intensions, show the role this distinction plays in Jackson's account of conceptual analysis and explain why the thesis that we have a priori access to A-intensions is crucial to Jackson's program. Once this has been covered, I will present an argument against our a priori access to A-intensions and then show that this argument is too strong by providing two thought experiments. Next, I will suggest another argument against our a priori access to A-intensions from the a posteriori nature of our theories. However, I will show that this argument need not pose a problem for Jackson provided that Jackson's A-intensions consist of the right kind of description. In this way, because Jackson can evade the skeptical arguments, Jackson can still claim that we have a priori access to A-intensions and, therefore, that conceptual analysis can still be considered a necessary condition of "serious metaphysics".

Serious Metaphysics and the Location Problem

Metaphysics seeks to explain the world and what the world is like. Furthermore, metaphysics seeks a *complete* account of the world, such that everything in the world is explained in terms of a limited set of more or less basic notions. Otherwise, metaphysics would be involved in no more than drawing up big lists. For this reason, Jackson defines "serious metaphysics" as a metaphysics that explains the world and everything in the world in the terms of some *limited* vocabulary, where this vocabulary is the most relevant vocabulary to the metaphysical theory that describes the basic notions of the metaphysics[1]. Jackson notes, however, that if we are committed to "serious metaphysics", then we must also be committed to solving, what Jackson calls, the location problem.

In order to understand what Jackson means by the location problem, let us assume that physicalism is true. Because physicalism is an instance of "serious metaphysics", if physicalism is true, then the world and everything in the world can, in principle, be explained in the physical vocabulary, i.e. the vocabulary of the natural sciences. However, the vocabulary of the natural sciences does not *explicitly* contain statements about terms like "belief", "meaning", "consciousness", etc. Therefore, these kinds of terms are not explicitly a part of the physicalist's theory. In this way, statements about the terms not explicitly included in the vocabulary of the "serious metaphysics" will not be accounted for by the "serious metaphysics". This is an instance of the location problem and can be generalized for any "serious metaphysics". Call the set of all true statements in the limited vocabulary of a "serious metaphysics" the *T*-statements and the set of all apparently true statements not explicitly contained within that vocabulary the *D*-statements. If one

1 For example, in the case of physicalism, the limited vocabulary would be the vocabulary of biology, chemistry, physics and neuroscience. Furthermore, it is important to note that, while I have defined "serious metaphysics" in linguistic terms, "serious metaphysics" can be equally well defined in ontological terms in the following way: a "serious metaphysics" is a metaphysics that explains the world and everything about the world in terms of a limited set of entities.

is committed to doing "serious metaphysics", then one must show that *everything* in the world can be explained by the *T*-statements. Therefore, because "serious metaphysics" leads to the location problem, a theorist of a "serious metaphysics" has two options: 1) be an eliminativist about the objects the *D*-statements refer to or 2) show that the *D*-statements are somehow included in the *T*-statements.[2]

Entry by Entailment and the Need for Conceptual Analysis

Jackson believes that we need to offer an account of how the *D*-statements can be included in the *T*-statements. He does this by suggesting that while the *D*-statements may not *explicitly* be contained in *T*-statements, they may still be *implicitly* contained in the *T*-statements. Jackson, therefore, distinguishes the explicit and implicit parts of a story. For example, I may *explicitly* tell you that Glenn Branca is better than every other composer. However, in stating this I have *implicitly* told you that Glenna Branca is better than Mozart. This is because the explicit statement *entails* the implicit statement, affording the implicit statement a part in the story. In the same way, the *T*-statements can implicitly contain *D*-statements because the *T*-statements entail the *D*-statements. Hence, Jackson's solution to the location problem is to suggest that *D*-statements are entailed by the *T*-statements. This is what Jackson calls entry by entailment. Furthermore, because entry by entailment claims that the *T*-statements entail the *D*-statements and that, because of this, the *T*-statements provide a *complete* account of the world, Jackson is also committed to the ontological thesis that the entities *picked out* by the *T*-statements *supervene* on the entities picked out by the *D*-statements. In this way, a theorist of a "serious metaphysics" can claim that there is nothing over an above the entities picked out by the *T*-statements. Hence, Jackson suggests commitment to the following inter-world global supervenience thesis:

> B) Any world that is a minimal[3] *T*-statement satisfying[4] duplicate of the actual world is a duplicate simpliciter.

Therefore, B) is true if and only if at any world in which the *T*-statements are true, the *D*-statements are true as well. In this way, commitment to B) will prevent independent variation between the *T*-statements and the *D*-statements relevant to each *T*-statement satisfying duplicate of the actual world. Again, this is because the *T*-statements entail the *D*-statements. In this way, Jackson solves the Location Problem by suggesting that the *D*-statements can find a place in the story of a "serious metaphysics" by being entailed by that story[5].

However, if this is to be convincing, then Jackson must have *some* story to tell about *how* the *T*–statements entail the *D*-statements, for, as it stands now, there is an explanatory gap between showing that because the *T*-statements are true, the *D*-statements are true as well. According to Jackson, in order to fill this gap, we need to *define the subject*. Defining the subject is the *a priori* process of taking a term *K* and deriving the necessary and sufficient conditions for *counting as* a *K* by imagining the various possible situations in which something would count as a *K*. This process is guided by our intuitions concerning whether or not, if certain conditions obtained, these conditions would count as *K*. Insofar as our intuitions about *K* coincide with the folk intuitions about *K*, these

2 It is important to note that while eliminativism about some areas of discourse might be a plausible position, Jackson believes that, with respect to the location problem, eliminativism is not an option. For example, "rivers", "explosions", "buildings" and a variety of other terms are not explicitly described in the language of natural science. In this way, if eliminativism was a plausible solution to the location problem, then we would be committed to the belief that explosions, buildings, rivers, etc. do not exist and this is clearly false.

