Warwick Graduate Conference in the Philosophy of Mind

27–28 November 2004

organised by
Jenny Booth, Alli Mitchell and Gloria Ayob

www.warwick.ac.uk/go/mindconf

**in association with the
Consciousness and Self-Consciousness
Research Centre**

www.warwick.ac.uk/go/consciousnesscentre

THE UNIVERSITY OF
WARWICK

Department of Philosophy

Contents

Social Externalism, Segal and Contradictory Intuitions

Alex Buckley

Department of Philosophy, University of Leeds

phl1ajb@leeds.ac.uk

In his recent publication, *A Slim Book About Narrow Content* (2000)[1], Gabriel Segal has attempted to argue that the social externalism of Tyler Burge is inadequate as an account of the fundamental nature of folk psychological explanation. The first part of this paper will comprise a rebuttal of this argument, on the basis that Segal assumes the truth of a disputed premise. In part two I will discuss the ramifications of the rebuttal of the argument. Briefly: the rebuttal leaves both internalists and externalists with a messy situation concerning our intuitions, and the natural way to create some order can raise problems for the social externalist. So the position expressed in this paper as a whole is that whilst Segal's argument, as it stands, fails to seriously call social externalism into question, it does indirectly mount a serious challenge. Finally, I shall outline a response the externalist can make.

I – Segal and Social Externalism

1. Segal's argument

Segal's argument is intended to show that a certain form of social externalism, attributed to Tyler Burge, cannot give an acceptable account of belief ascription when faced with certain crucial puzzle cases. "Consumerism" (the label is from Kaplan 1989) is the theory that one's words have the meaning they do because of the linguistic conventions of the community of which one is a member. One's ability to be a functioning (competent and appropriately deferential) member of that linguistic community endows one's words with certain properties, and affects the beliefs and concepts one may correctly be attributed.

> Words come to us prepackaged with a semantic value. If we are to use *those words*, the words we have received, the words of our linguistic community, then we must defer to *their* meaning…To use language as language, to express something, requires an intentional act. But the intention that is required involves the typical consumer's attitude of compliance, not the producer's assertiveness. (Kaplan 1989, 602)

This position is exemplified in the intended conclusion of Burge's well-known Twin-Earth thought experiments: intrinsic duplicates can be in different mental states, purely because of a difference in the linguistic practices of their communities. What beliefs a person has is not fixed by what is going on in their heads – social environment plays a determining role. Whilst

---

[1] All unattributed references are to this work.

acknowledging the intuitions that support such a conclusion, Segal believes it is a mistake to rely on them:

> There is, indeed, a tendency to think that…Alf believes that he has arthritis in his thigh, while Twin Alf does not. The mistake is to take this tendency too seriously and then to draw from it a general conclusion about the extension of concepts. (122)

The scenario Segal claims to demonstrate the inadequacy of consumerism is a variation on Burge's own thought-experiment. "Arthritis" and "inflammation of the joints" are synonymous, and so anyone, and in particular any expert, who knows enough about medical ailments to know that they are synonymous will express the same concept with the two expressions. Alf is not such an expert, and is not aware of the synonymy, and actually thinks he has arthritis in his thigh. Not only, therefore, is he not aware of the synonymy of "arthritis" and "inflammation of the joints", he would deny that, if asked. "Alf therefore has two different concepts that he expresses by 'arthritis' and 'inflammation of the joints'" (65). Where the experts have *one* concept, Alf has *two*. But given the assumptions that a) Alf has some minimal competence with the term "arthritis", and can use "inflammation of the joints" to mean what it should mean, and b) Alf would certainly defer to a room full of arthritologists, consumerism tells us that Alf gets to express the same concept/s with the expressions as the experts do. But plainly, Segal argues, he doesn't. The "prepackaged semantic value" of "arthritis" and "inflammation of the joints" is the same, and yet Alf clearly attaches different semantic values to the expressions. So what consumerism tells us is false. Despite his competence Alf has his own idiosyncratic concept that he expresses with the term "arthritis", that he does not share with the experts.

It is clear that the crucial step in Segal's argument is the claim that by not realising that arthritis is just an inflammation of the joint, Alf comes to express different concepts with the expressions. The support for this step is not explicit in Segal's exposition of the argument, but we can plausibly say that it consists in the following thought: Alf will assent to the sentence "I have arthritis in my thigh", and will assent to the sentence "I do not have an inflammation of the joint in my thigh". Alf is, by presumption, rational, and would not assent to both "a is F" and "a is not F", and so Alf uses the expressions "arthritis" and "inflammation of the joint" to express different concepts. Whilst this key claim is implicit and unsupported in Segal (2000), it is cited in a different context:

> *Fregean Principle of Difference*: If a subject, s, rationally assents to $P(t_1)$ and dissents from or abstains on the truth value of $P(t_2)$ then t_1 and t_2 have different meanings in s's idiolect and s associates different concepts with them. (2003, 417)

With such a principle included, Segal's argument consists in the claim that the following premises lead to contradiction.

> *Consumerism*: if a subject is minimally competent and deferential with an expression then they express the same concept with that expression as experts.

> *Synonymy*: experts in the use of synonymous expressions express the same concept with these expressions.

> *Fregean Principle of Difference* (**FPD**): if a rational subject sincerely and on reflection assents to "a is F" and "b is not F" then the subject expresses different concepts with the expressions "a" and "b".

In response to Burge's (1978) claim that synonyms are not inter-substitutable in belief attributions, and thus denial of *Synonymy* above, Segal formulates a version of the argument without that premise: a deferential and minimally competent subject thinks that one term, for instance "coffee", expresses two different concepts, and assents to "all coffee [type 1] comes from exotic countries" and "all coffee [type 2] does not come from exotic countries". By FPD, the subject expresses two concepts with the term "coffee", where the experts surely express one. But again, given the minimal competence and deference that the subject displays despite misunderstanding the term, Consumerism gives us a conflicting answer.

Segal's argument is thus that FPD and Consumerism together entail a contradiction. The fact that FPD is implicit and unsupported is therefore problematic. In another place he claims that it is a "[n]ormal, roughly Fregean, criteri[on] for the individuation of meaning" (2003, 416, my emphasis). I suggest that to view Segal's argument as plausible, we should charitably interpret him as claiming that the principle is an intuitively supported rule of normal practice, and that Consumerism as a philosophical theory should therefore be abandoned.

Consumerism as characterised above is a much simplified version of Burge's full account of belief and concept attribution[2], but on certain understandings of the key terms "competence" and "concept", Burge would endorse it fully. The reading of "competence" under which Burge would endorse Consumerism is the same *minimalist* reading that is used by Segal. The readings of "concept" that a) Segal is working with and b) under which Burge would endorse Consumerism, are more complex, and will be the focus of later discussion.

We should note that on Burge's account the reference to "experts" is inessential. Burge holds that a competent speaker defers to their linguistic community, whether or not there are experts for all particular terms used by that community. So we can say that through their competence the speaker expresses the *standardly* expressed concepts. Given that experts are competent speakers, Burgean Consumerism is a broader thesis than the consumerist principle stated above – the former entails the latter. Construed in that way, Consumerism is *prima facie* similar to the *Disquotational Principle* of Kripke (1979):

> DP: if a normal English speaker sincerely assents to "p" then they believe that p.

---

²   As given primarily in (1979), (1986) and (1993).

2. Objection to Segal's argument

Under the charitable interpretation suggested above, Segal's argument relies for its force on the strength of the claim that the Fregean Principle of Difference is intuitively powerful (maybe even obvious) whilst Consumerism is a controversial philosophical theory.

However, in employing our intuitive reactions to the famous Twin-Earth thought-experiments as support for their thesis, what externalists claim is that those intuitions reveal that our normal procedures for ascribing beliefs are deeply socially externalist. Not only that, but many externalists also claim that the truth of social externalism, as a principle of normal practice, casts significant doubt on principles such as FPD, which rely heavily on the controversial claim that we have privileged access to the contents of our beliefs. That is, some externalists claim that principles such as FPD depend on a *philosophical theory*, most famously formulated by Descartes, about the extent to which our mental life is transparent to us.

Given all this, it is possible that a social externalist may propose the very same argument as Segal, showing that FPD and Consumerism contradict, and draw the exact opposite conclusion: FPD, and its philosophical supports, must be abandoned. One example of such an argument is that put forward by Joseph Owens. He begins by accepting, like Segal, that FPD conflicts with social externalist principles such as Consumerism and the Disquotational Principle. He then highlights the extent to which our normal practice of belief ascription conforms to social externalism: if one examines normal practice, the Disquotational Principle is revealed as fundamental to the way in which we ascribe beliefs.

> It is hard to see how someone could reasonably dissent from [the Disquotational Principle], suitably qualified; it seems to be a part of our customary understanding of what is involved in sincerely assenting to a sentence one understands. Kripke goes so far as to say that [it] appears to be a self-evident truth. (1989, 291)

If one looks at The Fregean Principle of Difference, the situation is different – what is certainly undeniable is that no rational person would sincerely and *knowingly* assent to contradictory beliefs. But FPD requires more than that, it requires that a rational subject will not sincerely and *on reflection* assent to contradictory beliefs. To move to this stronger principle requires the claim that through reflection, a subject has the "more primitive ability to introspectively determine sameness and difference in belief" (1992, 158). Owens argues that this principle is far from intuitively obvious, and situates it firmly within a school of philosophical thought, calling it the "Cartesian model of access" (1992, 147).

Having claimed that the Disquotational Principle is supported by our intuitive understanding of normal practice, and that FPD ultimately relies on a (controversial) Cartesian philosophical thesis, Owens rejects FPD on those grounds. The intuitive force of the Disquotational Principle is such that if the Cartesian model of access contradicts it, then so much the worse for that:

> These principles [such as the Disquotational Principle] serve to "define" or at least delimit a non-Cartesian notion of belief, and to the extent that we operate with them we ought to abandon the conception of ourselves

as having some simple introspective way of discerning contradictions among our beliefs. (Owens 1989, 295)

In summary, Owens accepts two of the claims that Segal appears to make: a) that FPD and social externalism contradict, and b) that if a controversial philosophical claim contradicts our intuitions about normal practice, then it should be rejected. However, whereas Segal claims that social externalism is controversial philosophy and FPD is obviously true, Owens ultimately makes exactly the opposite claim. Their arguments share a common structure and some common premises, but give mirrored conclusions. For Segal, social externalism is a philosophical thesis whilst FPD is a principle of normal practice. For Owens, FPD is a philosophical thesis whilst social externalism is a part of normal practice.

Segal's argument, as presented by me, relies almost entirely on the intuitive force of the Fregean Principle of Difference: on the claim that it is so central to the way we think about belief, that a principle such as Consumerism that conflicts with it must be discarded. But what is clear from arguments such as Joseph Owens' is that many externalists deny the intuitive force of FPD. In fact, Owens makes a persuasive case that an acceptance of the force of the Twin-Earth thought-experiments can lead directly to a denial of the force of FPD. Segal's argument ultimately begs the question against the externalist, by assuming as intuitively obvious a principle regarded as controversial by many people. Consumerist is not seriously threatened at this point.

II – Ramifications Of Our Rebuttal of Segal

3.   Messy intuitions

While this completes my discussion of the argument presented by Segal, it is clearly not any sort of satisfactory conclusion. Segal relies in his argument on the Fregean Principle of Difference. Given that he does not explicitly argue for it, I have suggested that the most charitable interpretation is that Segal believes that principle to be an intuitively supported principle of normal practice, that we shall accept if we reflect on our practices of belief ascription. However, some externalists, and I have used Owens as an example, explicitly claim that FPD is revealed as resting on controversial philosophical assumptions, by the strong intuitions that support the claim of social externalism to be a principle of normal practice. The situation thus seems to be that we have two principles, FPD and Consumerism, that are claimed to be supported by intuition as principles of normal practice. And those two principles contradict in certain key puzzle cases. Such a situation is undoubtedly messy, and is downright unacceptable on some plausible assumptions.

Firstly, we want our intuitions about normal practice to be good guides to normal practice. That is not to claim that our *prima facie* reflections on our practices of belief ascription are always correct, but that extended philosophical analysis can deliver accurate information about the ways in which we attribute mental states. Secondly, we want folk psychology and normal practice to broadly be a good guide to the *truth* about beliefs and

mental states. This is clearly controversial, but I am content to say that anyone who rejects this assumption would probably not find the issues being discussed here of paramount importance in the first place. Thirdly, and finally, we take there to be a fact of the matter about the beliefs and other mental states which a subject is in. So if we assume that the actual state of affairs is not contradictory and that folk psychology is a good guide to that state of affairs, then we should be highly reluctant to suffer a contradiction in our intuitions about normal practice.

4.   A natural way out

Consumerism and the Fregean Principle of Difference both tell us what to say about the *concepts* that a subject expresses with their words when they make utterances. We can say therefore that the principles, and the intuitions that support them, contradict only insofar as they are concerned with the same *kind* of concept. This observation inspires an obvious solution to the problem of their apparent conflict: place the blame on the notorious flexibility of the term "concept", and say that both Consumerism and FPD are parts of normal practice, but that they concern different types of concept. This is the familiar "dual-content" approach. [3]

We have at hand a natural and intuitive dual-content approach, in the comments made by Tyler Burge himself. I employ his version as it is reasonably straightforward and allows us to highlight clearly the crucial points of debate, and also the contentious issues will turn on what is common to dual-content views, not on the peculiarities of the model.

Burge's distinction is between, in normal terms, a) the (roughly, public or linguistic) concept that a subject expresses with a term, and b) the (roughly, private and subjective) explication of the concept that on reflection the subject would give: the distinction is between the concept, and the conception of the concept. The concept is the *translational* meaning of the term, the *translational concept*, and the conception of the concept is the *explicational* meaning of the term, the *explicational concept*.

> I distinguish between a lexical item and the explication of its meaning that articulates what the individual would give, under some reflection, as his understanding of the word. Call the former "the word" and the latter "the entry for the word". I also distinguish between the concept associated with the word and the concept(s) associated with the entry. Call the former "the concept" and the latter "the conceptual explication". Finally, I distinguish between a type of meaning associated with the word, "translational meaning", and the meaning associated with its entry, "explicational meaning". For our purposes, the explicational meaning is the semantical analogue of the conceptual explication. (1989, 181)

It is important to note that whilst the explicational meaning of a term *can* be communal, in the sense of an explication that has been agreed upon and codified, perhaps as a dictionary definition, it can also be idiolectical, in the

---

[3]   The dual-content approach has been developed most famously by Block (1986), Fodor (1987) and Loar (1988) but it was present at the very beginning in Putnam (1975).

sense of the individual's personal explication of the term: the entry in that individual's personal dictionary, so to speak. Neither kind of explicational meaning is identical to the translational meaning, which can persist through changes in both idiolectical and communal explications of a concept. Associated with the two kinds of meaning are two kinds of understanding:

> [O]ne must distinguish the sort of understanding of a word in being able to use it to express a concept or translational meaning from the sort of understanding that is involved in being able to give a correct and knowledgeable explication of it. One may think with a concept even though one has incompletely mastered it, in the sense that one associates a mistaken conception (or conceptual explication) with it. (1993, 316f)

Burge's distinction clearly provides us with an account that accepts and explains the intuitions behind Consumerism and FPD, without having to conclude that normal practice is contradictory. The move from the principle that a rational subject will not sincerely and *knowingly* assent to contradictory beliefs, to FPD relies on the Cartesian idea that a subject is introspectively aware of whether her beliefs contradict. On Burge's distinction, what is clear is that a subject will certainly be aware if her beliefs contradict in their *explicational* concepts (personal dictionary definition), but may not be aware if they contradict in their *translational* concepts. We can now say that intuition supports FPD concerning explicational concepts, and not translational concepts.

> FPD-E: If a rational subject sincerely and on reflection assents to "a is F" and "b is not F" then they express different explicational concepts with "a" and "b".

Turning to Consumerism, we can use Burge's distinction here as well to formulate two principles where previously we have talked of one. It is plausible to say that intuitions support the principle concerning translational concepts, but certainly not explicational concepts. A subject does not, through their ability to be a normal, competent member of a linguistic community, get to be attributed the standard explicational understanding, dictionary definition, of a term[4]. Conceptual explications vary greatly from speaker to speaker, and where there is a codified and agreed communal explication of a term, it may not be known by many or even most of the competent speakers of the language, who are nevertheless perfectly able to use that term in the standard way to express the standard linguistic concept. However, Consumerism-T is plausibly supported by intuition. We do, as a fact of normal practice, ascribe standard linguistic concepts to competent speakers on the basis of their sincere utterance or assent.

> Consumerism-T: if a subject is minimally competent and deferential with an expression then they express the standard translational concept with that expression.

---

[4]   If we don't make the optional change and remove talk of experts, it is even clearer that DP-E is false. I do not get an expert's understanding of "arthritis" by becoming competent enough to use it, and be ascribed beliefs involving it.

So our natural solution is that if we correctly distinguish between the two aspects of meaning that speakers can express with terms, the contradiction that appeared to exist between our intuitions evaporates. Whilst the undifferentiated principles Consumerism and FPD do seem to contradict, it can be shown that intuition does not support them. Rather, it supports, Consumerism-T and FPD-E, which do not contradict. To reiterate: the consumerist principle that is a part of normal practice concerns the translational meaning of terms, the linguistic concepts they express; and the version of The Fregean Principle of Difference that is a part of normal practice is not a test for difference in translational concept expressed, but for difference in explicational concept, conceptual explication.

Now we can apply this account of the intuitions to Segal's scenario that we considered in section 1: the fact that Alf assents to "arthritis is not an inflammation of the joints" may reveal that Alf associates different explications with the translational concepts expressed by the expressions "arthritis" and "inflammation of the joint". But that does not mean that the translational linguistic concepts themselves differ – that is fixed, by his linguistic competence. Equally, Alf associates the same linguistic concept with the term "arthritis" as the experts, as they both belong, as functioning members, to the same linguistic community. But of course Alf does not share the same explicational understanding of the term as the experts. FPD-E and Consumerism-T adequately account for the situation described by Segal, without any threat of contradictory intuitions.

Clearly this is only a very rough gesture at a satisfactory account, and (as will be discussed later) its plausibility will depend on how the detail is filled in, but the bare outline is suggestive of a viable, and natural, solution.

5.   Problems for the social externalist

The account given above – which I shall call the "Modification Strategy" – seems to be a natural and intuitive method of dealing with our apparently contradictory intuitions, but such dual-content accounts have normally been proposed by those with internalist inclinations. The question concerns the issue of *cognitive content*. All the parties to the debate are concerned to describe the nature of that property of belief which is relevant to and operative in normal everyday folk-psychological explanation. An externalist is an externalist precisely because she believes that cognitive content is individuated externalistically. And an internalist is such because she believes that cognitive content is individuated internalistically. Given that fact, the issue that arises most prominently from the Modification Strategy is this: which aspect of belief – the socially individuated translational concept relevant to Consumerism, or the introspectively accessible explicational concept relevant to the Fregean Principle of Difference – is relevant to cognitive content?[5]

---

[5]   In posing this question I am rejecting the possibility that *both* explicational and translational concepts can contribute to cognitive content. Normal practice seems to assume a single notion of psychologically operative content, and we have seen that explicational and translational concepts pull in different directions in certain situations. It seems, and I will assume, that we can therefore countenance only one of the aspects

Social externalists place great emphasis on the role of their famous thought-experiments, and the intuitions that they inspire. The idea is that two intrinsically identical individuals, Bert and Twin-Bert, can be in different mental states, purely because of a difference in their linguistic communities. Bert, in our community, believes that you can get arthritis in your thigh. Twin-Bert, in a different community, where "arthritis" is used for a condition that can occur in joints and muscles, does not believe that you can get arthritis in your thigh – despite saying that "I have arthritis in my thigh", and being in all intrinsic respects identical to Bert. This thought-experiment supports the claim that the twins differ in the translational concepts they express with their terms. But it is written into the situation that both twins associate the same conception with the term "arthritis" – their personal dictionary definitions would be the same. So the thought-experiment so heavily relied upon by social externalists supports externalism for translational concepts, and not for explicational concepts. The twins share explicational concepts, and differ in translational concepts, so in so far as the thought-experiments support social externalism about the respects in which the twins differ, the thought-experiments support social externalism about translational content. So in order to use the those thought-experiments in such a way, the social externalist needs to show that translational concepts compose cognitive content. There are however difficulties with this requirement.

It is often held that the idiosyncratic, subjective component of beliefs is more suitable for the purposes of psychological explanation. There are a variety of ways to pursue this claim, and I shall adopt only one. I shall pursue the argument using the issue of *rationality*[6].

The first step in the argument is to claim that issues of rationality are central to folk psychological explanation. We predict people's behaviour under the assumption that they are rational, and make sense of their behaviour in the light of their rationality. If it turns out that someone is acting irrationally, we seem to be barred from being able to engage in folk psychology concerning them. Rationality is central to folk psychological explanation.

The next step in the argument is to claim that what matters to considerations of rationality is how a subject conceives of the world. It is rational for person A to do action X because of the way person A sees the world, their subjective view on the world – the way the world actually is seems to take a back seat. To use a concrete example: where would it be rational for Bert and Twin-Bert to apply an ointment labelled "arthritis cream"? Given that they both conceive of "arthritis" as being an ailment in their thighs, to be rational they both need to apply the cream to their thighs. This is despite the fact that in Bert's (our) world, arthritis just cannot occur in muscles, only joints. That last fact seems irrelevant to the question of what it is rational for Bert to do.

---

of belief being fundamentally contributive, in a substantive sense, to that property of belief which normal practice takes to be operative in folk psychological explanation.

[6] There are many complexities in the subject of rationality which are not directly relevant to this argument. The issue of rationality is just used here as an intuitive example of a case in which the subjective quality of belief is relevant to psychological explanation. The question of how rationality as a whole relates to externalism as a whole is a subject in itself.

To use another example. Alf says "arthritis is not the same thing as an inflammation of the joint". Alf is, by stipulation, rational. Arthritis is the same thing as an inflammation of the joint, so it can't be that, the way the world is, that renders his utterance rational. What renders his utterance rational is the fact that the conceptions he associates with "arthritis" and "inflammation of the joint" differ.

If we add the agreed premise that it is explicational concepts that capture a subject's conception of the meanings of her words, we can see that this argument, in showing that what is relevant to questions of rationality is the subjective aspect of beliefs, shows that it is explicational concepts that are central to questions of rationality. Further evidence for this claim is that rationality is explicitly referred to in our formulation of the Fregean Principle of Difference, and so the two are deeply linked.

So, explicational concepts (and not translational concepts) are relevant to rationality. And we said above that rationality is central to folk psychology. So we have to conclude that it is therefore explicational concepts that are central to folk psychology, and cognitive content. Translational concepts, on the other hand, seem to have therefore no central role in folk psychology.

So while the thought-experiments support social externalism about translational concepts, they do not seem to compose cognitive content. The conjunction of a) the natural way of dealing with the apparent contradictions in our intuitions (the Modification Strategy) and b) the argument just given, seems to render the traditional thought-experiments incapable of supporting social externalism. A social externalist who adopts the Modification Strategy "is effectively conceding that [Consumerism] is inadequate to account for [Alf's] behaviour….[T]he difference between the taxonomy of belief contents offered by [Consumerism] and the taxonomy for explaining actions is being recognised" (Patterson 1990, 321).

6. Social externalist response

What kinds of responses are open to the social externalist at this point? She could give up the dual-content approach and try and explain away the intuitions supporting the Fregean Principle of Difference. Such intuitions are very powerful, and the task seems daunting. Alternatively, the social externalist could attempt to retain the Modification Strategy, and defuse the argument just given, and claim that rationality is not as central to folk psychology as has been claimed. This could be done by showing that whilst our conception of the world is what matters when we explain our own behaviour, when we explain the behaviour of others we do not focus so much on their subjective viewpoint. We take into account many things which the subject may not be aware of, or are not aware of in the same way as us. In short, we may take into account the *context* in which the subject is embedded, and explain such behaviour without explicit concern for the rationality of the subject. We may, for instance, care about the suitability of the behaviour. This is a large issue, but it seems a well-supported claim, that what we care, in some sense, *most* about when we explain and predict people's utterances and physical behaviour is the state of their conception of the world and the meanings of the words they use, which I am tying to the issue of rationality.

The response I want to formulate is a different one, and concerns the possibility of accepting the argument concerning rationality, but spelling out the detail of the Modification Strategy in such a way as to avoid the apparent problems for social externalism.

The problem, recall, is that it is the kind of content that Alf and the experts do not share – the explicational concept – that seems to be relevant to folk psychology. Alf and the experts associate different explicational concepts with the term "arthritis": Alf believes that you can get arthritis in muscles, while experts do not, and they have many further technical beliefs about arthritis that Alf does not share. What appears to be the case here, however, is that it is quite *easy* to specify the differences in their explicational concepts, using perfectly standard words with perfectly standard meanings. The idiosyncrasy seems not to go beyond the surface.

> [This sort of case] demonstrates the *strength* of linguistic content, not its commonly supposed weakness: even in those bizarre cases the appeal to linguistic contents and other non-supervenient facts about history and belief acquisition is perfectly sufficient to explain what's going on. (Frances, 1999, 63)

This sort of view is suggested by Burge in his own comments on similar puzzle cases:

> You ask me what Al would think. It would be misleading for me to reply that Al would think that you do not have a contract, if I know that Al thinks a contract must be based on a formal document....In such cases, it is incumbent on us to cite the subject's eccentricity: "(He would think that you do not have a contract, but then) he thinks that there is no such thing as a verbally based contract." (1979, 91)

In describing Alf's explicational concept of arthritis, we consider it perfectly acceptable to ascribe certain standard concepts, without any hint that they themselves are idiosyncratically understood, or "private" in any sense – such as "he thinks one can have arthritis in a muscle". Public concepts attributed to competent language-users according to principles sensitive to linguistic conventions seem perfectly adequate to explicate the most bizarre and idiosyncratic understanding of a term. On this view, a subject's idiosyncratic explicational concept can be specified using words expressing perfectly standard public linguistic concepts, i.e. *translational* concepts. What this response does concede is that translational concepts as ascribed by Consumerism, in the form given above, are often not sufficient for psychological explanation. We may need to ascribe translational concepts in a different way in order to explain how the subject idiosyncratically groups them together in their conception of the meaning of a term:

> *Of course*, relying on simple disquotation will get it wrong in such cases. Being more complicated than the normal case, they require different, and often more complex, that-clauses if they are to be described accurately. (Biro 1992, 288)

We can cast this response in terms of the privacy and publicity of concepts: the externalist can concede that explicational concepts are private in the sense that they may indeed not be shared by anyone else, but they are public

in that they can be, and often are, clearly communicated. So whilst explicational concepts are paramount in folk psychological explanation, their use is underpinned by that of translational concepts, which are, as we have seen, individuated according to social externalism.

We can draw out the core issue by reconsidering the original thought-experiment, with the twins Bert and Twin-Bert. Their shared explicational concepts seem to be doing the work in any folk psychological explanation we may want to give of them. And we can specify that shared explicational concept. So whilst we can't say that they both believe that one can have arthritis in their thigh – as Twin-Bert has no beliefs about arthritis – we can say that they both believe that one can have "arthritis" in the thigh, that old people get it, that their Auntie Doris has it, and all the other things that warrant the claim that they share the same explicational concept. That is, we are inclined to say that they have some important property of belief in common because we are told that they both think that one can have "arthritis" in the thigh etc.

But what is now apparent is that there is an assumption that they don't really inhabit different linguistic communities. For the purposes of all terms except "arthritis", they inhabit the very same linguistic community, so we can specify, without difficulty, what it is that they share. What the social externalist may be able to claim is that if the difference in their linguistic communities were expanded to include certain other terms along with "arthritis", then the impulse to say that they shared any important belief content in common would recede. Specifically, if their linguistic communities used the terms "thigh", "old", "Auntie" in a systematically different way, consonant with their intrinsic identity, then we may lose the ability to claim that they share any explicational concept. We would have to say that they both believe that one can get "arthritis" in the "thigh", that "old" people get it, that their "Auntie" Doris has it. And this seems to say next to nothing about what they believe, and be of next to no use in any folk psychological explanation. When we lack translational concepts with which to specify a subject's explicational concept, the pull to claim that such an explicational concept actually exists is not there. The drive to say that Bert and Twin-Bert share a psychological important property of belief is a product not of their intrinsic similarities, but their shared linguistic environment and their responsibility to shared linguistic norms.

This response claims that the thought-experiments seem only to support social externalism concerning translational concepts because they are not really cases of subject being in different linguistic communities. If we expand the thought-experiment in the way described, it becomes apparent that explicational concepts are socially externalistic, in that they are composed of the translational concepts that have been socially externalistic all along. That is, explicational concepts themselves are sensitive to changes in linguistic environment.

Conclusion

If the contradictory intuitions raised by the social externalist rebuttal of Segal do force us to adopt a tactic such as the Modification Strategy, what is clear is that there is a lot to be done in order to make a social externalist approach to

folk psychological explanation plausible. What seems to be the case is that a simplistic version of social externalism such as Consumerism (or the Disquotational Principle) is inadequate to capture the complex relationships between a subject and their environment. Those externalist efforts focussed on such uncomplicated principles may be misguided, and as we have just seen, success for the externalist may lie in distancing themselves from such principles and developing detailed externalist accounts of a subject's idiosyncratic conception of the world.

References

Biro, J. (1992), "In Defence of Social Content", *Philosophical Studies 67*: 277-293.

Block, N. (1986), "Advertisement for a Semantics for Psychology", *Midwest Studies in Philosophy 10*, French et al eds.

Burge, T. (1978), "Belief and Synonymy", *Journal of Philosophy 75(3)*: 119-138.

Burge, T. (1979), "Individualism and the Mental", in P. French, T. Uehling, and H.

Wettstein, eds., *Studies in Epistemology*, Midwest Studies in Philosophy 4 (Minneapolis, University of Minnesota Press): 73-121.

Burge, T. (1986), "Intentional Norms and the Foundation of Mind", *Journal of Philosophy 83(12)*: 697-720.

Burge, T. (1989), "Wherein is Language Social", in A. George, ed., *Reflections on Chomsky* (Oxford, Blackwell): 175-191.

Burge, T. (1993), "Concepts, Definitions, and Meaning", *Metaphilosophy 24(4)*: 309-325.

Fodor, J. (1987), *Psychosemantics* (London, MIT).

Frances, B. (1999), "On the Explanatory Deficiencies of Linguistic Content", *Philosophical Studies 93*: 45-75.

Kaplan, D. (1989), "Afterthoughts", in J. Almog, J. Perry, and H. Wettstein, eds., *Themes From Kaplan* (Oxford, OUP): 565-614.

Kripke, S. (1979), "A Puzzle About Belief", in A. Margalit, ed., *Meaning and Use* (Dordrecht, Reidel). Reprinted in N. Salmon, and S. Soames, eds., (1988) *Propositions and Attitudes* (Oxford, OUP): 102-148.

Loar, B. (1988), "Social Content and Psychological Content", in R. Grimm, and D.

Merrill, eds., *Contents of Thought* (Tucson, University of Arizona Press). Reprinted in D. Rosenthal, ed., (1991) *The Nature of Mind* (Oxford, OUP): 568-575.

Owens, J. (1989), "Contradictory Belief and Cognitive Access", in P. French, T. Uehling, and H. Wettstein, eds., *Contemporary Perspective in the*

Philosophy of Language 2, Midwest Studies in Philosophy 14 (Minneapolis, University of Minnesota Press): 289-316.

Owens, J. (1992), "Cognitive Access and Semantic Puzzles", in Anderson, and Owens, eds., *Propositional Attitudes* (CSLI Press): 147-173.

Patterson, S. (1990), "The Explanatory Role of Belief Ascriptions", *Philosophical Studies 59*: 313-332.

Segal, G. (2000), *A Slim Book About Narrow Content* (London, MIT Press).

Segal, G. (2003), "Ignorance of Meaning", in A. Barber, ed., *Epistemology of Language* (Oxford, OUP).

Transcendental Arguments in the Philosophy of Mind

Mathilde Byskov Jakobsen

Department of Philosophy, University College London

m.jakobsen@ucl.ac.uk

Abstract

In this paper I defend the use of transcendental arguments within the philosophy of mind. I argue that we can use transcendental arguments to establish certain necessary features and capacities of the mind, given that it is the mind of a certain kind of experience. This means that as well as formulating belief-directed transcendental arguments, we can also formulate truth-directed transcendental arguments as long as they are directed at certain features of the mind. I present an argument for the unity of consciousness to illustrate this form of transcendental argument. I argue that this is a truth-directed transcendental argument which is not subject to Stroud's objection to the use of such arguments outside of an idealist framework.

1. Introduction

My aim in this paper is to defend the use of transcendental arguments in the philosophy of mind. I will identify a certain kind of belief-directed transcendental arguments that can be of some limited use in the philosophy of mind. This use of transcendental arguments fits with the view that all transcendental arguments can ever hope to reach are conclusions about beliefs. However, the main claim which I will defend in this paper is that a certain kind of truth-directed transcendental argument can also be of use in the philosophy of mind. I will defend this claim primarily by considering the methodology of transcendental arguments directed at the mind, but also by developing an example of a self-directed transcendental argument. I call this type of argument a "self-directed transcendental argument" following Quassim Cassam, as it is a name which reflects its purported subject matter. My paper has three main parts. First, I will consider the methodology of transcendental arguments in general and present Barry Stroud's well-known objection to them. Second, I will present two kinds of self-directed transcendental arguments of use in the philosophy of mind, which are not subject to this objection. The first of these is a kind of belief-directed transcendental argument. The second kind is a kind of truth-directed transcendental argument. The possibility of formulating truth-directed transcendental argument of this kind is exclusive to the philosophy of mind, and it is dependent on an assumption about the nature of mind which is broadly functionalist. In the third part of the paper I will further evaluate this strategy by considering an example of a self-directed transcendental argument. This argument will be developed from some of Kant's claims about transcendental apperception in the 'Transcendental Deduction'. I conclude that this methodology can identify significant constraints on the form of the mental

operations that our minds must implement, given that our experience has the character that it does, and so tell us something about the mind – or the self.

I

2. What are transcendental arguments?

Transcendental arguments argue from premises about experience to a conclusion about reality. Typically, they start with a premise which states that we have experience of a certain kind. The second premise determines that this form of experience is conditional upon some unobvious and philosophically interesting claim about reality. Which we can then conclude must hold. In one traditional use of such arguments, the aim of the argument is to refute the philosophical sceptic. In this traditional use the first premise is selected in order to be immune from Cartesian scepticism about the external world. This premise describes experience in a way which cannot be doubted from the first person perspective. However, this aspect is optional: transcendental arguments are not, primarily, intended to refute the philosophical sceptic even if they have, traditionally, been used in this way. Rather they are concerned essentially with identifying necessary conditions for experience, hence the force of the description "transcendental". This means that instead of having the following form:

> Necessarily P
>
> Necessarily, (if P then Q)
>
> Therefore, necessarily Q

The transcendental arguments I will discuss have this logical form:

> P
>
> Necessarily, (if P then Q)
>
> Therefore, Q

If the premise we start with is only contingent, - if it is taken as a matter of fact that we have a certain kind of experience, then the conclusion reached is not necessary. And if we cannot establish the first premise with certainty, then the conclusion reached is only conditional. This means that the argument cannot be used to refute philosophical scepticism. On the other hand it is not dependent on the claim that the experience we have is in some sense necessary.[1]

An example of a transcendental argument can be found in Kant's refutation of idealism (B275):

> I have experience of myself as determined in time;
>
> Time determination is dependent on the experience of something persistent, and this cannot be merely a representation in me;

---

[1] Harrison (1982) p. 216.

Therefore, there must be something persistent outside me of which I have direct consciousness by which I can judge the passing of time.

3.   Stroud's Objection

In a well-known article from 1968 Stroud objected to this entire class of arguments. His main objection is that transcendental arguments cannot draw conclusions about reality, because it is always open to an objector to insist that it is enough that the world appears to us a certain way or that we believe it to be a certain way in order for us to have experience – it is therefore not necessary that the world is in fact in this way. We can apply this objection to Kant's argument above. It draws a conclusion about reality – that there is in fact something persisting outside of me of which I can be conscious, but, the objection goes, it is in fact enough if it just appears to me as if there is something persistent outside of me and that I believe that it is independent of me, in order for me to have experience of myself as determined in time.[2] If this is true, then the transcendental argument has reach a conclusion about some belief or experience that a thinker must hold, given that the thinker holds other beliefs or have other kinds of experiences. It then does not in fact argue from a premise about experience to a conclusion about reality. TA's are therefore only belief-directed, rather than truth-directed and do not establish objective conclusions about how reality must be.

This objection, however, does not hold against the use of transcendental arguments within the framework of transcendental idealism, as Stroud has also made clear.[3] Such a framework secures the bridge between experience and reality, because reality, as phenomena, is dependent on our experience of it. Therefore, in drawing conclusions about necessary features of our experience using transcendental arguments one is drawing conclusions about phenomenal reality. As Kant puts it "The conditions of the possibility of experience in general are likewise conditions of the possibility of the objects of experience, and (…)[for this reason they have objective validity in a synthetic a priori judgement" (B197). This means that within the framework of transcendental idealism we can formulate *truth*-directed transcendental arguments.

II

4.   Psychological   belief-directed,   self-directed   transcendental arguments

We can concede to Stroud that transcendental arguments must be interpreted as *belief-directed* if they are to be valid outside the framework of transcendental idealism, and still preserve the usefulness of such arguments. We can do that by re-deploying these arguments, to be not about the world,

---

[2]   Stroud (1968).

[3]   Stroud (1999) p. 160.

but about how we must take the world to be – not about facts but about beliefs.