3 Where "minimal" suggests setting the *T*-statement satisfying nature of the world and doing nothing more.

4 Where satisfying the *T*-statements is making the *T*-statements true.

5 Jackson, F. (1994) Armchair Metaphysics. In Michael, M. & O'Leary-Hawthorne, J. (eds.) *Philosophy in Mind: The Place of Philosophy in the Study of Mind*, pp. 23-34.

necessary and sufficient conditions will isolate the folk theory of *K*. Furthermore, because this process is a priori it is a species of conceptual analysis[6]. Jackson maintains that, once our folk theory of *K* has been a priori defined, we will know that *K* is associated with a rigidified definite description[7] consisting of necessary and sufficient conditions for being a *K*. If *K* is not explicitly contained within of the vocabulary of the relevant "serious metaphysics", then knowledge of *K*'s description will explain, provided we have found some term in the vocabulary of the relevant "serious metaphysics" that satisfies *K*'s description, how *K* is actually contained within the relevant vocabulary. In this way, because statements about *K* would be included in the *D*-statements, Jackson can use conceptual analysis to explain how the *D*-statements are entailed by the *T*-statements. Therefore, conceptual analysis is a necessary part of solving the location problem and, hence, a necessary part of doing "serious metaphysics".

Two-Dimensional Semantics and the A Priori

The claim that discovering that the *T*-statements entail the *D*-statements occurs by *a priori* conceptual analysis might seem overly contentious. For example, say that "gold" is not in the vocabulary of a "serious metaphysics" and, therefore, statements about gold will not be contained within the *D*-statements; while the symbol "Au" is in the vocabulary of the serious metaphysics and, therefore, statements about Au are contained within *T*-statements. If Jackson is correct, then our a priori knowledge of gold as the actual stuff that plays the gold-role should be sufficient to determine the referent of gold, namely Au. However, as Putnam and Kripke have shown, our a priori knowledge is not sufficient to show that gold is necessarily Au, rather our knowledge of this necessity is an a posteriori matter. Therefore, it might be objected that the apparent fact of a posteriori necessity is sufficient to show that a priori conceptual analysis is not a necessary part of explaining entry by entailment and, hence, solving the location problem.

In response to this claim, Jackson distinguishes between two different kinds of intensions, or functions from worlds to extensions. This distinction arises out of the different ways in which one can consider possible worlds. C-intensions are functions from worlds to extensions where the actual world *w@* is taken as fixed and the intension is used to pick out extensions in counterfactual worlds $w_1 \ldots w_n$ with respect to *w@*. The C-intension, therefore, picks out the same extension in $w_1 \ldots w_n$ as it does in *w@*. In this way, because gold is Au in *w@*, the C-intension of "gold" will pick out Au in all counterfactual worlds, regardless of the properties or descriptions associated with "gold" at those worlds. This is the intension that concerns the Kripke-Putnam cases. A-intensions, by contrast, are functions from worlds to extensions in which whatever world the A-intension is being used to pick out an extension in is *taken to be w@*. Moreover, Jackson believes that the A-intension of a natural kind term like "gold" corresponds to a rigidified definite description: The actual *X* that plays the gold-role. Therefore, the A-intension of "gold" at a world w_1 where XYZ, instead of Au, performs the gold-role will pick out XYZ instead of Au. Jackson uses this distinction to evade the above criticism by claiming that all the criticism shows is that we do not have a priori access to C intensions. In spite of this, Jackson claims that we do have a priori access to A-intensions. Hence, we have a priori access to the rigidified definite description of "gold" and, therefore, we know a priori that:

C) Gold is the actual stuff that plays the gold-role.

Moreover, Jackson maintains that our a priori knowledge of A-intensions allows for our a priori understanding that the *T*-statements entail the *D*-statements. For example, if statements about "gold" are members of the *D*-statements and statements about "Au" are members of the *T*-

6 A paradigm instance of this would be the discourse on the Gettier cases concerning the necessary and sufficient conditions of knowledge.

7 A rigidified definite description would usually correspond to a conjunction of the stereotypical features of a referent, a sortal, a uniqueness clause and a operator that specifies the actual world.

statements, then we can come to know that statements about "Au" entail statements about "gold" in the following way:

A) Au is a precious metal. (Premise)
B) Au is the actual stuff that plays the gold-role. (Empirical fact)
C) Gold is the actual stuff that plays the gold-role. (A priori)
D) Therefore, gold is a precious metal.

Notice that the above argument is valid a priori. This is because, once we have a priori access to the A-intension in C) and know all the relevant facts about the terms *within* the *T*-statements (premise B)), we can discover a priori that the *T*-statements entail the *D*-statements. Thus, Jackson is able to vindicate a priori conceptual analysis and it's role in the entry by entailment thesis despite the Kripke-Putnam cases. In this way, Jackson can claim that conceptual analysis is essential to the entry by entailment thesis, solving of the Location Problem and, therefore, is a necessary part of doing "serious metaphysics". This claim, however, *depends* on the claim that speakers have a priori access to A-intensions[8].

Objection from Epistemic Access to A-intension Stereotypes

Laurence and Margolis (LM) object to Jackson's claim that conceptual analysis plays an essential role in "serious metaphysics", by suggesting that our epistemic access to A-intensions is not a priori, but rather a posteriori. LM claim that knowledge of

C) Gold is the actual stuff that plays the gold-role,

Requires knowing the stereotypical elements associated with gold. The gold-stereotype would presumably include that gold is a shiny-yellowish metal, traditionally involved in currency, etc. This is what knowledge of the "gold-role" consists in.