So far I have discussed transcendental arguments as directed at either the world or a belief about the world. However, we can also direct transcendental arguments at the mind or the self or at a belief about the mind or the self. I will call these arguments self-directed arguments, following Cassam.[4] Self-directed transcendental arguments argue from premises about experience to a conclusion about a belief or a fact about a feature or a capacity of the mind or the self which is necessary for the experience in question.

We can produce the following grid of the two kinds of transcendental arguments so far identified:

World or Self/Mind	1: Truths
	2a: Beliefs
	2b: Experiences

The lesson learned from Stroud is that in a non-idealist framework we can only formulate transcendental arguments of the second, belief-directed, kind.

Cassam's argument in *Self and World* is an example of a belief-directed, self-directed transcendental argument. The claim of the transcendental argument is that it is necessary that a subject of self-conscious experience has a psychological belief with a certain content: oneself qua subject as a physical object. Cassam develops both a belief and an experience version of this argument. He develop the experience version of the argument because the belief-version is subject to the objection that people who believe that they are Cartesian souls are still able to ascribe their representations to themselves is solved by saying that they in fact still have an experience or an intuition of themselves as physical objects, however confused they are in their beliefs.

Belief-directed transcendental arguments clearly have a use in the philosophy of mind. Arguments of this type can show the necessity of a certain psychological belief that a subject must have in order to have experience. I call these kinds of transcendental arguments 'psychological belief-directed arguments' for this reason. Such arguments can show that a subject, who has experience of a certain kind, must have some actual belief with a specific content. The belief is psychologically necessary for the possibility of experience. However, all such arguments are subject to what Cassam calls the 'misconception' objection (*Self and World)*. The misconception objection turns on the fact that if a transcendental argument can show that some belief is necessary for experience, then it should not be possible to have a misconception about the subject matter of this belief. If a belief-directed transcendental argument for instance shows that you cannot have experience of yourself in time without believing that there are objects

---

[4]   Cassam (1999).

outside of you, then the existence of sincere idealists, who experience themselves as determined in time, is impossible.

Psychological belief-directed, self-directed transcendental arguments can only be used to show that some belief, which we in fact all hold, is necessary. This obviously does not make such argument useless: if successful, the transcendental argument will tell us that it is not a contingent matter that we all hold the same belief. However, it does limit the subject area of such arguments radically.

However, the use of transcendental arguments in the philosophy of mind is not limited to establishing the status of certain beliefs that we all hold. Rather, philosophy of mind is an area where we can formulate *truth*-directed transcendental arguments, or so I will argue. In order to argue for this claim, I will first identify an alternative way in which transcendental arguments can be belief-directed and met Stroud's objection. Second, I will argue that in the area of the philosophy of mind some such belief-directed, self-directed transcendental arguments are in fact also truth-directed transcendental argument. If this is right, then transcendental arguments have the potential to draw conclusions about facts about the mind or the self, not just about beliefs about the mind or the self.

5. Theoretical belief-directed, self-directed transcendental arguments

We introduced psychological belief-directed transcendental argument as a reaction to Stroud's objection to truth-directed transcendental arguments. Psychological belief-directed transcendental arguments met Stroud's objection by making their aim less ambitious. Psychological belief-directed arguments aim to draw the conclusion that a subject of experience must have a certain belief, not that a certain fact must be true of the world or the subject.

However, that is not the only way in which we can make our arguments less ambitious and so avoid the problem Stroud points out. We can re-deploy transcendental arguments to reach conclusions about our conceptual scheme, instead of about the world. This solution can be used on Strawson's arguments in *Individuals*. On this view transcendental arguments can aim to show us that a certain belief is part of our conceptual scheme. This designates a different way in which transcendental arguments can be belief-directed than the way discussed above. Transcendental arguments can aim to show conceptual connections between the way in which we, as theorists, think of experience, and the way in which we, given that, have to use related concepts. This importantly does not place any psychological constraints on the beliefs of individual people. I call this second sense of a belief-directed transcendental argument, a *theoretical belief-directed argument*, because it shows the theoretical necessity of a belief, rather than the psychological necessity of the belief. A transcendental argument could for instance attempt to establish that given our conceptual scheme it is necessary that we take objects to be in space and time. The conclusion, that objects are in space and time, is a belief rather than a fact or truths about the world. However, the argument does not establish that single individuals must hold this belief to have experience. This distinguishes theoretical belief-directed transcendental arguments from psychological belief-directed transcend-dental arguments.

Cassam claims that all belief-directed transcendental arguments are subject to the misconception objection [*Self and World*]. However, in fact the misconception objection is only effective against psychological belief-directed arguments. Sincere idealists are not a problem for the transcendental philosopher who formulates theoretical belief-directed arguments. She is committed only to saying that a certain belief is (theoretically) required given our conceptual scheme not that it is psychologically required, and so allows that people can be confused and still have experience. A theoretical belief-directed version of Cassam's argument for the necessity of having the belief that one is a physical object to have self-conscious experience would not be subject to the misconception objection. It would just show that we have good philosophical reasons given our concepts of experience and subjects to think that subjects have to be physical objects, without demanding that the subject must actually believe itself to be one. This argument meets Stroud's objection. It does that because it allows that, though it is part of our conceptual scheme that we take ourselves to be physical objects, we cannot prove that we really are physical objects. It allows for the possibility that the world may in fact be very different from the way we have to take it to be.

We can add this last type of transcendental argument to the grid from before, so that it now looks like this:

	1: Truths	
World or Self/Mind	2a: Beliefs	} psychological
	2b: Experiences	
	3: Beliefs	} theoretical

Theoretical belief-directed transcendental arguments are not very interesting, when directed at the world in a non-idealist framework. Outside of transcendental idealism conditions of cognition are not also conditions of objects. The way in which we have to think of things does not tell us anything about how they are. In the case of world-directed transcendental arguments, we can for instance attempt to show that the belief in causation plays a central role in our conceptual scheme, so that in a reflection on the possibility of experience, we have to think of the world as causally ordered. However, the question will always arise whether independent reality actually corresponds to this necessary way of conceiving of it. The fact that our conceptual scheme demands that we make sense of the world only as causally ordered does not tell us what we want to know, namely whether the world is in fact causally ordered or not.

6.   The nature of mind

However, my suggestion is that in the case of at least some of the transcendental arguments we can direct at the mind, we do not find the same gap between how we must think of things and how they really are. Say that we have established that we must necessarily interpret or conceive of the mind as having certain features and characteristics in order to understand it as a mind which has experience of a certain kind. Now, we cannot allow that it

may in fact not really have these features at all – whilst still being interpretable as a mind of this kind of experience.

In cases where our argument shows that we must necessarily think of the mind as realising certain *functions or features related to functions* there is no sense in driving a wedge in between the claim that the mind must be interpreted or conceived as having these features and capacities, and the claim that the mind really realises these features and capacities. The picture of the mind required to support this view is broadly functionalist in that a mind is defined by its functions.[5] If something is defined by its function, and it performs these functions, then there is no question as to whether its real nature fits our functional picture of it.

This assumption about the mind opens up the possibility that *if* we can formulate transcendental arguments about features of the mind necessary for it to perform its functions, then we can consider them as identifying aspects of the mind and not just of our conception of the mind.

This assumption about the mind must not be confused with the claim that the mind is transparent to itself either in the sense that we know infallibly the features of our minds or the content of our mental states by direct introspection. It is not a claim about people's psychological relation to their empirical minds.

It also does not suggest that we can find out everything there is to know about the mind by the use of a transcendental reflection on the possibility of experience. First, there are many empirical facts about the mind which require empirical methods. Second, transcendental philosophy cannot tell us how a certain necessary feature or capacity is realised in the mind – if there is more than one way in which this could be done. A transcendental philosophy of mind uses self-directed transcendental arguments to ascribe features to the mind (necessary conditionally dependent on that the mind in question is a mind of a certain kind of experience). However, while it tells us that the mind must realise these features, it does not tell us how it realises them. Just like other kinds of functionalism, transcendental philosophy of mind is ontologically uncommitted. It can tell us nothing about either the substrata of mind or about the actual realisation of the very general constraints we can place on the mind.

This suggests that the philosophy of mind allows for transcendental arguments to be used to establish facts about the mind and not just beliefs about the mind.

III

7.   An example
To take a concrete example of this strategy, consider Kant's account of transcendental apperception in the B-edition of the 'Transcendental

---

[5]   It is only 'broadly' functionalist as it is not committed to any of the more specific requirements of contemporary functionalism.

Deduction'. A self-directed transcendental argument can be developed out of some of these considerations. The starting point of the B-Deduction is the observation that experience is made up of complex representations, where both the individual elements of a complex representation and different complex representations are synthesised together to form one experience. The next step is to say what this requires in the subject of such experience. The answer for Kant is the unity of transcendental apperception:

> It must be possible for the 'I think' to accompany all my representations; because otherwise something would be represented in me which could not be thought at all and that is equivalent to saying that the representation would be impossible, or at least would be nothing to me. (B132).

This quote is saying that all conscious parts of one experience (all representation that are not noting to me) must be self-ascribable to one 'I'. The idea that *a bundle of representations are all in some sense mine* corresponds to one understanding of the unity of consciousness in contemporary philosophy of mind. We have to distinguish this understanding of the unity of consciousness from one that refers to a feature of our experiences themselves. The term 'the unity of consciousness' can be used to refer to a phenomenal feature of experience: the fact that our experiences are unified in the sense that it feels different to experience *a* together with *b*, than it does to experience *a* alone (and that it is the experience of *a,* which is different).[6] Unity of consciousness, as I am interested in it, is not a claim about the feel of experience. If it was the argument would be trivial as the claim that experience is unified (at least partly) is already involved in the claim that it contains complex representations. Saying that a bundle of conscious states must all be ascribable to the same 'I' , saying that a bundle of representations all in some sense belong to me, is not a claim about experience but about the mind. So the possibility of complex representations requires that the parts of this representation are in some sense all mine – which is equivalent to the claim that they belong to one unity of consciousness. I will suggest three ways in which we can understand the unity of consciousness understood in this way.

One way of understanding what it means to be self-ascribable to one 'I' is via a causal relation to a body – a bundle of experiences are part of a unity of consciousness if they are all causally dependent on the same body. We can reject this view because there are no good reasons to think that there is a one-one relation between a body and a unity of consciousness – and split brain and multiple personality disorder cases in fact give us reasons to think that this is not necessarily the case. The conceivability of a unity of consciousness supported by more than one body – perhaps by a succession of bodies as in Kant's billiard ball example also supports this point. The fact

---

[6]   Dainton and Chalmers and Bayne take the unity of consciousness to express a phenomenal quality about our experiences themselves at a time. For Dainton the unity of consciousness is a phenomenal characteristic that lies in the co-conscious experiences themselves, for Chalmers and Bayne it is the existence of a state, whose phenomenal character subsumes the phenomenal characters of all the states the subject is in at that time.

that a bundle of representations all belong to a body in being causally dependent on it does not mean that they can all be self-ascribed together. And the possibility that a bundle of representations are causally dependent on two bodies does not rule out that they could be self-ascribed to one subject together. And it is this latter sense of belonging we need to make complex representations possible. This shows that just because something is true of the subject it is not necessarily true for the subject - it is not enough that it is true that the parts of a representation are mine if I don't take them to be mine.

Another option is to understand the unity of consciousness as a claim about actual self-ascription. Under this suggestion a number of conscious states are part of a unity of consciousness if there exist in the subject a state of the form 'I think that…' which subsumes all of them. However, the problem with basing unity of consciousness on actual self-ascription is that it limits it to include only states where such a higher order state exists, which seems too restrictive. It is after all supposed to be necessary for the possibility of complex representations.

We therefore have the further option of basing the unity of consciousness on *possible* self-ascription, which avoids this problem. According to this suggestion a bundle of experiences are parts of a unity of consciousness if they *can* be self-ascribed together under one 'I think'. Some people object to the use of dispositions or possibilities in an account of the unity of consciousness because it is hard to make the claim substantial. In situations where a subject at a time has a whole range of experiences, only some of which are ever actually taken together, how do we know that it could actually have done so to all of them? However, the solution cannot be to avoid talk of possibilities in favour of actualities. Accounting for the unity of consciousness in terms of a disposition or a capacity for self-ascription is still the best option, in my view, because it captures the sense of experiences belong to me, that we need, without being unnecessarily restrictive.

If experience contains complex representations, then the subject of experience realises a unity of consciousness – that is, the parts that make up the complex representations can all be self-ascribe under one 'I think'. Kant illustrates the necessity of the unity of consciousness with the example of the consciousness of a verse of a song – in order for there to be *a consciousness of a verse*, it is necessary that all the parts of the verse is taken by the subject as all its experiences and as belonging together. If a conscious complex representation is to count as such, then the subject must be able to take its part as a whole and self-ascribe it as one. There is no need for the subject to actually so self-ascribe it, which shows that it is the possibility not the actuality of self-ascription which is important. So if experience contains complex representations, then the unity of consciousness is necessary for experience.

8. The status of the conclusion

In this argument we go from an observation of the kind of experience we have to a conclusion about a necessary feature of the mind that has such experience, which in this case is the unity of consciousness. This conclusion of the argument is not that the subject must believe in the unity of consciousness in order to have experience. If the unity of consciousness is necessary for the possibility of experience containing conscious complex

representations, then it is neither necessary nor sufficient that the subject *believes* that it realise a unity of consciousness, it must actually do so. The argument does make a claim our conceptual scheme. The conclusions we draw about the mind on the basis of self-directed transcendental arguments link ways in which we think of experience with ways in which we, given this, must think of the mind. The argument identifies a conceptual connection between thinking of experience as unified and thinking of the subject of these experiences as realising a unity of consciousness. Describing this connection as conceptual does not reduce it to the trivially linguistic. It involves a reflection on experience, and a reflection on how we must think of the experiencer given that it is the subject of this kind of experience. However, the conclusion is not restricted to a claim about how we must think of the mind. In the case of functionally defined aspects of the mind, if we can show that we must necessarily think of the these features in a certain way, given that it is the mind of a certain kind of experience, then this identifies a real aspect of the mind.

It seems to me plausible to defend this general claim in the particular case of the argument about the necessity of the unity of consciousness.

This particular use of a self-directed transcendental argument, then, is not subject to Stroud's general methodological strictures: It is not restricted to showing only the necessity of a psychological belief that we all hold and it is not restricted to showing only something about our general conceptual scheme – if we take this to mean that it has not also been established as a feature of reality. The argument is therefore truth-directed.

I conclude that the philosophy of mind allows for the use of transcendental arguments to determine the features of the mind that are conditionally necessary for our experience to take the form that it does.

Bibliography

Brook, Andrew: Kant and the Mind. Cambridge University Press 1994.

Cassam, Q: *Self and World*. Oxford University Press 1997.

Cassam, Q: 'Self-directed Transcendental Arguments' in Stern, R: *Transcendental Arguments*. Oxford University Press 1999.

Chalmers, D. and Bayne, T: 'What is the Unity of Consciousness'. Here from Chalmers' webpage, but published in A. Cleeremans (ed.), *The Unity of Consciousness: Binding, Integration, Dissociation*. Oxford University Press 2003.

Dainton, B: *Stream of Consciousness*. Routledge 2000.

Gardner, S: *Kant and the Critique of Pure Reason*. Routledge 1999.

Harrison, R: 'Transcendental Arguments and Idealism' in Vesey, G: *Idealism, Past and Present*. Cambridge University Press 1982.

Kant, I: *Critique of Pure Reason*. (Translated by Kemp Smith) The Macmillan Press 1929 (second edition 1933).

Rosenthal, D: 'Unity of Consciousness and the Self' Proceedings of the Aristotelian Society 2003.

Strawson, P. F: *Individuals*. Routledge 1959.

Stroud, B: 'Transcendental Arguments' The Journal of Philosophy vol. LXV, no. 9. 1968.

Stroud, B: 'The Goal of Transcendental Arguments' in Stern, R: *Transcendental Arguments*. Oxford University Press 1999.

Stroud, B: 'The Synthetic A Priori in Strawson's Kantianism' in Glock, H-J: *Strawson and Kant*. Oxford University Press 2003.

Conceptualism and The (Supposed) Non-Transitivity of Colour Indiscriminability
Charles Pelling

Department of Philosophy, University of Reading

Abstract

In the philosophy of perception, the conceptualist asserts, and the non-conceptualist denies, that the representational content of experience is exclusively conceptual content. In this paper, I show that Delia Graff's recent arguments for the traditionally unpopular view that colour indiscriminability is transitive have important implications for the conceptualism versus non-conceptualism debate.

This is because (1) conceptualism can be true only if we possess context-dependent demonstrative colour concepts, and (2) only if colour indiscriminability is transitive can we possess such concepts.

This paper addresses (2). In order to establish its truth, I consider two accounts of demonstrative colour concept possession, those given by the two most prominent conceptualists, John McDowell and Bill Brewer. McDowell and Brewer each propose a condition that a subject must satisfy in order to possess a demonstrative colour concept. In the bulk of the paper, I am concerned to establish two things: first, that unless colour indiscriminability is transitive, neither of the conceptualists' proposed conditions are satisfiable; and second, that at least one of these conditions must indeed be satisfied by any genuine demonstrative colour concept possessor.

1.   Introduction

Delia Graff has recently argued – effectively, I think - for the traditionally unpopular view that colour indiscriminability is transitive.[1] My aim in this paper is not, however, to defend or even to discuss Graff's arguments, but rather to investigate their implications for an issue in the philosophy of perception that has generated considerable debate in recent years.

The debate is between the *conceptualists* – such as John McDowell and Bill Brewer - and the *non-conceptualists* – such as Christopher Peacocke and the late Gareth Evans. After explaining briefly what these opposing positions amount to, I will show that there are good reasons for the conceptualists to accept Graff's view that colour indiscriminability is transitive.

2.   Conceptualism and Non-Conceptualism

Conceptualism is the view that the representational content of (normal human adult) perceptual experience is exclusively conceptual content. *Non-*

---

[1]   In her 2001. Graff's arguments extend also to other forms of indiscriminability. I will focus on the issue of colour indiscriminability in order to simplify the discussion.

conceptualism is the view that the representational content of perceptual experience is at least partly non-conceptual content.

As I use the terms 'conceptual content' and 'non-conceptual content', insofar as the content of a subject's experience is conceptual, the experience has the content it does in virtue of its drawing into operation concepts that the subject himself possesses at the time of the experience; insofar as the content of a subject's experience is non-conceptual, by contrast, the experience has the content it does in virtue of something else – *not* in virtue of its drawing into operation concepts that the subject possesses at the time of the experience.

3.   Conceptualism and Fineness of Grain

An immediate consequence of the conceptualist view is that we must have concepts for all the various objects, properties, and relations which feature in our experiences. A traditional worry here is that it seems unlikely, initially at any rate, that we have a concept for every colour shade (say) that features in our experience when we are looking at a spectrum. It seems tempting to say that in such a situation, our colour concepts are not as *fine-grained* as the shades that feature in our experience.[2]

However, the conceptualists insist that this worry rests on the mistaken assumption that we can possess only those context-independent concepts expressible by such general words as 'red', 'green', or 'terracotta'. According to the conceptualists, such an assumption overlooks the possibility that while one is enjoying an experience in which a *particular* colour shade features, one can express a concept of exactly this shade by using a phrase which includes a demonstrative.[3] This response, the conceptualists contend, is sufficient to undermine the worry that stems from the issue of fineness of grain since it seems that one could have such a demonstrative concept for any shade, or any other property or relation, which could conceivably feature in an experience.

4.   Demonstrative Concepts and Recognitional Capacities

But the conceptualist, as McDowell himself notes, must ensure that our supposed demonstrative concepts really are bona fide concepts.[4] McDowell draws our attention to Wittgenstein's case in which a subject proclaims 'I know how tall I am!' and lays his hand on top of his head to prove it.[5] McDowell agrees that such a subject would lack a genuine demonstrative concept of his own height. But given this, what exactly *is* required for a subject to possess a demonstrative concept of a particular height (or a particular shade, etc.)? McDowell's suggestion is that:

---

2   cf. Evans, 1982, p. 229, and Heck, 2000, pp. 489-90.

3   See McDowell, 1994, pp. 56-7, and Brewer, 1999, pp. 170-4.

4   McDowell, 1994, p. 57.

5   Wittgenstein, 1958, § 279.

'We can ensure that what we have in view is genuinely recognizable as a conceptual capacity if we insist that the very same capacity…can in principle persist beyond the duration of the experience itself…What is in play here is a recognitional capacity, possibly quite short-lived, that sets in with the experience.'[6]

To establish how this proposal is supposed to work, let us return to the case where a particular colour shade is featuring in a subject's experience. I think that McDowell's idea, then, amounts to this: the subject possesses a concept of the shade only if he is capable, after the shade no longer features in his experience, of telling whether or not a shade that features in a *new* experience falls under the very same concept.

So for instance, suppose that my experience represents an object as being a particular shade of blue. According to McDowell's proposal, if I possess a concept for this exact shade, I must be capable, after the shade no longer features in my experience, of telling whether or not something is being represented to me in a new experience as being this exact shade of blue.

However, not all philosophers have been convinced by McDowell's account of our possession of such demonstrative concepts. One common suggestion is that, as a matter of empirical fact, shades[7] just *do* feature in our experiences even when we lack the corresponding recognitional capacities that McDowell envisages: contra McDowell, a shade can feature in my experience even when I would not be able to tell, after the shade no longer features in my experience, whether or not a shade featuring in a new experience is identical to the original shade. Sean Kelly voices something like this worry when he claims that:

'there's nothing in the nature of perception to keep it from being true, that our capacity to discriminate colours exceeds our capacity to re-identify the colours discriminated.'[8]

And if Kelly is right, then either McDowell is wrong to suppose that demonstrative colour concept possession requires such recognitional capacities, or (as Kelly thinks) he is wrong to suppose that, as conceptualism implies, we must have demonstrative concepts for all the shades which feature in our experiences.

In response to this kind of concern (and unwilling to abandon his conceptualism), Bill Brewer has proposed a rather different account of demonstrative colour concept possession.[9] According to Brewer's proposal, a possessor of a demonstrative concept of a shade featuring in an experience need not have the ability to tell, after the shade no longer features in the experience, whether or not a shade featuring in a new experience is identical to the original shade. Rather, the concept possessor must have the ability

---

6   McDowell, 1994, p. 57.

7   Or heights, etc. Although I will focus on shades from now on, the points that I will make apply to other properties and relations.

8   Kelly, 2001, p. 411.

9   Brewer, 2004, pp. 14-18.

'to keep track of the same shade…during a single extended period of observation.'[10]

The key idea here then is that the demonstrative colour concept possessor need only keep track of the relevant shade during the course of a *single* experience: there is no requirement that he should be able to reidentify the shade after a complete *break* in (the shade's featuring in the) experience. And *this* account, one might think, is one which is rather less likely to be shown empirically to be incompatible with conceptualism: from the philosopher's armchair at least, it seems plausible that although (as a matter of empirical fact) our capacity to discriminate colours exceeds our capacity to reidentify the colours discriminated after a *break* in experience, our capacity to discriminate colours does *not* (as a matter of empirical fact) exceed our capacity to *keep track* of the colours discriminated over the course of a *single* experience.

5. The Incompatibility of Two Claims

But if the theorist decides to accept either McDowell's stricter constraint or Brewer's more liberal alternative condition on demonstrative colour concept possession, he will need to give an account of the *identity conditions* for the shades that feature in our experiences. For if a subject is to *recognise* that some shade which features in a new experience is identical to a shade which featured in a previous experience, or if he is to *keep track* of a particular shade over the course of a *single* experience, then it must be that there is only one shade in play: there must be only one shade that he reidentifies, or one shade of which he keeps track.

However, I do not think that the theorist should look to the conceptualists' *own* accounts for an answer to this question, since this is a stage at which they themselves go wrong. For in providing such accounts, the conceptualists make two claims that are incompatible.

First, take McDowell's suggestion that:

'We might lay down the rule that something counts as having *that* shade just in case it is indiscriminable in colour from the indicated sample.'[11]

Call this *the identity of indiscriminables claim*. So if the claim is true, in order to possess a demonstrative concept of a shade that is featuring in my experience:

(1)   According to McDowell, I must have the capacity to tell, after the shade no longer features in my experience, whether or not a shade that features in a new experience is *indiscriminable* from the original shade; or,

(2)   According to Brewer, I must have the capacity to tell, at a later stage of the *same* experience, whether or not a shade that is featuring in the experience is *indiscriminable* from the original shade.

---

[10]   Ibid., p. 16.

[11]   McDowell, 1994, p. 170.

But as McDowell goes on to tell us, embracing the identity of indiscriminables claim threatens to raise a problem due to the (supposed) non-transitivity of colour indiscriminability.[12] To say that colour indiscriminability is non-transitive is to say that there are shades A, B, and C where B is indiscriminable in colour from both A and C, but C is discriminable in colour from A. Call this *the non-transitivity claim*. Such a claim threatens to raise a problem if we accept the identity of indiscriminables claim. For if *both* claims are true, such shades as A and C would seem, despite their being discriminable, to have to be identical since they are both indiscriminable from (and hence identical to) a third shade, B.

McDowell's proposed solution to this problem,[13] which Brewer also accepts,[14] is to insist that a shade A need *not* be identical to a shade C just because both shades are indiscriminable from a further shade B; rather, shade A must *itself* be indiscriminable from shade C if the two shades are identical. The reasoning here seems to be that once the theorist has ensured that this further rule is in place, there will no longer be a danger that he will have to count shades that are discriminable as nevertheless identical.

However, this apparent benefit comes at too great a price. For if the theorist accepts the conceptualists' story here, he will be forced also to accept a wholly bizarre view of the nature of identity. For according to the conceptualists' proposal, we should not count shade A as identical to shade C (since they are discriminable) despite the fact that both shades are identical to (since they are indiscriminable from) a further shade B. But this is absurd: if shade A is *identical* to shade B, then we don't have *two* shades – shade A on the one hand, and shade B on the other - rather we have a single shade which is going by two names. And this single shade cannot be *both* identical to, *and* different from, some further shade C.

For this reason, then, it seems to me that the apparent tension between the non-transitivity claim and the identity of indiscriminables claim is in fact perfectly genuine: there really is no right way to commit oneself to both.

6.   Options for the Conceptualist

This is where Graff's arguments become relevant, since if (as I suspect) they do indeed establish that colour indiscriminability is transitive, it is clear what action the conceptualist should take: he should drop the non-transitivity claim.

However, if the conceptualist insists on rejecting Graff's view and retains his commitment to the non-transitivity claim, he will face serious consequences: he will then need to reject the identity of indiscriminables claim and with it *both* McDowell's *and* Brewer's constraints on demonstrative colour concept possession, since if the identity of indiscriminables claim is false, both constraints would be unsatisfiable. I shall now explain why this is.

Recall that according to McDowell's constraint, a possessor of a demonstrative concept of a shade featuring in an experience must have the capacity to tell, after the shade no longer features in the experience, whether

---

[12]   Ibid., p. 170-1.

[13]   Ibid., p. 171.

[14]   Brewer, 1999, p. 175.

or not a shade featuring in a new experience is identical to the original shade. However, *at best*, a subject might be able to tell if a shade featuring in a new experience is *indiscriminable* from a shade which featured in a previous experience. But if the indiscriminability of two such shades does not imply their identity, it seems impossible to see how a subject could, even in principle, have the capacity to tell if two shades are (not merely indiscriminable but also) identical. So if the conceptualist decides to drop the identity of indiscriminables claim*,* he must also concede that McDowell's constraint is unsatisfiable.

Precisely the same problem applies to Brewer's constraint: at best, the subject of an experience in which a particular shade features might have the capacity to tell, at a later stage of the experience, whether or not a shade featuring in the experience is *indiscriminable* from the original shade. But again, if the indiscriminability of the two shades does not imply their identity, it seems impossible to see how the subject could, even in principle, have the capacity at the later stage to tell if the relevant shade is (not merely indiscriminable but also) identical to the original shade. So again, if the conceptualist decides to drop the identity of indiscriminables claim*,* he must also concede that Brewer's constraint is unsatisfiable.

7.   Possible Objections (1)

For these reasons, then, my conclusion will be that the conceptualist should accept Graff's view that colour indiscriminability is transitive. In order to establish that this conclusion is correct, however, I now want to consider two objections to it that the conceptualist might try to make.

To anticipate, the first objector suggests that rejecting the identity of indiscriminables claim is actually quite consistent with maintaining that McDowell's and Brewer's constraints are satisfiable. The second objector, by contrast, concedes that rejecting the identity of indiscriminables claim *does* mean admitting that the two constraints are unsatisfiable, but he suggests that demonstrative colour concept possession need not require satisfaction of either constraint in any case.

Firstly, then, I want to investigate the prospects for the conceptualist should he maintain that he can drop the identity of indiscriminables claim without thereby admitting that McDowell's and Brewer's constraints are unsatisfiable. One way in which he might try to do this is by conceding that since he has rejected the identity of indiscriminables claim, he must indeed admit that it is impossible for a subject in whose experience a particular shade is featuring:

(1)   To tell, after the shade no longer features in his experience, whether or not a shade that features in a new experience is (strictly speaking) identical to the original shade; or,

(2)   To tell, at a later stage of the *same* experience, whether or not a shade that is featuring in the experience is (strictly speaking) identical to the original shade.

But the conceptualist might then claim that the (strict) non-identity of two such shades need not stop them *counting* as the same for the purposes of the proposed constraints. On this view, the constraints revolve around the idea

that in order to possess a demonstrative colour concept, a subject need not have the capacity to tell if some later shade is *strictly speaking* identical to an earlier shade; rather, he must have the capacity to tell if the later shade *counts* as the same as the earlier shade, where the indiscriminability of the two shades ensures that they do indeed at least count as the same.[15]

For after all, the conceptualist might point out, the key motivation which underlay the imposition of the constraints in the first place was that they were supposed to ensure that a possessor of a demonstrative colour concept must have some bona fide ability to recognise when a shade falls under that concept, which would in turn ensure that he can properly be said to *know what it is* for something to fall under that concept. And if two shades that feature in a subject's experience *count* as the same, in virtue of their indiscriminability, then there seems nothing to prevent that subject correctly bringing them under a single demonstrative colour concept.

The response to this proposal is immediate, however. For if the two shades in question are not (strictly speaking) identical, then they make a (strictly speaking) different contribution to the content of the experiences in which they feature. So if the two shades are brought under a single demonstrative colour concept, this means that the features are more fine-grained than the concept under which those features are brought. And this is *precisely* the original difficulty which the conceptualist's appeal to demonstrative concepts was supposed to solve. Hence this first objection fails: if the conceptualist decides to drop the identity of indiscriminables claim, he must indeed accept that McDowell's and Brewer's constraints on demonstrative colour concept possession are unsatisfiable.

8.   Possible Objections (2)

The second strategy that the conceptualist might pursue would involve conceding that if he rejects the identity of indiscriminables claim he must also admit that McDowell's and Brewer's constraints are unsatisfiable, but he might then question whether it follows from this admission that possession of demonstrative colour concepts is impossible. For there are good reasons to think, the conceptualist might claim, that demonstrative colour concept possession need *not* in fact require satisfaction of either constraint.

The idea here would be this. The original problem posed by Wittgenstein's case seemed to be that it placed the onus on the conceptualist to explain why there is a disanalogy between this case, on the one hand, and the case of the demonstrative colour concept possessor, on the other: it seemed that the conceptualist needed to tell us what it is that the demonstrative colour concept possessor *can* do, but Wittgenstein's subject *cannot* do, that ensures that the relevant concept is genuine.

---

[15]   Some of the conceptualists' remarks suggest that they themselves might be tempted by this view. For example, McDowell says: 'Something that counts as having a shade does not thereby count as a *sample* of the shade…' (1994, p. 171) One way (though not the only way) of interpreting this is as a concession that something which merely *has* a shade need not be strictly speaking identical in colour to something which is a *sample* of the shade. Note also Brewer's repeated talk of 'counting as a shade' in preference to talk of 'being a shade' (1999, p. 175).

But perhaps the correct response for the conceptualist to make here is not to impose the constraints on demonstrative colour concept possession that McDowell and Brewer do, but rather to claim that the very fact that the demonstrative colour concept possessor's experience features a particular shade is *itself* enough to put him in a position to know what it is for something to be that shade. (The conceptualist can happily admit that when the shade ceases to feature in his experience, he no longer has the relevant knowledge.) Now in Wittgenstein's case, the subject clearly does not know what it is for something to be 'this tall'. Or at least, this would be so if we imagine him (say) sitting in an armchair while performing his gestures. But in this case, the conceptualist might point out, the relevant property does *not* feature in his experience: if we rather imagine him gesturing in front of a mirror and so the property *were* featuring in his experience, it is by no means clear (the conceptualist might claim) that we would be so keen to deny that he had the relevant demonstrative concept.

I concede that there is indeed some initial plausibility in this suggestion. My response to it is to insist that (if the identity of indiscriminables claim is false, as the conceptualist is currently conceding) the mere fact that a subject's experience features a particular shade is not enough to enable him to know what it is for something to be that shade. For given that he cannot tell the difference between *that* shade (call it 'S1') and some *other* shade, call it 'S2', that is indiscriminable to him from S1, he simply does not know which shade S1 is.

Some may suspect that my response here is ineffective. For although the subject cannot tell the difference between S1 and S2, this is not in itself sufficient, the conceptualist might claim, to establish that he does not know exactly which shade S1 is. For he may be able to identify S1 *demonstratively*, the conceptualist might suggest, in such a way as to allow him to know exactly which shade it is.

The conceptualist might draw an analogy with the case of identical twins: suppose that I am looking at Amy, who has an identical twin, Becky, who looks so similar to Amy that I would be incapable of telling them apart. Does it follow that I do not know exactly who I am looking at? Seemingly not. If I am looking at Amy, it would seem plausible that I do know exactly who I am looking at, since I know that I am looking at *her* (as I might put it). And just as the fact that Amy has an identical twin is irrelevant to the question of whether I know exactly who I am looking at, the conceptualist might go on, so the fact that S2 is indiscriminable from S1 is irrelevant to the question of whether our envisaged subject knows exactly which shade it is that is featuring in his experience.

I have a two-part response to this line of argument. The first part is to suggest that the conceptualist's current appeal to demonstrative *identification* may be vulnerable to the same concerns that plagued his original appeal to demonstrative *concepts*. The conceptualist's original problem was to find a way to assure us that his envisaged demonstrative concepts really were genuine concepts. But his appeal to demonstrative *identification* at this later stage of the dialectic threatens to generate essentially the same difficulty: we still need a reason to believe that the supposed demonstrative identification really is a genuine form of identification.

The second part of the response is to suggest that the conceptualist's analogy with the case of identical twins does not supply us with such a reason. For although it is indeed plausible that I can identify twin Amy demonstratively in a way that allows me to know exactly who I am looking at, there is a key disanalogy between a case of this kind and the shades case.

The disanalogy consists in the fact that in the twins case, but not in the shades case, the relevant subject satisfies some version of Brewer's 'tracking' constraint on demonstrative concept possession: in the twins case, I can keep track of twin Amy while she continues to feature in my experience; in the shades case, by contrast, the subject cannot keep track of shade S1 while it continues to feature in his experience.

I shall illustrate this by describing two experiments that a scientist might perform, one involving the twins, and the other involving the shades. Suppose that in the first experiment, I am told to keep track of Amy continuously as she sits in front of me. At some point during the experiment, the scientist then sits Amy's sister Becky down beside her. The scientist then points to Becky and asks 'Did I tell you to keep track of *her*?'. My likely response to such a question seems clear: pointing to Amy, I would surely reply 'No – you told me to keep track of *her*.'

Suppose now that the scientist tells me to keep continuous track of shade S1, which happens to be instantiated by a particular card placed in front of me. (We can imagine the scientist emphasising the need for me to keep track of the shade that is instantiated by the card rather the card itself.) Then, after a period, the scientist places a second card down next to the first, but this second card instantiates shade S2. Pointing to the second card, the scientist then asks 'Did I tell you to keep track of *this* shade?'. Here my answer would surely be rather different: since S2 is indiscriminable to me from S1, I would be unable to tell that the shade instantiated by the second card is *not* the shade which I was supposed to be tracking.

Now recall that according to Brewer's constraint, in order to possess a demonstrative concept of a shade featuring in an experience, the subject must have the capacity to tell, at a later stage of the experience, whether or not a shade featuring in the experience is identical to the original shade. This condition seems not to be met in the envisaged case. For despite the fact that the original shade, S1, has featured continuously in my experience throughout the experiment, I am not in a position to tell that the second shade, S2, is different from the original shade. So according to Brewer's constraint, I lack a demonstrative concept of the original shade, S1.

An analogue of Brewer's constraint that applies to possession of demonstrative concepts of *objects* would look like this: in order to possess a demonstrative concept of an object featuring in an experience, the subject must have the capacity to tell, at a later stage of the experience, whether or not an object featuring in the experience is identical to the original object. Now this constraint *does* seem to be satisfied in the twins case: so long as Amy has featured continuously in my experience, I will immediately be in a position to tell that any *other* person (even her identical twin) that features in my experience is someone other than Amy.

Ultimately, then, the onus remains firmly on the conceptualist to explain how it is that in the shades case, even though neither McDowell's nor Brewer's constraint on demonstrative colour concept possession is met, I can

nevertheless identify shade S1 demonstratively. Certainly, the fact that demonstrative identification is possible in the twins case, when (an analogue of) Brewer's constraint *is* satisfied, gives us no reason to suppose that it is also possible in the shades case, when the constraint is *not* satisfied.