LM suggest that we cannot have a priori access to a description which picks out the referent of "gold" in each world *w* considered *w@*, because we don't even have a priori access to a description that picks out the referent of "gold" in the actual *w@*. This is because *all* the elements of a natural kind stereotype are open to revision in light of empirical findings. This is because 1) the stereotype for a natural kind term *might* be based on atypical or idiosyncratic samples and 2) the conditions of observation *might* affect the characteristics of the natural kind, therefore, allowing for these characteristics to change over time. For example, for all we know, the introduction of a new gas into the atmosphere at a future time *tF* might cause gold to have a dull-red colour rather than a shiny-yellowish colour. Moreover, scientists and historians might discover that the "gold" that has traditionally been involved in various economic matters was actually a kind of fools gold rather than Au. If these cases obtained, then we would need to revise our gold-stereotype. Jackson might suggest that, because only a sufficient number of the elements associated with gold need to be satisfied, the fact that *some* of the elements of the gold-stereotype are a posteriori revisable should not pose a problem for his view. However, LM suggest that, *all* of the elements of the gold-stereotype are in principle revisable in this way, thus, blocking Jackson's suggestion. Therefore, because revision of the gold-stereotype in light of empirical findings is a species of a posteriori knowledge, LM conclude that our knowledge of the gold-stereotype and, therefore, C) is not a priori but rather a posteriori. Therefore, because our knowledge of A-intensions is an a posteriori matter, conceptual analysis, conceived as an a priori process, is not a necessary part of "serious metaphysics"[9].

8 Jackson, F. (1998) *From Ethics to Metaphysics*, Oxford: Oxford University Press, pp. 29-85.
9 Laurence, M. and Margolis, E. (2003) Concepts and Conceptual Analysis**.** *Philosophy and Phenomenological*

LM's argument is, however, too strong. The argument is too strong because LM claim that *all* of the elements of the gold-stereotype are in principle a posteriori revisable. In order to show that this claim is too strong, I will first present a thought experiment showing that 1) definite descriptions of some sort are necessary for a agent to know the referent of a term and 2) that these descriptions need not be anything substantial in terms of stereotypical properties, they merely need to delineate *some* kind of role that the natural kind plays. The first thought experiment is as follows:

> TE1) Imagine two qualitatively identical steel spheres; the one on the left-hand side you have named "Jonny" and the one on the right-hand side you have named "Amanda". The spheres are then shuffled when you are not looking. Suppose that I ask you now which sphere is "Amanda" and which one is "Jonny". Can you refer in this case?

It should be obvious that, in this case, you will not be able to tell me which is "Amanda" and which is "Jonny" even though there is a fact of the matter that goes with the distribution of properties: Amanda is the one on the left-hand side that you named "Amanda" and "Jonny" is the one on the right-hand side that you named "Jonny". Moreover, the reason why you cannot know the referent in thismcase is precisely because you would have *no* description associated with either of the steel spheres[10]. It follows that 1) having *some* associated definite description about an object is a necessary condition being able to refer to that object. Furthermore, TE1) shows that 2) the description need not have anything to do with the stereotypical elements typically associated with the referent. All that is needed is that the description delineates *some* kind of role that the referent plays, in this case the role of being named either "Jonny" or "Amanda". In this way, Jackson's initial response to LM seems more plausible[11]: We only need *some* elements associated with the natural kind to be a part of the rigidified definite description.

We are now in a position to show that LM's claim is too strong. Remember, because LM claim that, in principle, *all* of elements of a natural kind's definite description can be revised in light of empirical findings, it should be the case that our description of gold could be completely revised and yet we would still be talking about gold. Consider the following thought experiment about another natural kind term "water":

> TE2) Scientists have declared that, as we all know, water is H_2O. But suppose that at some future time *tF* scientists discover that, contrary to what we thought, H_2O was not the stuff that filled the lakes, came from the taps or had *any* of the properties typically associated with water. Scientists even discovered that H_2O was not the object that caused us to say water when we talked about it and was not the thing that played the water-role in everyday life. H_2O is actually always a black gas, it never caused us to say anything until recently, plays no role in nourishment, etc.

Would we say that, provided the above obtained, in talking about H_2O, we are still talking about water? TE2) should make it clear that once *all* the elements of a referent's definite description have been revised in light of empirical findings, we would not say that we would still be talking about the relevant natural kind. Rather if TE2) obtained we would be compelled to say that we have changed the subject. In this way, LM cannot claim that all of the elements of a referent's rigidified definite

Research, 67, No. 2, 260-263.

10 Jackson, F. (2009) Replies to My Critics. In Ravenscroft, I. (ed.) *Mind, Ethics and Conditionals: Themes from the Philosophy of Frank Jackson*, Oxford: Oxford University Press, pp. 411-12.

11 Of course, this *alone* does not prove that Jackson's response is correct, for it could still be the case that parts of the definite description are all a posteriori revisable.

description can be revised in light of empirical findings. Furthermore, as TE1) has shown, we must have access to some kind of definite description that delineates some role that the natural kind plays that is a priori, and not a posteriori revisable in order to refer to the natural kind at all[12]. In this way, because LM's universal claim is false, LM's argument does not go through. Therefore, not only can Jackson claim that only a sufficient number of the elements associated with gold need to be satisfied in order to refer, but also that we can still have a priori access to C) and, therefore, to A-intensions in general.