In the absence of any further conceptualist explanation, then, I conclude that my original response to the current objection remains correct: if the identity of indiscriminables claim is false, the mere fact that the supposed demonstrative colour concept possessor's experience features a particular shade is not enough to put him in a position to know what it is for something to be that shade, since it does not enable him to know which shade it is. And this in turn casts serious doubt on the idea that the supposed demonstrative concept is a genuine concept at all.

9.   Conclusion

I will end by recapping on the structure of my argument. I have been arguing that the conceptualist should accept Graff's view that colour indiscriminability is transitive. He should do this since only by so doing can he retain a commitment to the identity of indiscriminables claim. This in turn is essential since only by retaining such a commitment can he accept the idea that (either of) the constraints that McDowell and Brewer impose on demonstrative colour concept possession are satisfiable. Finally, he does indeed need to accept this idea since there seems good reason to think that demonstrative colour concept possession is not possible if neither constraint is met.

References

Brewer, B. 1999. *Perception and Reason*. Oxford: Oxford University Press.

Brewer, B. 2004. 'Do Sense Experiential States Have Conceptual Content?' Forthcoming.

Campbell, J. 2002. *Reference and Consciousness*. Oxford: Oxford University Press.

Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.

Graff, D. 2001. 'Phenomenal Continua and the Sorites.' *Mind* 110, 905-936.

Heck, R. 2000. 'Non-Conceptual Content and the "Space of Reasons".' *Philosophical Review*, 109, 483-523.

Kelly, S. 2001. 'Demonstrative Concepts and Experience.' *Philosophical Review*, 110, 397-420.

McDowell, J. 1990. 'Peacocke and Evans on Demonstrative Content.' *Mind* 99, 255-66.

McDowell, J. 1994. *Mind and World*. Cambridge, Mass.: Harvard University Press.

Peacocke, C. 1991. 'Demonstrative Content: A Reply to John McDowell.' *Mind* 100, 123-33.

Peacocke, C. 1992. *A Study of Concepts*. Cambridge, Mass.: MIT Press.

Wittgenstein, L. 1958. *Philosophical Investigations*. Oxford: Blackwell.

Relational Looking and the Reversibility Constraint

Richard Price

Department of Philosophy, University of Oxford

richard.price@ccc.ox.ac.uk

In this paper I am interested in cases of *relational looking*; cases where one object looks to you to bear a relation to another object. In particular, I am interested in whether you have to *see* objects A and B in order for A to look to you related to B. Here are two examples that bring out what I have in mind. In the first example, you see two lines, A and B, before you, and you say that A looks longer than B. In the second example, you see only one building in front of you (that is not the Eiffel Tower), and you say that *this building* looks to you the same height as the Eiffel Tower. In this paper I shall discuss the relation between saying that A looks to you longer than B, and saying that *this building* looks to you the same height as the Eiffel Tower. In particular, I will discuss whether both statements are true in the same sense of 'looks'.

This is an issue to do with how rich the content of experience is. When you look at a building and say 'this building looks to me the same height as the Eiffel Tower', the question is whether it is part of the content of your experience that the building before you is the same height as the Eiffel Tower. In this paper, I develop a constraint on relational contents of representational states, which I call the 'reversibility constraint', and I argue that the case of the building before you looking to you the same height as the Eiffel Tower does not satisfy this constraint—hence the relation of being the same height as the Eiffel Tower is not part of the content of experience. By contrast, when you see lines A and B, and A looks longer to you than B, this case does satisfy the reversibility constraint, and so it is an open possibility that it is part of the content of experience that A is longer than B.

In the second section of the paper, I introduce a principle I call the 'Content Principle', which I use to define the notion of the content of experience. I then argue that the content principle seems to entail that the relation of being the same height as the Eiffel Tower can be in the content of experience, even when you do not see the Eiffel Tower. I then suggest an intuitive amendment of the content principle to resolve this puzzling conflict between the reversibility constraint and the content principle.

In the third section of the paper, I discuss how the reversibility constraint can do useful further work in demarcating the content of experience, especially in contentious areas. For instance, it is a matter of controversy whether the property of having a back, and, by extension, the property of being a three-dimensional physical object can occur in the content of experience. I will argue that the reversibility constraint rules out from the content of experience the property of having a back, and, by extension, the property of being a three-dimensional object from being in the content of experience.

In addition the reversibility constraint rules out the property of being to the left of me, the property of being far away from me, the property of being circular-and-at-a-slant-from-me from the content of experience. This raises a

question about what kinds of location properties objects can look to have. I discuss the options, absolutist and relationalist location properties, and I reject both. Using the reversibility constraint, I also reject the notion that objects look to be at certain visual angles from us. This leads to a puzzle that lies unresolved at the end of the paper, namely what kinds of location properties objects do look to have.

I

Consider S's belief that A is the same height as B. This is a belief with a relational content. When S believes that A is the same height as B, she believes something about A, and she also believes something about B. What she believes about B is that it is such that A is the same height as it. So, if A is believed by S to be the same height as B, then B is believed by S to be such that A is the same height as it. This principle holds as a matter of necessity: it would be nonsensical to say that you believe that A is longer than B, but you don't believe anything about B. Necessarily, in believing that A is longer than B, you believe that B is such that A is longer than it. This principle seems to hold for all representational content. If A is represented as being taller than B, then B is represented as being such that A is taller than it; i.e. one can always reverse the order of A and B, to say how B is represented as being. Let us express this constraint on relational contents as follows:

Reversibility Constraint:
Necessarily, if A is represented as bearing relation R to B, then B is represented as being such that A bears R to it.

A stronger version of the Reversibility Constraint may be true as well. You might think that, if someone believes that A is taller than B, then this is exactly the same as believing that B is less tall than A. Or that, if one believes that A is the same colour as B, then this is the same as believing that B is the same colour as A. Or that, if one believes that James hit Fred, then this is the same as believing that Fred was hit by James. The idea is that for every relation, R, that holds between A and B, there is a reverse relation, reverse-R, that holds between B and A. According to the stronger version of the reversibility constraint, if you believe that A bears R to B, this is the same as believing that B bears reverse-R to A. The weaker reversibility constraint does not require that you believe that B bears reverse-R to A. The weaker version merely requires that you believe that B is such that A bears R to it.

It is a difficult question whether the stronger version of the reversibility constraint is true. There is some temptation to say that, if a subject believes that A bears a very complex relation R to B, she may not know whether this complex relation is symmetric or not, so she may not have a belief about whether B bears R to A. The stronger version of the reversibility constraint cannot accommodate this possibility. According to the stronger version, if R is a symmetric relation, then in virtue of believing that A bears R to B, S believes that B bears R to A. The counter-intuitiveness of this conclusion may, however, be tempered by the fact that the stronger version of the reversibility

constraint does allow for the possibility that the subject does not have any beliefs about whether R is a symmetric relation.

This puzzle is not restricted to relational beliefs. If Susanna believes that it is not the case that Jim isn't hot, is that the same as Susanna believing that Jim is hot? These are difficult issues to resolve, but fortunately we do not need to resolve them for our purposes. For our purposes, all that needs to be true is that, if A is believed to be taller than B, then B is believed to be some way: namely, such that A is taller than it. (I am grateful to Stephen Kearns here. The above are developments of an objection that he put to the stronger version of the reversibility constraint that made me think that a weaker version was needed).

Take the example of the two seen lines, A and B, where A looks longer to you than B. For it to be in the content of your experience that A is longer than B, that content would have to satisfy the reversibility constraint. And that content does satisfy the reversibility constraint. For if A looks longer to you than B, then B looks to you to be such that A is longer than it.

But consider the case where you are looking at a building before you and say '*this building* looks to me the same height as the Eiffel Tower'. Here the relational looks statement does not satisfy the reversibility constraint. One cannot say 'the Eiffel Tower looks to me to be such that this building is the same height as it', because the Eiffel Tower does not look any way at all to you at the time, since you don't see it.

Of course, there may a sense of the word 'looks' in which the Eiffel Tower looks to me such that this building is the same height as it. If this sense exists, it is the sense in which the Eiffel Tower *tends to look to me* such that this building is the same height as it. But this wasn't the sense of 'looks' in the original looks-statement. When it was said that this building looks the same as the Eiffel Tower, this did not mean that this building *tends to look to me* the same height as the Eiffel Tower—after all, this building may *not* tend to look to me the same height as the Eiffel Tower.

The point illustrated in the last paragraph can be made for the beliefs too. Imagine that the Eiffel Tower is believed by Fred to be taller than Canary Wharf is normally believed to be. (i.e. Fred believes that: the Eiffel Tower is taller than Canary Wharf is normally believed to be). The reversal of this content in accordance with the reversibility constraint is not: Canary Wharf is normally believed by Fred to be such that the Eiffel Tower is taller than it. After all, Canary Wharf might not be normally believed by Fred to be any way at all. Rather, the correct reversal is: the height Canary Wharf is normally believed to be is believed by Fred to be such that the height of the Eiffel Tower is greater than it.

One might think that when we say 'this building looks to me the same height as the Eiffel Tower', what we really mean is that this building looks to me the same height as the Eiffel Tower *normally looks*. The above application of the reversibility constraint may have demonstrated that the perceptual content does not contain the relation of being the same height as the Eiffel Tower, but it may yet allow that the perceptual content contains the relation of being the same height as the Eiffel Tower normally looks.

However, 'this building looks to me the same height as the Eiffel Tower normally looks' does not satisfy the reversibility constraint either. The correct reversal of the supposed content is not: the Eiffel Tower normally looks the

same height as this building. The correct reversal is rather: the height the Eiffel Tower normally looks to have looks to be such that the height of this building is the same as it. Even if it is intelligible for the heights of objects to look certain ways to you (as opposed to the objects themselves), the problem is the same one as before: the height that the Eiffel Tower normally looks does not look any way to you right now, since you do not see that height.

What has been demonstrated thus far is that, when it is true that this building looks the same height as the Eiffel Tower, and you don't see the Eiffel Tower, there is no experience with the content <this building is the same height as the Eiffel Tower>. And when it is true that this building looks the same height as the Eiffel Tower normally looks, and you don't see the height that the Eiffel Tower normally looks, there is no experience with the content <this building is the same height as the Eiffel Tower normally looks>.

II

In section III, I intend to consider how the reversibility constraint can do further work in demarcating the content of experience. In this section, however, I will discuss a principle, that I call the 'Content Principle', that seems to conflict with the conclusion of section I; that is, the content principle seems to entail that even when you don't see the Eiffel Tower, you can have an experience which includes as part of its content the relation of being the same height as the Eiffel Tower. This is puzzling, since both the reversibility constraint and the content principle seem to be true, and yet there is evidence that they conflict.

Why do we believe that perceptual experiences have contents? What seems to be central to the idea that experiences have contents is that there is such a thing as an object being the way it looks, or failing to be the way it looks. For instance a square may look triangular, and thus not be the way it looks. The same idea makes us think that beliefs have contents: there is such a thing as an object being the way it is believed to be, or failing to be the way it is believed to be.

Thus, central to the idea that experiences have content is the idea that how objects look places a condition on the objects, namely that the objects have to be a certain way to be the way they look. Let us spell out this intuition in the following principle:

Content Principle:
Necessarily, there is a sense of the actual sentence 'A looks F' which is true iff:
    (i)   necessarily, if A looks F and A is not F, then A is not wholly the way it looks.
    (ii)   necessarily, if A looks F and A is F, then A is partly the way it looks.

In using the conditions in the content principle to give necessary and sufficient conditions for perceptual content, I have assumed that the veridicality conditions of a given experience are equivalent to the content of that experience, and hence there are no modes of presentation in the content of experience. I believe this assumption to be correct, though I do not have the

space to defend it here. In addition, the puzzle I set out below could be rephrased in terms of the veridicality conditions of experience, rather than its content.

There will be some expressions of the form 'A looks F' which do not satisfy the two conditions in the content principle—i.e. cases where it is true to say 'A looks F', even though A does not have to be F to be the way it looks. The project of demarcating the content of visual experience is equivalent to the project of deciding for which statements of the form 'A looks F' the conditions in the content principle are satisfied. For example, the first section of this paper has argued that, if the Eiffel Tower is not seen, the sentence 'A looks the same height as the Eiffel Tower' cannot correctly report the content of visual experience. Therefore, when the Eiffel Tower is not seen, the statement 'this building looks the same height as the Eiffel Tower' will not be true in the sense of 'looks' defined by the content principle: i.e. it will not be the case that A has to be the same height as the Eiffel Tower to be the way it looks.

In a way, this result is quite intuitive. Imagine a situation in which it is true that A looks the same height as the Eiffel Tower, even though we don't see the Eiffel Tower. We wouldn't count ourselves as misperceiving A if, in the course of our experience, the Eiffel Tower suddenly shrunk so that it was no longer the size of A—i.e. so that even though A looked the same height as the Eiffel Tower, it wasn't the same height as the Eiffel Tower. This suggests that the truth that A looks to me the same height as the Eiffel Tower does not require that A be the same height as the Eiffel Tower to be the way it looks. At this stage, the results of the reversibility constraint seem to fit well with our intuitions about what would count as a misperception of A and what would not.

However, the problem is that a respondent might object: but if the Eiffel Tower had suddenly shrunk during the course of our experience, then from then onwards, it would *no longer* be true that A looked the same height as the Eiffel Tower. The content principle only applies to cases where A does look the same height as the Eiffel Tower. In these cases where A does look the same height as the Eiffel Tower, we then ask, does A have to be the same height as the Eiffel Tower to be the way it looks? And it seems that the answer to this is 'yes'. For the respondent will say that if A looks the same height as the Eiffel Tower, then if the Eiffel Tower has a height of H, it follows that A looks H. If A looks H, then A must be H to be the way it looks; and since H is the same height as the Eiffel Tower, it follows that A must be the same height as the Eiffel Tower to be the way it looks. Moreover, any world in which H equal to the height of the Eiffel Tower, but A looks H, will not be a world, according to the respondent, in which A looks the same height as the Eiffel Tower, and hence will be irrelevant to the question of whether the conditions in the content principle are satisfied. The respondent will argue that the point in this paragraph shows that condition (i) of the content principle is satisfied.

The respondent has a good case for saying that condition (ii) of the content principle is satisfied too. If the Eiffel Tower is H, and A is the same height as the Eiffel Tower, then that means that A is H too; given that A looks H, it follows that A is partly the way it looks. Hence both conditions of the content principle have been met. Since the content principle was introduced to capture what we mean by perceptual content, it seems to entail that the

relation of being the same height as the Eiffel Tower is in the content of one's experience.

The puzzle, therefore, is that the reversibility constraint rules out the relation of being the same height as the Eiffel Tower from being in the content of visual experience when the Eiffel Tower is not seen. And yet the content principle, which is used to capture the notion of the content of experience, seems to allow that the relation of being the same height as the Eiffel Tower is in the content of experience even when the Eiffel Tower is not seen. I shall now suggest a possible resolution of this puzzle.

Let us say that an object, A, looks F. When considering whether A has to be F to be the way it looks, what we must do is consider all those worlds in which A looks the same as it does when, in the actual world, A looks F. Call that set of worlds S1. Now we must consider all those worlds within S1 in which A is not F, and ask whether, in all those worlds, A is or is not the way it looks. If there are ~Fa-worlds in S1 in which A is nonetheless the way it looks, then the conditions in the content principle have not been satisfied, and 'A looks F' does not report the content of visual experience.

What is in common amongst the worlds in S1 is that A looks the same as it does when, in the actual world, A looks F. Let us define S2 to be the set of worlds in each of which it is true to say that A looks F. What is key to resolving the puzzle is recognizing that S1 and S2 are not necessarily the same set. This is because there may be worlds in S1 in which A does not look F, even though in these worlds A looks the same as it does when, in the actual world, A does look F. This may sound bizarre. There is an intuitive principle, that, if two objects, A and B, look the same, then if it is true that A looks F, then it must be true that B looks F. Call this the 'Looks the Same Principle'.

What the puzzle brings out, I believe, is that the Looks the Same principle is false. An object, A, might look the same from t1 to t3. At t1, it was true to say 'A looks the same height as the Eiffel Tower'. At t2, the Eiffel Tower shrinks in height, so that at t3, it is not true to say 'A looks the same height as the Eiffel Tower'. However there is no change whatsoever in the way A looks from t1 to t3. This example is sufficient to show that the Looks the Same principle is false. Recognizing the falsity of this principle helps us to see how to resolve the puzzle.

When considering whether A has to be the same height as the Eiffel Tower to be the way it looks, when it looks the same height as the Eiffel Tower, contrary to the content principle above, we should not consider the set of worlds in which A looks the same height as the Eiffel Tower. Rather, we should consider the set of worlds in which A looks the same as it does when, in the actual world, A looks the same height as the Eiffel Tower. This set, call it S3, will contain worlds in which A does not look the same height as the Eiffel Tower (because the Eiffel Tower is much smaller or bigger in those worlds compared with ours), but it looks the same as it does in worlds in which it does look the same height as the Eiffel Tower. This is the first step to resolving the puzzle.

In S3, there will be a world, *w*, in which the Eiffel Tower has a height of ½ H, whereas in our world the Eiffel Tower has a height of H. Let us say that in *w*, A looks H, and is H, so it is the way it looks. *w* shows that the statement 'A looks the same height as the Eiffel Tower' does not report the content of experience, since in *w*, A looks the same as it does in the actual world, A is

also the way it looks, and A is not the same height as the Eiffel Tower. In general, if A looks F, and we ask whether A has to be F to be the way it looks, we clearly have to keep the 'look' of A fixed. But the way to do this is not to keep fixed all of the truth-values of looks-statements involving A. That is what created the puzzle above. Rather, what we have to do is draw on our grasp of what it is for A to look the same that is independent of keeping all the truth-values of looks-statements about A fixed. This is another way of saying that the Looks the Same principle articulated above is false. The correct formulation of the content principle is as follows:

> **Content Principle***:
> Necessarily, there is a sense of the actual sentence 'A looks F' which is true iff:
>
> (i)   necessarily, if A looks the same as it does when, in the actual world, it looks F and A is not F, then A is not wholly the way it looks.
>
> (ii)   necessarily, if A looks the same as it does when, in the actual world, it looks F and A is F, then A is partly the way it looks.

Rather than conflicting with the reversibility principle, the content principle* supports the conclusion that, when the Eiffel Tower is not seen, one's visual experiences do not include in their contents the relation of A being the same height as the Eiffel Tower. This is because of worlds such as *w* above, in which A looks the same as it does when, in the actual world, it looks the same height as the Eiffel Tower, and in which A is the way it looks, and in which A is not the same height as the Eiffel Tower.

III

In this section I will discuss some further work that the reversibility constraint can do in demarcating the content of experience. There is a question as to whether experience represents things as tomatoes, tables and persons, or whether experience represents a much sparser range of properties, for instance colours and shapes. As the content principle illustrates, what these questions amount to is asking whether it is ever the case that an object has to be, say, a tomato to be the way it looks. Of course, we do say sentences such as 'A looks to be a tomato', but the question is whether these sentences are ever true in the sense that A has to be a tomato to be the way it looks.