Objection from Epistemic Access to the Theories that Determine A-intensions

While LM might not have convincingly shown that we do not have a priori access to A-intensions because the elements of a natural kind's rigidified definite description are a posteriori revisable, Laura Schroeter (LS) has presented an argument suggesting that we do not have a priori access to A-intensions because the *theories* which determine the rigidified definite description of a natural kind term are a posteriori revisable. Therefore, if LS is correct, then we cannot have a priori access to A-intensions and, therefore, conceptual analysis cannot plausibly be held to play an essential role in "serious metaphysics".

LS begins her argument with an analysis of the component parts of an A-intension[13] for a natural kind. She distinguishes between two distinct parts of a rigidified definite description: 1) a sortal and 2) an actual-world description. The sortal specifies what sort of object or property would qualify as a candidate for reference, while the actual-world description specifies the properties that must be satisfied in order for an object to fall into the extension of a concept in the actual extension of the concept. The actual-world description was the focus of LM's criticism. LS focuses on how one would come to know the sortal part of the rigidified definite description and, therefore, does not inherit the problems of LM's criticism. Furthermore, LS suggests that the sortal is a necessary part of an A-intension's rigidified definite description, for, if Jackson is to underwrite a priori conclusions about gold, then the analysis available to the subject on the basis of a priori reflection must make a substantive claim about the *kind* of object that gold is. This can only be done with a sortal.

Once this has been established, LS considers the kinds of intuitions that might a priori determine the nature of the sortal. Jackson further distinguishes first-order from second-order intuitions[14]. Our first-order intuitions are those intuitions which Jackson claims define the subject. By contrast, our second-order intuitions are our intuitions about how we should revise our first-order intuitions. These are, affectively, our best intuitions about how to theorize about natural kind terms. LS suggests that, in order to account for a sortal that is a) narrow enough to specify a determinate class of referents and b) broad enough to accommodate all the ways we think we might be mistaken about the nature of what we are referring to, our sortal must be determined by our second-order intuitions.

We are now in a position to articulate LS's claim that we do not have a priori access to A-intensions. LS claims that, because Jackson claims that we can know the rigidified definite descriptions for natural kind terms a priori, Jackson is committed to the claim that our best second-order intuitions are infallible, i.e. that they cannot be revised in light of empirical evidence. However, our second-order intuitions amount to our best theories about how to determine what gold is. In this way, because second-order intuitions determine what *counts as* a the kind of thing gold is, if we change our theory about how to determine what gold is, then Jackson is committed to that

12 It may be the case that certain elements of the natural kind are more essential than others to the rigidified definite description. My arguments do not take a stance on which elements these might be, for all I intend to do is show that LM's argument is false and that Jackson can keep his original claim: that we need a priori knowledge of an A-intension in order to refer.

13 LS calls this a natural kind's reference-fixing conditions.

14LS calls these first and second-order dispositions.

claim that we are changing the subject. LS, however, points out how implausible this claim is by considering the way Aristotle theorized about water. Aristotle had a radically different theory about the kind of thing that water is than we do today, and, therefore, Jackson would be committed to the claim that when Aristotle spoke of "water" he was not referring to what we refer to when we talk about "water" today. This, however, seems absurd. The more realistic story to tell is that our theories can, and should be, revised in light of empirical evidence and that, because theory determination is an a posteriori matter, we need not suggest that our second-order intuitions be infallible. Hence, we can still claim that, despite the difference in second-order intuitions, Aristotle referred to the same object that we do in our talk of water. In this way, because the theories that determine the sortal of the A-intensions associated with natural kind terms are a posteriori revisable, it follows that we cannot know the A-intensions of natural kind terms a priori[15].

Theory-nesting A-intension Response and Further Vindication of Conceptual Analysis

LS's objection, however, rests on the assumption that the sortal of the rigidified definite description of the A-intension of a natural kind is determined by our best theories. It is for this reason that Jackson's claim to our a priori access to A-intensions does not accurately account for our ability to refer in the face of changes in scientific theory. In this final section, I will show that by adopting a certain kind of rigidified definite description, one that *nests* theories and is not determined by theories, then Jackson can evade LS's argument and claim that A-intension can be known a priori.

David Braddon-Mitchell[16] has developed the kind of rigidified definite description that Jackson needs in order to save his claim. Consider our folk theory of a natural kind term, in the sense specified above, to be a level-1 theory *T1* that says that gold is whatever plays the actual gold-role by some possibly unknown true theory *TT*. Furthermore, let *P1, P2* and *Pt* be terms in theories *T1, T2* and *TT* respectively. We are now in a position to state the theory-nesting rigidified definite description of gold:

> TN) Gold is whatever plays the gold-role according to *T1* (Folk theory) that contains a term *P1* (gold) and a clause that associates with *P1* whatever second-order intuitions are associated with the term *Pt* of some true theory *TT* that explains the nature of what plays the gold-role actually[17].

If Jackson adopts the rigidified definite description in TN) as the A-intension of gold, then he can evade LS's objection. For example, call Aristotle's theory about gold *T2*. Aristotle would consider *T2 = TT*, for he would consider *T2* the true theory. Years later, empirical evidence suggests that *T2* is false and that our current theory *T3 = TT*. In this case, the a posteriori change in theory does not entail that in moving from *T2* to *T3* one changes the subject, for, in each case, the rigidified definite description refers to whatever actually plays the gold-role. Rather, only our second-order intuitions change from *T2* to *T3*. In this way, TN) can be said to nest theories. Therefore, contrary to LS's assumption, the A-intensions of natural kinds are not necessarily determined by our theories. Thus, a posteriori changes in theory do not present a challenge to our a priori access to A-intensions, provided Jackson adopts TN).

Conclusion

Therefore, because Jackson can avoid the skeptical arguments against our a priori knowledge of A-intensions presented by both LM and LS, Jackson can still claim that a priori conceptual analysis is

15 Schroeter, L. (2004) The Limits of Conceptual Analysis. *Pacific Philosophical Quarterly*, 85, 427-448.
16 Braddon-Mitchell, D. (2005) The Subsumption of Reference. *British Journal for the Philosophy of Science* 56, 157-60.
17 The rigidified definite description has been altered a bit to respond directly to LS's objection.

an essential part of doing "serious metaphysics". This claim, however, can only be made provided he puts certain constraints on the kind of rigidified definite description used as an A-intension. In this way, Jackson's vindication of conceptual analysis is, perhaps, much more difficult to debunk than the skeptics have assumed.

Bibliography:

Braddon-Mitchell, D. (2005) The Subsumption of Reference. *British Journal for the Philosophy of Science* 56, 157-78.

Harman, G. (1994) Doubts About Conceptual Analysis. In Michael, M. & O'Leary- Hawthorne, J. (eds.), *Philosophy in Mind: The Place of Philosophy in the Study of Mind,* 41-8.

Jackson, F. (1994) Armchair Metaphysics. In Michael, M. & O'Leary-Hawthorne, J. (eds.) *Philosophy in Mind: The Place of Philosophy in the Study of Mind*, pp. 23-42.

Jackson, F. (1998) *From Ethics to Metaphysics*, Oxford: Oxford University Press.

Jackson, F. (2009) Replies to My Critics. In Ravenscroft, I. (ed.) *Mind, Ethics and Conditionals: Themes from the Philosophy of Frank Jackson*, Oxford: Oxford University Press, pp. 387-475.

Laurence, M. and Margolis, E. (2003) Concepts and Conceptual Analysis**.** *Philosophy and Phenomenological Research*, 67, No. 2, 253-80.

Lycan, W. (2009) Serious Metaphysics: Frank Jackson's Defense of Conceptual Analysis. In Ravenscroft, I. (ed.), *Mind, Ethics and Conditionals: Themes from the Philosophy of Frank Jackson*, Oxford: Oxford University Press, pp. 61-85.

Schroeter, L. (2004) The Limits of Conceptual Analysis. *Pacific Philosophical Quarterly*, 85, 423-53.

Stalnaker, R. (2001) Metaphysics Without Conceptual Analysis. *Philosophy and Phenomenological Research,* Vol. 62, No. 3, 631-6.

H.P. Grice and the Great Pragmatics Predicament

Fenner Tanswell

H. P. Grice is widely accredited with the discovery of implicature, that which is not literally said by a sentence but is nonetheless conveyed when used in a conversational context, creating a theory which has had a tremendous influence on the study of pragmatics. However, in this essay I shall be arguing that beyond the very intuitive notion that implicature exists, the system that Grice constructs to explain and predict it in *Logic and Conversation*[1] is incomplete in several devastating ways, which eventually leads to the need for its extensive refinement and additional elements to form a complete whole. It is often commented that Grice's system is too vague to commit him to anything, however, Grice does make several definite claims and it is my aim to show that these cannot properly or fully characterise how implicature work.

I shall now first discuss how the different elements of Grice's project are used to construct a theory explaining implicature and its use in language. Grice's system of implicature focuses particularly on what he labels conversational implicature, characterised chiefly by his Cooperative Principle and its four associated maxims, which is a "subclass of non-conventional implicature"[2]. The Co-operative Principle is the following rough guide Grice lays out: "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged"[3]. Grice believes that fundamental to our ability to create implicature in conversations is the fact that they are not a disjoint series of sentences, but mutually inter-related contributions to a collective effort. He argues that we can distinguish these cooperative linguistic transactions by the fact that participants have a common immediate aim to appropriately exchange the correct information, that contributions are dependent on other participants and responsive to their input and that both parties expect the conversation to continue in a suitable manner unless both parties are agreeable that it should terminate.

The assumption that the Cooperative Principle is being observed can then be combined with his more specific maxims to deduce what is being implicated in a sentence. These maxims are Quantity, Quality, Relation and Manner and are described thus:

Quantity: Use the correct level of informativeness when speaking, namely:
1) Make your contribution as informative as is required (for the current purposes of the exchange).
2) Do not make your contribution more informative than is required.[4]

Quality: Try to make your contribution one that is true, namely:
1) Do not say what you believe to be false.
2) Do not say that for which you lack adequate evidence.

Relation: Be relevant.

Manner: Be perspicuous, namely:
1) Avoid obscure expressions.
2) Avoid ambiguity.
3) Be brief (avoid unnecessary prolixity).
4) Be orderly.

1 Putnam, H. P., "Logic and Conversation" in A. P. Martinich (ed.) The Philosophy of Language pp. 171-181.
2 Ibid., pp. 173.
3 Ibid
4 This second part of the maxim of Quantity is not held to be as certain by Grice, who suspects that the same effect is assured by the maxim of Relation.

Another vital component of Grice's scheme is the ability of either participant to fail to fulfil a maxim. This can occur when a person violates a maxim quietly either intentionally to mislead or unwittingly; when a person makes it clear that they are opting out, such as in the case of secret-keeping; when a clash of maxims occurs forcing the person to violate one; or finally they may intentionally flout a maxim by blatantly failing to fulfil it.