It seems that part of what would be involved in representing an object as a tomato would be representing the object as a three-dimensional physical object; and part of what is involved in representing an object as a three-dimensional physical object would be representing the object as having a back.. This is where the reversibility constraint does some work. Just as the statement 'I have a hand' expresses a relation between me and one of my parts, so the statement 'this object has a back' expresses a relation between it and one of its parts. For that relation to be part of the content of experience, it would have to be reversible, i.e. if the sentence 'this object looks to me to have a back' is true, then the sentence 'a back looks to me to be such that this object has it' will be true. Yet, the latter sentence will not be true, since if I

do not see the back of the object, a back does not look any way to me at all. Therefore the reversibility constraint entails that, when we do not see the backs of objects, our experiences do not represent that the objects have backs.

We often say sentences such as 'this objects looks to the left of me'. Is it part of the content of my visual experience that the object is to the left of me? According to the reversibility constraint, this relation is part of the content of experience only if one can reverse the relata in the relation and say 'I look to be such that this object is to the left of me'. However, if I do not see myself, I do not look to myself to be any way at all. For similar reasons, the sentence 'A looks far away from me' will not truly report the content of experience, since I do not look to myself to be such that A is far away from me. And the same points apply to 'A looks in front of me'. The reversibility constraint also rules out the sentence 'this coin looks to me circular-at-a-slant' from reporting the content of experience, since that sentence is elliptical for 'this coin looks to me circular-at-a-slant-from-me', which will not be reversible if I do not see myself.

One might think that the reversibility constraint allows the sentence 'A looks further away than B' to report the content of experience, if A and B are both seen. But in fact this statement is elliptical for 'A looks further away *from me* than B', and this statement is not reversible if I do not see myself. Similarly, even if one sees A and B, that does not mean that the sentence 'A looks to the left of B' can correctly report the content of experience, since that sentence is elliptical for 'A looks further to *my* left than B', and this statement is not reversible if I do not see myself.

If one is unable to use properties such as being to the left of me to characterize the locations objects look to be in, how should one characterize these locations? One idea would be to use the notion of a visual angle. An object might look to be at a certain angle, say 90 degrees. But being at 90 degrees is really the relation of being at 90 degrees from me (I am grateful to Hemdat Lerman here). Hence, the reversibility constraint entails that objects do not look to be at 90 degrees from me, if I do not see myself. This raises the question: when we talk about objects looking to be at certain locations, what kinds of properties do we mean?

According to relationalists, the property of being at a certain location is a matter of standing in relations to other objects in the world. However, given that one does not see all the objects in the world, the reversibility constraint would rule out relationalist location properties from being in the content of experience. According to absolutists, space is considered to be analogous to a large container, and every point in space has an absolute xyz coordinate, where the relevant axes are taken to be analogous to the sides of the container. Might visual experience represent absolutist location properties? It seems that an absolutist could not provide a full account of the location properties in experience, for the following reason. Imagine that I am looking at a red square on my left and a green square on my right. If I then move round and look at them from the opposite perspective, the way things look change. However, if experience represents absolutist location properties, then the content of experience has not changed at all, since the objects will still look to be in the exactly the same absolute location as they looked to be in from the original perspective. The absolutist is unable to explain how the way things

look seems to change as one switches perspective on the same set of objects.

This latter example also makes another point clear. Whatever location properties visual experience does represent, they must be such that, when two people are looking at the two coloured squares from opposite angles, the location properties the squares look to have are such that the squares can actually have them. We do not want to be forced to say that there is only one perspective from which one can veridically perceive the squares' locations.

The points in the above two paragraphs can be combined to specify the two roles that must be played by the location properties that objects look to have. Firstly, they must be able to explain how the way things look changes as your perspective changes; secondly, the location properties that an object is represented as having by experiences from multiple perspectives must be capable of being had by one and the same object (i.e. experiences from multiple perspectives must be capable of all being veridical). The puzzle is that the properties that seem ideally suited to play these two roles are properties such as 'being to the right of me at such and such an angle'; and yet, the reversibility constraint rules out these kinds of properties from being in the content of experience. Thus, what kinds of location properties are represented by experience? Given how closely related size and location properties are, we can ask: what kinds of size properties are represented by experience? These seem to me to be extremely interesting questions, and ones that deserve some further investigation.

Before finishing I will mention two further accounts that are in conflict with the reversibility constraint. The first is the relational view of colour properties. Some hold that being red is the property of looking red to normal observers in normal conditions. If one accepts the assumption there are no modes of presentation in the content of experience (and, it seems that there is some pressure to do this if one believes in a nonconceptual theory of perceptual content—after all, what could these modes of presentation be other than concepts?), then, on this view, representing something as red is a matter of representing the thing as looking red to normal observers in normal conditions. Yet representing this latter relation would violate the reversibility constraint, given that you do not see normal observers whenever things look red to you.

The second account in conflict with the reversibility constraint is John Searle's account of perceptual content. In order to rule out the possibility of veridical hallucination, John Searle claimed that one's experiential contents have an implicit causal and self-referential component (Searle, 1983). He claimed that, when you are looking at a yellow car before you, the content of your experience is <there is a yellow car before me that is causing this experience>. However, given that one does not see one's own experiences, the reversibility constraint would rule out from the content of experience the relation of causing one's experiences. One's experiences do not look to one to be such that the object before one causes them.

The reversibility constraint seems an essential constraint on the relational contents of representational states. It has been shown to do much useful work in clarifying what belongs in the content of experience and what does not. It has also raised some very interesting questions regarding what

kinds of location and size properties experience represents, and these questions will be the subject of further investigation.

Reference

J. Searle, *Intentionality*, 1983, CUP.

Perception Involving Action

Susanna Schellenberg

Pittsburg/Oxford

susannaschellenberg@hotmail.com

It is striking that only self-movers are perceivers. I will argue that it is no coincidence. Although passivity on the part of the perceiver is an essential feature of perceptual experience, I aim to show that perceivers are not just passive receivers of information. The idea that perception is dependent on action has a long history in philosophy, but has barely been elaborated in detail. Aristotle can be read as arguing that only beings that are self-movers can perceive. Alva Noë goes farthest with the claim that perception is dependent on action. He writes: "What perception is … is a kind of skillful bodily activity".[1] I do not want to go as far. However, the connection I draw between perception and action is stronger than that perception is a *means* to action and action is a means to perception. No doubt our perceptions *guide* our actions and our actions facilitate us having perceptions of different objects. Action and perception are certainly related in such an instrumental manner. I aim to bring out, however, that the capacity to perceive is furthermore necessarily and intrinsically dependent on the capacity to act.

Two arguments will be considered for this thesis. In Part 1, I present a way of thinking about perception that will motivate what I will be arguing for in the rest of the paper and thereby will lay out what I will be taking for granted. In Part 2, I argue that the capacity to perceive objects in objective space involves practical knowledge of how one's perception changes as one's spatial relation to perceived objects changes. This could be called the sensorimotor knowledge argument. It is subject to a host of objections, but— as I bring out in Part 3—if one uncovers a more fundamental connection between action and perception that this argument depends on, the objections can be put to rest and what is attractive about the sensorimotor knowledge argument can be retained. In Part 3, I argue that the capacity to form sensorimotor knowledge is dependent on a perceiver being *aware* that she is the acting perceiving subject. The self-awareness in play is understood practically in the sense that a perceiver understands herself as occupying one location in space from which she both acts and perceives. This will be called the self-awareness argument.

Before I embark on this project, it is necessary to make two terminological points. When I speak of action I do not have a notion in mind that has anything to do with reason giving practices, as the notion is usually used today in philosophical debates on action. I use the term action since that is the term used in the lively discussion of these matters in cognitive psychology today. I will discuss in more detail what I have in mind towards the end of this paper. Among other things it will have to be discussed whether the activity in question must be self-activated and whether it must be intentional. I

---

[1]   Alva Noë (forthcoming), *Action in Perception* (Cambridge, MA: MIT Press), p.3.

will take for granted that intentional actions are not necessarily actions for reason.

As a second terminological remark, when I speak of perception, as I will explain in a moment, I mean *visual* perception of objects as three-dimensional space-occupiers. So I have a fairly rich notion of perception in mind. I hope that my argument holds for modes of perception other than visual spatial perception, but I will not venture into that in this paper.

I – Egocentricity Argument

Trivially, to perceive, one must perceive from a point of view. We perceive objects *from* our own position in space and *in relation* to our own position in space. We do not perceive objects simply to the left or to the right without perceiving them as being to *our* left or to *our* right and in this sense we do not just register how things are in perception, we perceive how things look from our point of view.

If perception is relational, then a perceiver's vantage point must play a role in her perception. Indeed, I want to say that the perceiver's point of view must figure in the content of perception for the content to present itself as perception of an objective spatial world.[2] But how can the location of the perceiver—which is simply a fact about the world—play a role at a semantic level? In order to play a role in perceptual content, the location of the perceiver must gain a place at a cognitive level. A first step in that direction can be to say that in perceiving objects as spatially related to herself, a perceiver gains awareness of her self as located in the perceived world. Understanding egocentric spatial perception as self-locating in this way involves an element of circularity, but a circularity that is not objectionable in itself. At the most basic level, the ability in question involves that a perceiver understands that she occupies the spatial location *between*, say, the bookshelf and the computer. The subject of perception is thereby conceived of as located in the world as a geometrical point of view on the world and the location of this geometrical point is defined by reference to the egocentric spatial content of the subject's perceptions.

But the capacity to locate oneself is not sufficient to perceive objects as three-dimensional space-occupiers. It is only if a perceiver can abstract from her position in space and understand herself as one object among others that she can gain the understanding of objective space necessary to perceive the spatial properties of objects independently of the point of view she happens to have in any particular situation of perception. How can this requirement be met? In order to abstract from her point of view and gain a conception of herself as one object among others, a perceiver must not only be able to understand that she occupies the spatial location between, say, the bookshelf and the computer, but furthermore needs to understand what it would mean to occupy the spatial location that is now occupied by the computer. Another

---

[2]   Christopher Peacocke (1999) can be read as developing this thought in his *Being Known* (Oxford: Clarendon Press).

way of expressing this thought is that perceiving objects in objective space involves being able to imagine having alter-ego points of view and conjoining these different *possible* egocentric points of view to gain an understanding of objective space. This idea must not be understood as suggesting that there are two kinds of space: egocentric and objective space. There is only one space. However, the positions of objects in space can be characterized in more or less egocentric terms. Similarly, there are not two kinds of perceivings: perceivings from a particular point of view and perceivings of objects as objects that fill out a certain space.

In this section, I laid out what I am taking for granted and presented a way of thinking about perception that will motivate the argument in the rest of this paper. In particular, I motivated the view that there is an interdependence between how objects appear to us in egocentric space and how we perceive them in objective space. In the next section, I will take a closer look at this interdependence.

II – Sensorimotor Knowledge[3]

The fact that objects are always perceived from a particular point of view does not challenge the objectivity of our perceptions. Our perceptions do not simply *record* how things appear to us. It is more complex than that. We see the plate as round, despite its appearing elliptical from our vantage point. How does the way objects appear to us play a role in how our perceptions represent the world? While the *actual shape* is determined by how an object fills out space, I will understand its *apparent shape* as determined by its actual shape and the perceiver's spatial relation to the object. This is what I will call the appearance of the object. Appearances are not mental items, but a fact about the world. They are relational, but they are not determined by relations between objects and their sensory effect in us, but rather by the shape of objects and the perceiver's location in relation to the object. Although the motivation for speaking about appearances is to say something about perception, the relational property that brings about an appearance exists independently of any perceiver. An appearance of a round object just is the shape of the object as projected on to a plane that has a specific spatial location to the round object in question. This is very important, since it is because of this that the view suggested will not lead into a phenomenalist view.

Although we always perceive an object from a particular vantage point, we are able to grasp how the object looks independently of the point of view we happen to have in any particular situation. As I argued in the previous

---

3   The view presented in this section is a variation of a standard view in cognitive psychology. Anyone familiar with Susan Hurley and Alva Noë's work will realize that I owe much of the specific formulation of the sensorimotor view to their accounts. The view as it is described here differs from Hurley's as well as Noë's view in that my focus is exclusively on relations between perception and action on a personal level. See in particular Susan Hurley (1998), *Consciousness in Action* (Cambridge, MA: Harvard University Press) and Noë (forthcoming).

section, there are not two kinds of space or two kinds of perceivings. Accordingly, the question that needs to be addressed is *not* how to bridge the gap between egocentric and objective space or between how objects appear to us and how we in fact perceive them. The question is rather how it is that we perceive things as in objective space, although we always perceive things from a particular point of view and how the way objects appear to us plays a role in how our perceptions represent the world.

I aim to bring out that the capacity to perceive an object is only intelligible together with the practical ability to spatially orient ourselves in relation to objects. When we visually experience an object, say as round, we do so because of implicit, practical knowledge of the way the object's appearance varies in the characteristic way that the appearances of round objects vary as our relation to the perceived object changes. The details of this point are crucial. The most modest claim is that perception involves implicit, practical knowledge of the effects of movement on perception. A stronger claim is to say that perceptual content is constituted by a perceiver's ability to exercise bodily skills. In elaborating on these ideas, it will be necessary to gain a clearer understanding of what it means to say that an object appears to me in a certain way. Of course the way an object appears to us should not be understood as an epistemic intermediary between what we encounter in experience and the perceptual content of our experience.

The idea I would like to exploit is that we perceive a plate as round *because* we have encountered its elliptical appearance. It is important that this should not be read as suggesting that we *first* see a plate as elliptical and only later come to realize that it is in fact round. Nor is the suggestion that we do not *actually* see plates as round. It is true that one appearance in isolation may be misleading. But we do not *learn* to see round objects as round and we do not reflect on how objects appear to us and then arrive at judgments about how we should see them. But I do want to say that we see the plate as round *because* it appears elliptical to one.[4] In order to have perceptions of objects in objective space a perceiver employs (implicit, practical) knowledge of the effects that changes in her spatial relation to objects have on her perceptions. More specifically, a person's perception of an object, say a cube, is determined by practical knowledge of the form "If I were to move to the right, my perception of the cube would change thus and so, namely in the characteristic way that the perception of cubes varies as a perceiver's spatial relation to the perceived object changes." The grasp of such practical conditionals between action and perception is a kind of practical knowledge. No doubt, it would be possible to have such knowledge explicitly. But what is involved in perception need not be explicit knowledge. I will call this practical knowledge "sensorimotor knowledge". It might be misleading to speak of knowledge in this context, even if one stresses the practicality of the

---

[4]   The idea underlying this thought is the same idea that motivates Leibniz's distinction between grandes and petites representings. On Leibniz's view our perception of a the ocean roaring, to use his example, is constituted of a multitude of micro perceptions of the noise that a grain of sand makes when water crashes on it. We are not aware of the noise that every single grain of sand makes when listening to the roaring of the ocean. Nonetheless, we hear the roaring of the ocean *because* we hear the noise of many grains of sand.

knowledge. An alternative term would be "sensorimotor skill".[5] The term "skill" suggests, however, that what is in question is *gained* through practice, which is a view I would like to avoid for reasons that I will lay out shortly.

To develop the specific way that I would like to understand the sensorimotor knowledge involved in perception, it will be helpful to think through some central ideas of Husserl's account of protention.[6] On Husserl's view, perception of time is not atomized into a series of discrete instants. Rather, our time-consciousness is a continuous flux: when listening to a tune, at any given time, we have a 'primal impression' of the note that is occurring now, note 1. When we hear the next note, note 2, we no longer have a primal impression of note 1, but we retain it: we are aware of it as just past. As the tune proceeds, the first note recedes further into the past and appears in ever changing 'retentional modifications'. Furthermore, at any given point in the tune we 'protain' its future course. To say that we 'protain' what is to come when listening to a tune does not mean that we hear into the future.[7] Husserl distinguishes retention from memory in that when remembering a note our attention is directed at the note past and thus our perception of the present phase is impaired. By contrast, when we retain a note of a tune our attention is not directed at that note, but rather at the note that is currently to be heard. In the very same way, Husserl distinguishes protention from expectation.

When perceiving a cube, we do not *see* the surface of the cube facing away from us. As on Husserl's account of protention, we do not *direct* our attention at the surface of the cube facing away from us. Although our attention is directed at the surface of the cube facing towards us, there is a sense, however, in which we perceive the surface facing away from us as well. We have expectations (expectations that do not involve actively directing our attention at what is imagined) of what an object looks like from points of view we do not have in the particular moment of perception. In this sense, our perception of objects is not limited to the information projected onto the retina. When we perceive a cube, of course, we never see the whole cube. But it is *as if* we perceive the whole cube since we know that *if* we were to move thus and so, we *would see* the cube from the other side and eventually be visually confronted with every angle of the cube. The sensorimotor knowledge that figures in the expectations we have of what an object looks like from other perspectives involves entertaining the possibility of having different vantage points to the object perceived.

This sensorimotor knowledge argument is subject to a host of objections. I will consider three objections.

---

[5]   This is the term that Alva Noë uses to describe a similar capacity.

[6]   See Edmund Husserl (1905), *Lectures on the Phenomenology of Internal Time-Consciousness*, tr. John Barrett Brough (Dordrecht: Kluwer Academic Publishers, 1991).

[7]   This understanding of perception is distinct in several ways from Merleau-Ponty's understanding. For an engaging discussion of the latter's views see Sean Kelly (forthcoming), "Seeing things in Merleau-Ponty".

Imaginative Mind Objection

Although one can make a case that to be a successful perceiver it *helps* to be able to move around objects, why must we say that sensorimotor knowledge *necessarily* figures in perception in objective space? Why not just say that we need a concept of object and a concept of objects as solid and temporally located three-dimensional space-occupiers in order to have perceptions of objects in objective space. Surely, we do not need to be able to actually move around objects in order to gain a sense of their three-dimensionality and recognize that they appear differently from different points of view. Indeed it would seem that all we need is an imaginative mind that can entertain the possibility of having different points of view. I will call this the imaginative mind objection.

The problem with this line of thought is that it requires at the same time too much and too little. It requires too much since a perceiver does not need the *concept* of an object to have perceptions of objects. No doubt an imaginative mind can do much of the work that sensorimotor knowledge can do, however, it seems important to take seriously that perception is a fairly primitive cognitive skill. A perceiver needs merely the *ability* to perceive objects as three-dimensional space-occupiers.

The imaginative mind objection requires too little since imagining an object and perceiving an object with practical knowledge are two very different activities. By replacing the practical knowledge that is in play by an imaginative mind the presentness and particularity of perception gets lost. I have argued that there is a sense in which I perceive the sides of an object that are not immediately presented to me from the point of view I happen to have in a particular situation of perception. The practical knowledge that figures in my perception is part of my perception proper. This is a very different idea than the idea involved in, say, having a *concept* of objects as solid and three-dimensional. The difference is of the same nature as the difference I have made with reference to Husserl between protention and expectation. A perceiver does not direct her attention at the back side of the object, but there is a sense in which she perceives an object as having a back side qua perceiving it as a three-dimensional space-occupier. What I have said so far is not sufficient to put the imaginative mind objection to rest. I will come back to this objection in Part 3.