Now Grice can define formally that a person A saying p has conversationally implicated q if and only if:

1) A can be presumed to be following the Cooperative Principle.
2) In order for p to be consistent with (1), it must further be supposed that A is aware that, or thinks that q.
3) A thinks that it is within the capacities of the listener to correctly deduce or intuitively grasp that (2) is required.

Furthermore, Grice employs what we shall refer to as the Calculability Condition, "the presence of conversational implicatures must be capable of being worked out; for even if it can be intuitively grasped, unless the intuition is replaceable by an argument, the implicature (if present at all) will not count as a conversational implicature; it will be a conventional implicature"[5].

Conversational implicature in its simplest terms then is just an acknowledgement that in conversation not everything needs explicit statement to be understood by both parties. A basic example of conversational implicature may be:

Anne: I'm thirsty.
Brian: There's some milk in the fridge.

Analysed just on what is explicit in Brian's utterance, it would be concluded that it is unrelated to Anne's. However, if it can be assumed for condition (1) that Brian is obeying the Cooperative Principle then we can see that in order for Brian's input to be consistent with it we must suppose that Brian is conversationally implicating something so (2) holds. By the maxim of Relation, we can suppose that Brian following the Cooperative Principle implies that his statement is relevant, therefore Anne can conclude that Brian means and is implicating that the milk in the fridge would quench her thirst, that it hasn't gone off, that she is allowed to drink it and all things which seem naturally implied by Brian's statement. These all seem obvious so clearly condition (3) holds, as Brian could reasonably expect Anne to work all of these implicatures out.

Grice also argues that conversational implicature can be generated by the flouting of a maxim of conversation. This occurs when it can still be assumed that the person you are talking to is obeying the Cooperative Principle but yet they blatantly flout one of the maxims, so for condition (1) to be consistent with what has been said it can be further assumed that their flouting of the maxim was being employed in order to conversationally implicate something. We shall briefly give examples of how this may be done for each maxim. The classic Gricean example of flouting the maxim of Quantity is that of the philosophy professor being asked to provide a reference for a bad student and writes "the student has excellent handwriting and regularly attended lectures", thereby providing insufficient information to satisfy the maxim of Quantity, thus implicating that there nothing better to say about the student without breaching the social convention of not writing negative references. An example of a breach of the maxim of Quality is sarcasm, where someone says something blatantly untrue with the implicatum that the opposite is in fact true, like "joining scientology is a really good idea" or "Britney Spears has produced some excellent music". An example of breaking the maxim of relation may be when the subject is changed suddenly, with the implicature that a taboo has been broached or that a person that was being talked about has just

5 Ibid., pp. 174.

entered the room. Breaking the maxim of Manner could be exemplified when adults are speaking in the presence of a child, if they speak in intentionally obscure language, then it may implicate that what they are saying is not intended for the child to understand.

We have thus laid out Grice's theory and seen how he believes that conversational implicature is generated. I shall now argue that despite the apparent plausibility of this system and the seemingly successful examples of its application, there are several key reasons that Grice's schema fails to hold generally. Firstly, the problem of differentiation shows examples of what is calculable as implicature according to Grice's schema but intuitively isn't. Secondly, the mutual knowledge assumption does not avoid this sufficiently. Thirdly, despite the crucial role Grice gives to the Calculability condition, there is no need for conversational implicature to obey it. Fourthly, the Cooperative Principle needs to be simply assumed to be being followed, but this assumption may plausibly fail. Lastly, the Cooperative Principle is an arbitrary way of distinguishing conversational implicature from similar phenomena which can be created without it, suggesting that it is not the reason for the generation of this type of implicatures.

The most widespread criticism of Grice's system is advanced by W. Davis, who argues that "for nearly every implicature correctly predicted by Gricean theory, others are falsely predicted"[6] and that conversely "implicatures exist that cannot be derived from conversational principles"[7]. Davis' criticism focuses mainly on the failure of the maxim of Quantity to be sufficient to explain implicatures which Grice's system would need the maxim to calculate. Consider, for example, Antti saying "Some papers at the Reading Party were good". By the maxim of Quantity, operating a Gricean calculation on this, it would be reasoned that if Antti can be assumed to be following the Cooperative Principle then he will be trying to be as informative as is appropriate for the current purposes of the conversation, so if it had been the case that all of the papers were good he would have said so, therefore we can take Antti to have implicated the denial of the stronger statement "All papers at the Reading Party were good". However, the problem of apparent plausibility applies here because although it appears that the reasoning is sound and that Grice's theory has adequately explained and predicted the implicature, we can easily consider other cases where like reasoning is applied to like cases but yet the derived prediction of implicature would be false. For instance, take the following sentences:

> All but one of the papers at the Reading Party were good.
> Three of the papers at the Reading Party were good.
> The papers by the German students at the Reading Party were good.
> Half of the papers at the Reading Party were good.

All of these are stronger statements and more informative than "Some papers at the Reading Party were good", but the fact that Antti used this sentence does not imply the denial of any of these other sentences in the same way that it implicates the denial of "All papers at the Reading Party were good". However, the reasoning is the same so by Grice's system these implicatures should also hold. It would be spurious to try to argue against this from any of the other maxims: it can't be said that these alternatives would have been over-informative so the implicature doesn't hold for them, since they are just as appropriate to the conversation as what Antti did say. It equally can't be said to be breaking the maxim of Manner for not being brief, since none of these sentences are particularly lengthy or complex. It seems we can conclude that Grice's theory over-predicts implicatures so we can accuse Grice of post hoc reasoning in the cases where the implicature of a weaker statement to the negation of a stronger statement is invoked because Grice would use the cases in which this is successful to support his claim without discussion of the cases which fail to hold. It could be argued

6Davis, W., "Implicature: intention, convention, and principle in the failure of Gricean theory", pp. 33.
7 Ibid.

that attributing successful implicatures to the maxim of Quantity is therefore incorrect because if the maxim was in general operation it would lead to far more incorrect implicatures being predicted.