Sentient Statue Objection

Assuming that perception involves sensorimotor knowledge as described, is the requirement merely that our perceptions be integrated into sensorimotor patterns allowing us to anticipate how our perceptions would change were our spatial relations to the perceived objects to change? If this were the case, it would only be necessary that a person's body can be moved in relation to perceived objects. Or must the movement at least at times be self-movement? The answer to these questions depends largely on how one understands the sensorimotor knowledge in play. Noë takes a radical position, arguing that bodily movement-perception *coordination* must be gained in order for

perceptual experience to acquire content.[8] If one says that coordination of action and perception is necessary to gain the sensorimotor knowledge involved in perception then the ability to self-activate movement becomes necessary for perception. Intentional movement or deliberate action play an ineliminable role on Noë's view, since it is only through such self-activation that one can figure out the sensorimotor interdependence.

I do not want to go as far. Although I do not take it to be necessary to be able to gain movement-perception *coordination*, I will argue in the next section, however, that it is necessary that a perceiver have practical awareness of being the perceiving subject, and that a perceiver is aware *that* she is perceiving in virtue of controlling *what* she perceives through action. From what has been said so far it is not obvious why it would not be enough that a perceiver be moved in relation to objects. But more needs to be in play to bring out what the ability to form sensorimotor knowledge amounts to in order to understand what is involved in having alter-ego points of view and conjoining these different possible egocentric points of view to thereby gain an understanding of objects as three-dimensional space-occupiers. So in order to understand why this sentient statue objection does not hold you will have to bear with me some more.[9]

Sense Data Objection

Finally, as a last objection to the view outlined so far: why is taking the way objects appear to a perceiver from her point of view into account not just a way of introducing sense data? I am arguing that there is a way in which perception presents the world as being independently of a perceiver's vantage point. In this respect, I perceive the plate as round. But there is also a way the world is presented in perception that incorporates a reference to how things appear from a perceiver's vantage point. In this respect, the plate appears elliptical to me. Such appearances, however, are not mental items. As I argued above, how things appear with respect to shape is a fact about the

---

8   See Noë (forthcoming), p. 24.

9   A further objection against the sensorimotor knowledge view is that a necessary condition for objective perception is for the perceiver to be in a position to regard diverse perceptions as perceptions of a single enduring and distinct object. In order to have perceptions of a single enduring and distinct object a perceiver must be able to recognize two distinct perceptions as successive perceptions of the same object and distinguish this case from cases in which successive perceptions are of two different objects. As was argued above, sensorimotor knowledge allows us to recognize any particular appearance of an object as only one of many possible ways that an object can present itself to a perceiver. But this thought is not the same as the thought that diverse perceptions are recognized as perceptions of a single enduring object.

Sensorimotor knowledge as the idea has been unraveled so far cannot account for the capacity to distinguish between successively perceiving one and the same object and successively perceiving qualitatively indistinguishable but numerically distinct objects, because in order to regard successive perceptions as perceptions of the same object, a perceiver must be able to ascribe them to a numerically identical subject whose route through the world anchors them to a single object. For a helpful discussion of related ideas see Quassim Cassam (1997), *Self and World* (Oxford: Oxford University Press).

object and the vantage point that the perceiver happens to have on the perceived object. Therefore, the appearances that I am arguing are crucial for an understanding of visual spatial perception have nothing to do with sense data. The thought that I think must be taken very seriously in an account of perception is that perception does not just represent the objective spatial properties of objects; it represents how things are *in relation* to the perceiver.

But taking this thought seriously must not be understood as introducing sense data. Accepting that there is a mutual dependency between apparent shapes and actual shapes need not involve understanding perception as a process of constructing an internal representation. The appearance of an object from a perceiver's point of view is difficult to isolate from how we perceive the object on the basis of our sensorimotor knowledge. It is only when we, for instance, learn to draw realistically that we train our eyes to see that a plate appears elliptical from most points of view.

When we take a realistic painter's point of view, we perceive the round object as elliptical. We perceive the object as elliptical *because* we are confronted with a round object, rather than perceiving a round object *because* we encounter an elliptical appearance. In both cases, there is an interdependence between apparent shape and actual shape. Whereas in the first case one abstracts from one's vantage point and perceives the shape of the object independently of one's point of view, in the second case one brings one's vantage point into the picture and perceives the shape of the object as it appears to one from one's point of view.

To say that there is an interdependence between apparent shape and actual shape does not mean that one is necessarily consciously aware of both the apparent shape and the actual shape. And to say that the actual shape and the apparent shape both play a role in perceiving the shape of an object does not involve saying that there is an inconsistency in the perceptual content, since we do not perceive the round object both as elliptical and as round, but rather as either *appearing* elliptical from our point of view or as *being* round.[10]

III – Self-Awareness Argument

I have argued that perception involves sensorimotor knowledge. I aim to show now that the capacity to form sensorimotor knowledge is dependent on a more fundamental connection between action and perception, namely that a

---

[10] I take the interdependency between apparent shape and actual shape to be an instance of the interdependence between how an object appears to one in perception and how the object actually is; an interdependence that is involved in modes of perception other than visual spatial perception. A difficult case is the perception of color, since the color of a surface is not independent of its appearance in the way the actual shape of an object is. Unlike shapes, one might argue, colors are themselves appearances. When perceiving a wall it might be just as correct to say that the wall appears uniformly white, as it is correct to say that the wall appears to have patches of pink and green. But here again there is an interdependency, in this case, between two ways that the wall appears to one:

perceiver be *aware* that she is the acting perceiving subject. Only through such awareness can we make full sense of what it means to understand ourselves as one object among other objects (egocentricity argument) and as keeping track of our perceptions as our spatial relations to perceived objects change (sensorimotor knowledge argument). Uncovering this dependency on self-awareness will put the objections that can be raised against the two previous arguments to rest and allow us to keep hold of what is attractive about the two arguments. Furthermore, it will bring out how the egocentricity argument and the sensorimotor knowledge argument are connected. Keeping track of how our perception changes as our spatial relation to perceived objects changes requires awareness of one's own position in space in so far as this position is the point of origin of our actions and perceptions.

If we take seriously the idea that how things look from here is a relational property that figures in the content of perception, then the perceiver's vantage point must figure in the content of perception. Encountering the elliptical appearance of the plate is what allows us to perceive the plate as round, but only because the vantage point from which we perceive the plate enters in our perceptual content. So the subject of perception plays a role in the content of her perceptions to the extent that she forms the point of origin of an egocentric frame of reference. Furthermore, to play a role in perceptual content, the perceiver's spatial location must gain a place on a cognitive level. (These are two of the ideas that I laid out in the first part and am taking for granted in the context of this paper.)

The location of the perceiving subject as the egocentric frame of reference, I will argue, can figure in our perception only in so far as we are aware of ourselves as acting beings. The idea I am trying to exploit is that perceivers have an understanding of their location in space, because it is the location from which they both perceive and act. Perception is essentially perspectival insofar as perceptual content is structured in subject-dependent terms. But the possibility for action that is involved in the egocentrical organization of perceptual content allows us to go beyond the perspectival representation of objects and to perceive them in objective space. So paradoxically, it is the egocentricity of perception that allows us to transcend our perspectival frame of reference.

It is important that both the capacity to perceive and the capacity to act in play are understood on a personal level. In this respect, the view I am suggesting differs from an argument that takes subpersonal interrelations between perception and action into account. Susan Hurley takes such an approach in her *Consciousness in Action*, arguing that the interdependence between the contents of intentions and of perception on a personal level can be understood as emerging from the codependence of perception and action on dynamically circular subpersonal relations. In particular, she argues that feedback from motor outputs to sensory inputs plays a critical role within such a subpersonal dynamic system. The point I am making is fundamentally different in that the focus of the interdependency between action and perception is on the perceiving subject, not its subpersonal system. But saying that there is an interdependence on a personal level need not involve denying that there is an interdependence on a subpersonal level; I am simply not taking a stance on it.

It must be added that Hurley and similarly Noë in his forthcoming book *Action in Perception* self-consciously slip back and forth between personal and subpersonal levels. Noë argues that bodily activity and the physical implementation of perception in the brain and nervous system are epistemologically on the same level of investigation and, thus, rejects any autonomy thesis, any claim that a philosophical analysis of perception is epistemologically independent of a scientific analysis. I believe that one must hold on to a distinction between bodily movement and the processes of an organism's nervous system that is not just a distinction in observability of the movement involved. But this is a big topic. In this paper, I will take for granted that there is a clean distinction between perception on a personal level and the processing at the concrete implementation of perception.

Now if the capacities to perceive and act are understood on a personal level, why does this account of self-awareness involved in perception not introduce a superfluous and potentially problematic intermediary stage in perception? This is a version of the sense data objection that I argued against in the previous section. It is important to keep in mind that when perceiving external objects the self is not the object of attention. Perceptual attention is focused on the objects perceived, not on the perceiver or the mental state of perception. So the self-awareness in play must not be understood in analogy to the awareness involved in perception of objects.

In this sense, the view presented here differs fundamentally from what could be called the experiential knowledge view. Awareness of what it *feels* like to be perceiving or any other form of introspection or awareness of one's inside is not what constitutes the essential awareness of oneself as the point of origin of perception. I neither feel *that* I am perceiving, nor is there something that is perceiving that I then realize is me. One's perceptions cannot be experienced as unowned or of uncertain ownership.[11] Indeed, the question whether awareness and felt ownership of perception are separate issues does not make much sense. Although one can imagine cases in which the thought that "I found myself perceiving x" makes sense, the situation in which one is aware that one is perceiving but unwilling to ascribe one's perception to oneself does not make sense. By understanding the self-awareness in play as immediate and non-relational in this way, I am rejecting the view that awareness of an object and awareness of myself as perceiving are two states of mind that in perception successfully come together.

How do these ideas of the ownership of perception relate to the thesis that the capacity to perceive is dependent on the capacity to act? I am arguing that a perceiver is aware *that* she is perceiving because she is aware of occupying one space from which she both perceives and acts. But why would perception alone or action alone not be enough to bring about the awareness of one's own location in space that is said to be necessary for perception of objects in objective space? Why does one need the two?

---

[11] There are interesting analogies between the sense of ownership involved in perception and the sense of ownership involved in action. For a discussion of cases in which sense of action and sense ownership can come apart see Anthony Marcel and Christopher Peacocke's (2003) discussion of the Anarchic Hand syndrome in Naomi Eilan, Johannes Roessler (eds.), *Agency and Self-awareness* (Oxford: Oxford University Press), pp. 48-110.

The answer to this question leads to the idea that spatial concepts are not simply related to the place we occupy, but rather to the specific way we occupy that space concerning our possibilities for action. We need at least an understanding of what it would mean to reach out to the glass to perceive the glass as within reach and we need to have an understanding of what it would mean to move our body upwards in order to have an understanding of the spatial concept of up.[12]

The idea behind the thought I am interested in is that a perceiver is aware *that* she is perceiving in virtue of controlling *what* she perceives through action.[13] When we perceive a cube we know that if we were to move thus and so, we would see the same cube from another angle. We are aware that we are perceiving rather than imagining an object because when we, say close our eyes, we cease to see the object and when we make a step to the left our perception changes in expected ways. Action alone or perception alone is not sufficient to gain the self-awareness that is said to be necessary for perception.[14]

Overintellectualization Objection

Now why does understanding the awareness that is said to be involved in perception along these lines not run the risk of over-intellectualizing perception? In reply to this question it needs to be insisted that the notion of self-awareness in play is understood in a radically practical way. The point of origin figuring in perception need not be conceptualized in any conceivable manner and the self-awareness in play need not involve *information* about the self, its states or their contents. A perceiver need not have the capacity to have first-person thoughts, nor does she need the capacity to ascribe

---

[12]   The idea I am interested in is related to Evans's thought that "an egocentric space can exist only for an animal in which a complex network of connections exists between perceptual input and behavioural output" (See his (1982), *Varieties of Reference* (Oxford: Oxford University Press), p. 154). Spatial concepts are not simply related to the place we occupy, but rather to the specific way we occupy that space concerning our possibilities for action. If I am tilting my head, I do not see objects on the verge of sliding off the surface of the earth. The reference of 'up' is not determined by the direction of my head, but rather by how I would move my body, given what I perceive. It is because our perspectival perceptions involve how we *would* move and act that perceptual content gains objectivity through its egocentrical structure. In this sense, the self enters the content of perception as the point of origin of an egocentric frame of reference only in so far as we understand ourselves as acting beings.

[13]   This way of thinking about the self-awareness in play is similar to the basic insight of the ecological understanding of agency according to which perceptual information involves self-awareness in virtue of its role in controlling actions. For a critical discussion of the ecological sense of self-awareness along Gibsonian lines, see Johannes Roessler (2003), "Intentional Action and Self-Awareness" in Eilan, Roessler, pp. 383-406.

[14]   This thought is related to a reading of G.E.M Anscombe (1957) according to which there are important connections between action control and knowledge of actions "from within". Anscombe can be read as arguing that an agent is aware of what she is doing in virtue of controlling her action, rather than on the basis of observation or introspection. The notion of control, involved in perception is naturally of a very different kind. See her *Intention* (Cambridge MA: Harvard University Press).

perceptions to herself. Furthermore, the suggested view is not that a perceiver locates her own position in space by *reflecting* on her spatial relation to perceived objects. As I argued above the self-awareness in play is understood as immediate and non-relational. A perceiver is aware of her location in a dual mode: the zero-point of perception presents itself as being the zero-point for bodily movement. One occupies *one* position from which one *both* perceives and acts.

When a cat perceives a chair, it might not see it *as* a chair, but it sees something that is located in a certain relation to itself and something that it can either choose or choose not to jump onto. Through perception it gauges the distance it must jump to land on the chair. Its location in relation to the chair must figure in its perception for it to be able to flex its muscles such that it ends up landing on the chair.[15]

Imaginative Mind Objection and Sentential Statue Objection revisited

Against the sensorimotor knowledge view the objection was raised that a perceiver must only be able to *imagine* having different points of view in order to acquire the sensorimotor knowledge necessary for objective perception. If it is not necessary that a perceiver actually *has* different points of view, it is unclear why a perceiver need to move in space or think of herself as being capable of moving in space to acquire sensorimotor knowledge. I called this the imaginative mind objection.

An objection could be raised on similar grounds against the self-awareness argument. I have argued that a perceiver needs the *capacity* for self-movement, but it is not clear why a perceiver must ever actualize this capacity. Furthermore, why not say we need perception, sensation, and cognitive abilities that allow us to have self-awareness and imagine occupying different locations in space. Why do we need the capacity to act at all? Finally, surely the connections between perception and action are not so tight as to exclude the possibility that a being that is not capable of self-movement can perceive objects as in objective space. So, why does bringing self-awareness into play refute the imaginative mind objection?

In order to discuss this question, it will be necessary to take a closer look at the notion of action in play. The notion of action in play in the self-awareness argument differs from the notion of action in play in the sensorimotor knowledge argument. The question how to understand the action on which perception is claimed to be dependent must therefore be answered differently with respect to the different ways in which perception is dependent on bodily activity.

As I argued in the last section, it is not obvious that actual token actions need to be involved for perceivers to have sensorimotor knowledge. What constitutes sensorimotor knowledge are not token actions, but rather practical knowledge of how perceptions *would* change, *were* the perceiver to change

---

[15] The example aims only to exemplify that the self-awareness in play is understood practically in a way that is unproblematic to ascribe to cats. The example is not meant to make any implications about cat-perception. It cannot be taken for granted that cats have perceptions of objects as three-dimensional space-occupiers.

her visual angle on the perceived object. By contrast, the self-awareness argument was formulated in terms of *capacities* to act. But again, it is not clear that actual actions need to be involved. Certainly a subject who is *temporarily* unable to act can have a full-fledged conception of her location in space from which she perceives and from which she *would* act *were* she able to act.

On the view suggested a perceiver must, however, know what it *means* to, say reach out to an object, when perceiving the object as within reach. It is not required that a person *actually* reaches out or *has* reached out in the past. But what does it mean to say that a perceiver knows what it means to act, what it means to reach out to something? Certainly it does not mean knowing what it *feels* like to reach out to something or any other form of proprioception. What is meant is rather knowing what it *takes* to reach out to something or knowing what the success conditions are to reach out to something. Since we must know what it would mean to act on our perceptions, being just moved in relation to objects, as say plankton are, is not sufficient to have perception of objects in objective space. This is why the sentient statue objection does not hold. But the alternative is not that the movement be self-activated. Rather I am taking a different line altogether, one that involves minimal cognitive skills, namely knowing what it would mean to act.

But why does this not mean that it would be enough to say that a perceiver need to be able to *imagine* herself to be able to act? The answer to this question depends on what one means by imagining oneself to be able to act. If knowing what it takes to do this or that counts as imagining oneself as being able to act, then indeed that is all that is required. But this is not how the imaginative mind objection was formulated.

The objection was that having a concept of object and a concept of objects as solid and temporally located three-dimensional space-occupiers is sufficient to have perceptions of objects in objective space. Having such concepts is not sufficient to know what it would mean to do this or that, say, reach out to something that one perceives as within reach. In this sense, as I argued in the last section, the imaginative mind objection requires at the same time too much and too little.

IV – Conclusion

In Part 1, I laid out what I am taking for granted, namely that perception is relational, that there is such a thing as perception of objects as three-dimensional space-occupiers, and that in order to perceive objects as three-dimensional space-occupiers a perceiver must have an understanding of objective space. To have such an understanding a perceiver must be able to abstract from her own point of view and understand herself as one objects among others.

The fact that objects are always perceived from a particular point of view does not challenge the objectivity of our perceptions. We are able to perceive the shape of objects independently of the point of view we happen to have in any particular situation. In Part 2, I argued that our perception of an object is determined by practical knowledge of how our perception changes as our spatial relation to the perceived object changes. This sensorimotor knowledge

brings out the interdependence between the apparent shape and the actual shape of the object. While the *actual shape* of an object is determined by how the object fills out space, its *apparent shape* is determined by the actual shape of the object and the perceiver's spatial relation to the object. The apparent shape, what I have called the appearance of the object, is not a mental item, but rather a fact about the world.

The egocentricity argument as well as the sensorimotor knowledge argument is subject to a host of objections. In Part 3, I argued that if one uncovers a more fundamental connection between action and perception that these two arguments depend on, the objections can be put to rest and what is attractive about both arguments can be retained.

Both arguments require that a perceiver be *aware* that she is the acting perceiving subject. Keeping track of how our perception changes as our spatial relation to perceived objects changes requires awareness of one's own position in space in so far as this position is the point of origin of our actions and perceptions. If we take seriously the idea that how things look from here is a relational property, then we must take seriously the idea that the subject of perception plays a role in the content of her perceptions to the extent that she forms the point of origin of an egocentric frame of reference: perceptual content is organized egocentrically representing the perceived object as being in a certain spatial relation to the perceiver. Encountering the elliptical appearance of the plate is what allows us to perceive the plate as round, but only because the vantage point from which we perceive the plate enters into the content of our perception.