We could consider the maxim of Quality to suffer the same problems: Grice attributes sarcasm, irony, metaphor, understatements, exaggeration and hyperbole all to flouting the maxim of Quality and thereby using implicature to create these phenomena. However, having already seen the problems for the maxim of Quantity, it seems unlikely that the same reasoning will hold in all cases that Grice's theory would predict it for. Our original example of flouting the maxim of Quality was sarcasm in which something blatantly untrue was used to imply the opposite, but it is straightforward to find examples where this doesn't hold. For example, if I say to my sister "Your face is absolutely hideous", this is obviously not true but doesn't implicate the opposite, instead it suggests I am trying to annoy her. Grice might respond to this that he acknowledges there may be "all sorts of other maxims (aesthetic, social or moral in character) such as 'Be polite,' that are also normally observed by participants in talk exchanges, and these may also generate non-conventional implicatures"[8]. Then it could be argued that when talking to my sister I was flouting the maxim of Politeness to annoy her. However, this response is inadequate since clearly what I am saying is also violating the maxim of Quality as already stated, so Grice's theory provides no clear reason to apply one maxim rather than the other when calculating the implicature. The only way that it is decided that in one case I was being sarcastic while in the other I was annoying my sister is because we intuitively decide which is which, not because Grice's system has any way of doing this. In fact, it seems that the Calculability Condition needs an extension such that it is not just required that there is a way of working out the implicature by argument, but also that there is only one applicable deduction in each case[9].

It has been replied to the general problems of differentiation that these can be overcome by contexts and mutual knowledge, which are included by Grice in his system. However, I will now argue that this is insufficient to make Grice's system a comprehensive theory of implicature. For example, the argument could run that in my earlier example of Antti implicating which papers were good at the reading party, it could be the case that if there was mutual knowledge between him and his listener that Antti hates Germans then he could have been implicating that it wasn't the case that the papers by the German students were good[10]. However, we could now ask how the listener is to conclude whether Antti was just implicating the negation of all papers being good, or whether he meant any one of the others: the mutual knowledge that Antti hates German students could give his listener another viable option for what Antti might be implicating but it does not in any way convince us that the listener will figure it out correctly. Mutual knowledge does seem to be essential for implicature to be created: without the mutual knowledge of particular facts, certain interpretations of what is being implicated by the speaker will seem less viable. However, it does not seem that this rescues the theory because it still does not avoid the need for post hoc reasoning where the correct implicatum is needed to be known before an adequate description of its calculation can be given, so it seems that Grice's project is incomplete on this point. Mutual knowledge does not directly imply that contextual implicatures will be deducible.

The next objection we have to offer against Grice's system is directed at the Calculability Condition. There is a curious gap in Grice's argument between speaker-meaning and listener-interpretation. It may be argued that conversational implicature is predominantly an act by the speaker: that although it is standard for it to be used in a Cooperative effort there is no strict need for it to be worked out. For example, someone smug may, in a conversation fully abiding by the Cooperative Principle, subtly implicate things that they are nearly certain the listener won't pick up

8 Grice, pp. 174.

9 There is the problem here that a sentence may be used for different things in different contexts ergo there shouldn't be a unique deduction clause. However, it does seem that for a sentence in one context there should be only one argument to apply to it in order to calculate the correct implicatum. Yet, literature and poetry often make use of these ambiguities in language, so maybe our new restriction would have to be in some way limited.

10 This response was raised by Antti himself at the reading party!

on just to feel self-satisfied and witty. It does not strike us that what would count as conversational implicature if the listener was more intelligent fails to be in this case because this isn't the case. The problem for Grice then resides in the fact that his calculation schema is directed at the listener, while instead, as H. H. Clark puts it "paradoxically, he expressed the maxims as exhortations to speakers"[11]. If the listener is incapable of grasping the implicature, and furthermore by Grice's condition replacing that intuition by a reasoned argument from the Cooperative Principle and the maxims, then Grice claims that it is not conversational implicature but conventional implicature. Yet it seems that what counts as conversational and conventional implicature should not be determined by the listener's ability to interpret, so Grice has been lead to an unappealing result. Also, there are many situations where listeners pick up on implicatures that the speaker may not have been initially aware of, such as accidental puns and innuendo, which would be inconsistent with Grice's scheme if required to be conversational implicature. Yet the only other person we could allow to be the judge of calculability is the speaker herself which cannot be acceptable or else thinking you were conversationally implicating something would be the same as actually implicating it. For example, a speaker who is distracted may leave out a vital part of the sentence in which he would have been using conversational implicature, thus believes he is implicating something while actually not doing so.

It could be argued then that the Calculability Condition in these problem cases should not be about either the listener's or speaker's capability to calculate, but rather that of some kind of objective impartial observer. However, it would be undesirable to introduce the need for an impartial observer, since it unclear as to the calculating power that this observer should be endowed with. For instance, if we take the example of a listener picking up on additional implicatures from before, but suppose that for the same sentence the listener does not pick up on this accidental innuendo: if the impartial observer is capable of working this out then there is conversational implicature that neither the speaker nor the listener are aware of, which seems flawed. The best suggestion along the same lines is the suggestion that it is not an actual observer so the condition could be phrased as "what a reasonable person could work out". This seems the best option we have considered, but is beyond the scope of Grice's own position and may give rise to further dangers in introducing a modal notion to characterise implicatures.

A further objection to Grice is that condition (1) is unashamedly phrased as a necessary presumption for any conversational implicature to occur. Yet it doesn't seem to be so unusual for at least some part of the Cooperative Principle to fail: for instance, someone may not be making their conversational contribution as is required in order to mislead or withhold information, or they may not accept the same direction or purpose as you do. This does not seem to be so extreme as to be a sceptical argument, since it actually strikes me as fairly commonplace that people are lying, holding back, talking at crossed purposes or past each other. To simply presume that this isn't occurring in order for the theory to work seems optimistic at best, naïve and unlikely to hold true. Also, it then seems that despite a failure in one of these ways, the cooperative elements required for conversational implicatures to hold may still occur, so it may be the case that Grice's definition of conversational implicature and how it is generated may need refining. How to do this within his system, however, is not entirely clear.

In fact, there may be more to this story. I shall now contend that the phenomena that Grice is trying to characterise within his system actually occur outside of its scope too, showing that Grice has over restricted his theory and that it thereby cannot be a full embodiment of the phenomena he is attempting to explain. More specifically, whichever way you interpret the Cooperative Principle it cuts out too many of the cases in which implicatures that intuitively belong in Grice's schema are present. I shall look at examples of the cross-examination of witnesses and interrogations. In both examples it is actually reasonable to expect uncooperativeness as described in the previous

11 Clark, H. H., "Using Language", pp. 142.

paragraph. In the case of cross-examination, consider a witness who is trying to avoid incriminating himself being examined by a lawyer:

Lawyer: What were you doing on the night of the 13th?
Witness: I was at a party until eight. I had work in the morning.

Although it is clear that the witness may be being vague deliberately to avoid incrimination, so has intentionally not explicitly stated that he went home for work in the morning, it also seems clear that there is implicature to the effect that the witness wants us to think that. However, the Cooperative Principle could be argued to not be in effect: the witness is definitely not accepting the same purpose as the lawyer as one wants to incriminate the other while the other is actively trying to avoid this happening. Furthermore, the witness clearly didn't make his contribution as is required by the lawyer's purpose for the conversation. It could be responded by Grice that in his theory conversational implicature is only a subclass of non-conventional implicature, and that this type of situation is in a separate subclass. However, this response doesn't seem to hold because here the method of generating implicature seems to be the same as in a situation Grice would describe as conversational implicature- the lack of Cooperative Principle has changed the context but not really affected the implicature itself, why then must it be moved to a different subclass of implicature? Grice's requirement for the Cooperative Principle to hold seems an arbitrary categorisation of implicature types.

A further response to our example could be that although the witness and the lawyer had differing motivations in the cross-examination, they both accepted some general direction of the talk like "the completion of the cross-examination in a just manner" which was furthermore the accepted purpose due to there being a court full of people as witness to this being completed. This exposes an ambiguity in the Cooperative Principle: who exactly is meant to be "accepting" the purpose of the talk? Is it the speaker, the listener the statement is aimed at, the extended crowd of all those listening or some combination of all of these? Although Grice would clearly prefer the latter option, our example has already demonstrated that this needn't hold since implicature was generated despite a lack of cooperation. To consider an even less clear example, I shall examine a torture scenario, where once again there is an attempt to extract information from a subject. This time there is no crowd to coerce the witness into an unwilling cooperation. In fact, even Grice's most general statement that "at each stage, some conversational moves would be excluded as conversationally unsuitable" doesn't seem to flow since a subject being tortured may come out with all kinds of gibberish at any stage, but yet this doesn't mean that the subject couldn't alternate between that and sentences which do conversationally implicate. The Cooperative Principle thus seems untenable as a condition for conversational implicature to hold to necessarily, especially if it to include all of the conversational phenomena which seem appropriate to include as the same type of implicature. Once again, the Cooperative Principle appears to be an arbitrary categorisation of implicature types.

In conclusion, we have shown that there are several crucial problems for the claims Grice makes. Firstly, the problem of differentiation shows that although there may be maxims at work, the ones that Grice makes explicit only capture a part of how these implicatures are created. Furthermore, there is a clear need for an explanation of how we do figure out implicatures correctly according to Grice's schema. Secondly, we showed that although theoretically mutual knowledge and context should play an important role in implicatures, the way they are invoked by Grice does not seem to capture their importance. Thirdly, the Calculability Condition is clearly one that cannot be sustained: not only is it unclear precisely who should be able to create the "reasoned argument" Grice requires, but it also seems perfectly plausible for the actual use of language to be of such complexity for this condition not to hold despite our ability to intuitively figure out implicatures. Finally, the Cooperative Principle is something that Grice strongly commits himself to, but yet does

not seem to be the crucial element for distinguishing non-conventional implicatures. So it can be concluded that although there are parts of Grice's theory which do capture aspects of how implicatures work, overall the theory is incomplete and several of the commitments that Grice thinks conversational implicatures do have to make do not hold up to closer scrutiny.

Bibliography

Clark, H. H., *Using Language*, Cambridge: Cambridge University Press, 1996.

Davis, W., *Implicature: intention, convention, and principle in the failure of Gricean theory*, Cambridge: Cambridge University Press, 1998.

Grice, H. P., 1975 "Logic and Conversation" in A. P. Martinich (ed.) *The Philosophy of Language*, 5th Edition, Oxford: Oxford University Press, 2008, pp. 171-181.