When I say that the self enters the content of perception as the point of origin of an egocentric frame of reference, I do not mean that a perceiver being *aware* of herself as an acting perceiver is what figures in the content of perception. It is rather the spatial location of the perceiver in relation to the perceived object that figures in perceptual content. But this spatial location can *only* figure in perceptual content because the perceiver is aware of herself as the acting perceiving subject. Perceivers understand themselves as occupying one space from which they both perceive and act and this self-understanding is crucial for perception in objective space, because it brings about the notion of self that figures in the content of perception and allows us to transcend the egocentricity of our perception.

The Rationality of Deductive Inference

Declan Smithies

NYU/Warwick

dgs224@nyu.edu

1.   The Question

What is required for forming a belief on the basis of deductive inference in a way that is rational?

2.   Fodor's Language of Thought Model

I want to approach this question by considering Fodor's claim that the language of thought hypothesis explains the rationality of deductive inference.[1] The explanation takes the following shape: (1) beliefs (and other propositional attitudes) are identified with physical structures that are causally efficacious; (2) deductive inference consists in the occurrence of causal transitions between these physical structures; (3) the causal transitions between them are sensitive only to their formal (i.e. non-semantic) properties; (4) there is an isomorphism between the formal properties of these physical structures and the semantic properties of their propositional contents; (5) the causal transitions between these physical structures are sensitive to their formal properties in such a way as to preserve relations of truth-functional consequence between their propositional contents.[2] Here is Fodor's summary of how the explanation is supposed to go:

> We can therefore build machines, that have…the following property:
>
>> The operations of the machine consist entirely of transformations of symbols;
>>
>> In the course of performing these operations, the machine is sensitive solely to the syntactic properties of the symbols;
>>
>> And the operations that the machine performs on the symbols are entirely confined to altering their shapes
>
> Yet the machine is so devised that it will transform one symbol into another if and only if the propositions expressed by the symbols that are so transformed stand in certain semantic relations – e.g. the relation that

---

[1]   There is an issue here about whether the language of thought (LOT) hypothesis, as opposed to what Fodor calls the representational theory of mind (RTM), is required to explain the rationality of deductive inference. In the appendix to his 1987, he argues for LOT on the basis of the productivity and systematicity of thought, but in Fodor and Pylyshyn 1991, this is supplemented by a further argument for LOT on the basis of the systematicity of inference.

[2]   Note on Fodor's official position of "meretricious metaphysical neutrality". So perhaps this explanation should not be ascribed to Fodor himself, but rather to "the Fodorian".

the premises bear to the conclusion in a valid argument. (Fodor 1987, p.19)

Intuitively, however, these conditions are not sufficient to explain the rationality of deductive inference, since a mere machine could satisfy them, but a mere machine would not have the capacity for forming beliefs on the basis of deductive inference in a way that is rational.[3] The question I want to consider, then, is what it is, if anything, that Fodor's account leaves out.[4]

3. Searle's Chinese Room

This question is closely related to the one that Searle considers in his famous article, "Minds, Brains and Programs". Searle's question is whether purely computational conditions (of the kind that Fodor articulates) are sufficient for genuine *intelligence* (which can be glossed as rational thought). The conclusion that he reaches on the basis of the Chinese Room thought-experiment is that purely computational conditions are not sufficient, because the man in the Chinese Room satisfies all the relevant computational conditions, but fails to understand the meanings of the Chinese symbols he is processing, and so the processing is not genuinely intelligent.

Searle's diagnosis is that what is lacking in this case of the Chinese Room, as well as in many other cases of so-called "artificial intelligence", is the presence of genuine intentional content:

> Because programs are defined purely formally or syntactically, and because minds have an intrinsic mental content, it follows immediately that the program by itself cannot constitute the mind.

In later work (cf. Searle 1992), he goes on to fill out this conception by arguing that genuine (as opposed to merely "as if") intentionality requires consciousness and that the satisfaction of purely computational conditions does not suffice for consciousness. Thus, the satisfaction of purely computational conditions does not suffice for genuine intentionality, and since genuine intentionality is required for genuine intelligence, the satisfaction of purely computational conditions does not suffice for genuine intelligence.

A number of comments are relevant here. First, it is not at all clear that all the relevant computational conditions have in fact been met in Searle's case. For instance, he does not consider the point that genuine intelligence must be structured and systematic in the right kind of way.[5] As he describes the case, the competence of the Chinese Room is due to its occupant's use of a mere look-up table. Suppose, then, that we adjust the case so that various states of the system make a systematic causal contribution to the behavioural responses exhibited by the Chinese Room. Now we can challenge Searle's contention that intentional states must be available to phenomenal

---

[3] Note also that the explanation is proposed by Fodor as an empirical, rather than a constitutive, one.

[4] Fodor's account is reliabilist in the sense that he takes it that an explanation of the rationality of deductive inference is merely required to explain its distinctive kind of reliability.

[5] Compare Block, "Psychologism and Behaviourism".

consciousness.[6] It can be argued that intentional contents need to be ascribed to the states of the system, regardless of whether or not they are available to consciousness, in order to explain the relational properties of its behavioural responses.[7] On this line, the ascription of intentional content is not a purely instrumentalist matter, but is answerable to perfectly objective constraints on explanation.

It seems to me that Searle has misdiagnosed the intuitive significance of his own case. What is intuitively lacking in the Chinese Room is not intentionality *per se*, but more specifically *understanding* of the Chinese symbols being processed. The claim that understanding requires consciousness (whether we consider linguistic understanding or merely nonlinguistic conceptual understanding) is much more plausible than Searle's own claim that intentionality requires consciousness. After all, there is a compelling intuition that a mere machine without any conscious states could not be credited with any kind of linguistic or conceptual understanding, even if it could be attributed intentional states of some other kind. But some kind of linguistic or conceptual understanding is surely a prerequisite for forming a belief rationally on the basis of deductive inference. Since the Fodorian account of the rationality of deductive inference is silent on the issue of consciousness, this suggests the beginning of an answer to the question of what that account leaves out. However, it is no more than a beginning: the claim that understanding requires consciousness needs both a sharper formulation and a principled explanation.

4.   Brewer's Non-Reflective Access Requirement

In the following passage, Brewer makes an intriguing suggestion about *why* consciousness might be a requirement for rational deductive inference:

> Any purely mechanical dispositional account of the matter is unacceptable. To start to see why, consider what is involved in following a valid deductive argument with real understanding, in a way that yields knowledge of its conclusion…. In cases where I do not go wrong, I am correctly compelled by the argument to believe its conclusion. In following and fully understanding the argument, this compulsion is not simply a blind and mysterious manipulation of my beliefs by some reliable mechanism, however well established by evolution, benevolent hypnosis, or whatever. I am not just a machine which runs along those rails. For if my following the argument is really to extend my knowledge, then my understanding of it must give me some appreciation of why I am right in believing its conclusion. I have to have some grip on how I thereby know the conclusion, and my belief should be guided by this understanding. A disposition to take beliefs on board in parallel with the steps of the argument is on its own insufficient for the argument to provide me with genuine knowledge. For such beliefs would come as a succession of mere hunches, wholly unsubstantiated *for me* by the de

---

[6]   Compare Searle 1990.

[7]   Compare Peacocke, "Content, Computation and Externalism".

facto validity of the argument propelling my endorsement of them. (Brewer 1995, p.242)

The suggestion is that forming a belief on the basis of deductive inference in a way that is rational, and thus a potential means for extending one's knowledge, requires *understanding* the argument that one is following. Understanding the argument, in turn, involves an appreciation of how the conclusion follows from the premises, and hence an appreciation of why it would be rational to believe the conclusion, given belief in the premises. And since such an appreciation requires consciousness, it follows that genuinely rational deductive inference requires consciousness.

More specifically, Brewer proposes that for a belief to be rationally formed on the basis of deductive inference, the formation of the belief must actually be *guided* by a conscious appreciation of why it would be rational to form it, which is provided by an understanding of the argument in question. If the belief is not formed on the basis of any such appreciation, then the suggestion is that it will seem, from the subject's own perspective, to be no different from blind guesswork or an ungrounded hunch. This, in turn, prompts an intuition that it can be nothing more than a matter of mere luck, from the subject's perspective, if the belief is true. But then, of course, the belief cannot be rational. This point emerges in the following passage:

> More generally, the problem for any purely mechanical dispositional account is that it is bound to ignore this sense of why one is right in exercising the capacity in question as one does, which is crucial if this is to make *cognitive* contact with the truth in the relevant area. It is the fact that the capacity has as its point ascertaining the truth on some matter…which sets up the norms for its correct exercise. Some appreciation of how what one is up to is onto this truth, is sensitive to the resultant norms, is therefore essential if exercising it is to be more than a blind mirroring of the norms, extrinsically, and in this sense only incidentally, in contact with the truth. (1995, p.243)

So, in general, if a belief is to be rational, then it must be formed in a way that is guided by some conscious appreciation of why it would be rational to form it.

It is on the basis of such considerations that, in later work, Brewer imposes the following generalized *access* requirement on rational belief and action:

> If a person's reasons are to be cited as *her reasons* for believing or doing what she does, then she necessarily recognizes them as such. (1999, p.166)

Moreover:

> …she is guided in making the transition by her recognition of her reason *as a reason for doing so*. (1999, p.165)

The difficulty here is to see how to avoid a familiar kind of regress problem. In fact, there are two kinds of regress problem to be avoided. Suppose that the way in which the subject recognizes her reason-giving states as such is by forming second-order beliefs about them to the effect that they give her reasons for belief or action. Now, these second-order beliefs must themselves

be formed in a way that is rational. But then the reason-giving states on which these second-order beliefs are based must themselves be recognized as such, which requires the subject to form third-order beliefs about these states to the effect that they give her reasons. And these third-order beliefs, too, must be rational. So, we are embarked on an infinite regress. Perhaps the regress is not vicious, but it does have the incredibly implausible consequence that the rationality of a single belief depends on an infinite hierarchy of higher-order rational beliefs.

However, we get a regress which is more clearly vicious if the only plausible account of the rationality of the second-order belief makes it dependent on the rationality of beliefs lower down the hierarchy. But certainly in the logical case, this can seem to be the only plausible account. Recognizing the rationality of a belief based on modus ponens presumably requires recognizing that modus ponens is a valid rule of inference. And we are supposing that this recognition is a matter of rationally formed higher-order belief. But the rationality of a belief that modus ponens is valid must surely depend on making a modus ponens step, as it does in the following piece of reasoning:

> An argument is valid just in case, necessarily, if its premises are true, then its conclusion is true. Now, suppose that *p* is true and that *if p, then q* is true. Well then *q* must be true, as a matter of necessity, whatever *p* and *q* stand for. So, an argument of the form – *p; if p, then q; therefore, q* – is valid.

But the rationality of making such a modus ponens step is precisely what we are trying to explain; so any purported explanation seems guaranteed to presuppose what it is trying to explain.

Brewer's strategy for avoiding such regress problems is to propose a distinctive account of the way in which the access requirement is satisfied. According to his proposal, the access requirement is not satisfied only at the reflective level by means of higher-order beliefs about one's reason-giving states; rather, it is satisfied at the unreflective level by means of one's reason-giving states themselves. On this view, it is an essential feature of reason-giving states that being in them suffices for satisfaction of the access requirement; no further, reflective thoughts about those reason-giving states are required. Thus, his version of the access requirement is a *non-reflective* access requirement.

Unfortunately, it is not clear how to make sense of the claim that reason-giving states embody a built-in conscious recognition *of their own reason-giving status*. This seems to require a kind of self-referentiality in the contents of reason-giving states which is obscure, at best, and possibly even incoherent.

In fairness, Brewer's (1999, Ch.6) account of the rationality of perceptual-demonstrative judgements suggests that he has something less demanding in mind. According to this account, perceptual experiences provide reasons for perceptual judgements about the external world because they embody an appreciation on the part of the subject that the experience is

causally dependent on the state of the external world in relevant respects.[8] It is in virtue of embodying such an appreciation that perceptual experiences provide reasons for perceptual judgements. It may be objected that this does not suffice, strictly speaking, for the claim that perceptual experiences embody an appreciation of the fact that *they themselves provide reasons* for perceptual judgements. But it may equally be replied that, loosely speaking, it suffices for the claim that perceptual experiences embody a sense of "why it would be right" to make perceptual judgements on their basis.

The analogous claim in the case of deductive inference would be that a thinker's understanding of the premises and the conclusion of an argument embody a conscious appreciation of the fact that the conclusion follows from the premises. Loosely speaking, this can be described as a sense of "why it would be right" to believe the conclusion of the argument on the basis of believing its premises, even in the absence of an explicit recognition of the fact that my belief in the premises provides me with a reason to believe the conclusion. The proposal, then, would be that the rationality of deductive inference depends on the fact that it is made on the basis of a conscious appreciation of the fact that the conclusion of the argument follows from its premises. This proposal has recently been endorsed by Christopher Peacocke, who expresses it as follows:

How can we elucidate the rationality of the thinker's judgements? One intuitive account is that in making a rational transition to a judgement that p [on the basis of deductive inference] a thinker must know what it is for it to be true that p, must appreciate that his grounds or reasons for the transition to the conclusion that p suffice for the truth of p, and must be making the judgement because of his appreciation that these grounds or reasons so suffice. (2003, p.176)

However, neither Peacocke nor Brewer says much about how the relevant kind of conscious appreciation might be constituted so as to play this role in grounding the rationality of deductive inference. But this issue turns out to be crucial in the evaluation of the proposal.

5.   Boghossian's Challenge

Suppose that appreciating the validity of a form of argument is a matter of explicitly articulating the form of the argument and rationally judging it to be valid. If this is what it is involved, then there are compelling objections against the claim that a deductive inference is rational only if the subject makes the inference on the basis of some such appreciation of its logical validity. First of all, the requisite conceptual capacities are lacked by most normal adults, not to mention animals and children; but it would be intuitively implausible to deny on this basis that they have any capacity for rational deductive inference. The

---

[8]   Note that this appreciation is not a matter of employing an explicit theory of the enabling conditions for perception, on Brewer's account, but is rather a matter of (1) being able to keep track of objects over time as they change in respect of the properties (such as appearance, or spatial location) in virtue of which the subject grasps their identity; and (2) being able to make sense of informative identities involving such objects in response to experientially discontinuous demonstrative identifications of them. (See especially 1999, p.197, fn.11 and p.210.)

second objection is that the rationality of judgements about validity itself depends on rational deductive inference, as we saw above. So, the rationality of deductive inference cannot depend on an appreciation of the validity of a form of argument; rather, the direction of dependence is the reverse.[9]

One line of response to these objections would be to claim that we have some kind of direct and non-inferential, quasi-perceptual access to the validity of forms of argument. The claim would be that such a quasi-perceptual faculty of "rational intuition" is possible in the absence of a sophisticated capacity for articulating general forms of argument and using the resources of deductive inference to evaluate their validity. Moreover, it could be claimed (by analogy with the perceptual case) that judgements about validity can be non-inferentially justified on the basis of the deliverances of such a faculty, without requiring any additional inferential backing.

However, there are compelling objections against assimilating our knowledge of logical validity too closely to the perceptual paradigm. The first is that perception of an object or an instantiated property requires the holding of a causal relation between the perceiving subject and the object or instantiated property. Valid forms of argument, however, are abstract objects, and as such, are incapable of entering into causal relations. So, there is no coherent notion of perceiving the validity of a form of argument. The second objection is that perception alone can only yield knowledge of contingent truths.[10] But knowledge of logical validity is knowledge of necessity. Therefore perception alone is incapable of explaining this modal dimension of our logical knowledge. The upshot is that the notion of *genuinely* perceptual access to the validity of forms of argument is incoherent, while the notion of *quasi*-perceptual access is merely obscure.

In addition, Paul Boghossian (2001, 2003) has argued that the appeal to rational intuition cannot do the work required of it. Suppose we grant that deductive inference is not required for forming a rational judgement to the effect that a certain form of argument is valid; still, it is presumably required in order to bring any such judgement to bear on the rationality of a particular inference. So, if we judge on the basis of rational intuition that any inference of the modus ponens (MPP) form is valid, then we still need to make the following inference:

(i)   Any inference of the form MPP is valid.

(ii)   This particular inference, from (1) and (2) to (3) is of MPP form.

Therefore,

(iii)   This particular inference from (1) and (2) to (3) is valid. (2003, p.233)

A similar objection applies if we grant that what rational intuition provides is not direct access to the validity of a general form of argument, but rather to the validity of a particular instance of that form. The thought is that inference is still required in order to bring the judgement to bear on the rationality of making a deductive inference in the particular case, along the following lines:

---

[9]   See Boghossian (2001), (2003) for a statement of these objections.

[10]   Ref. Peacocke (2003) on a posteriori knowledge of necessity.

(iv)   This particular inference from (1) and (2) to (3) is valid.

(v)   If an inference is valid, then anyone who is justified in believing its premises and knows of its validity is justified in inferring its conclusion.

Therefore,

(vi)   Anyone who is justified in believing the premises of the argument is justified in inferring its conclusion.

(vii) I am justified in believing the premises (1) and (2).

Therefore,

(viii) I am justified in inferring (3). (2003, p.234)

6.   Dummett on the Perception of Semantic Patterns

Despite the force of Boghossian's objections, it seems to me that rational intuition is both a genuine phenomenon and one to which we can usefully appeal in giving substance to the Brewer/Peacocke account of the rationality of deductive inference. In arguing the point, I will draw on some illuminating remarks made by Michael Dummett on the matter.

In the following passage, Dummett draws a distinction between *understanding* a proof and merely checking that it is correct:

> …an understanding of a proof demands more than an ability to recognize that it is correct. To verify that every line of a formal proof follows from earlier lines by one of a list of transformation rules is to be convinced, within the limits of human error, that it is correct; but it takes one very little way towards understanding the proof. The proof has an architecture that must be comprehended as a whole; but the first necessity for gaining such comprehension is to be intuitively convinced, for each step, that it genuinely follows from the earlier lines from which it was derived. On Frege's account, this will in general require a creative act. It is not enough merely to grasp the thought expressed by each line of the proof; in addition, one must perceive patterns common to those thoughts and others, patterns which are not given with the thoughts as a condition for grasping them but which require a further insight to apprehend. (1991, p.197-8)

To understand the proof, it is not enough to understand each line of the proof and to have checked that each line follows from the earlier lines by application of the inference rules. In addition, it is required that one should *perceive certain patterns* which hold between the propositions or thoughts expressed by the various lines of the proof. This "perception of pattern" is not something which is built into the requirements for understanding the various lines of the proof, but requires an additional, creative insight. Thus, on Dummett's account, *following* a proof is as much a creative matter as *discovering* a proof.

Dummett contrasts the case of understanding a proof with the imaginary case of an obedient community under the governance of an International Academy of Logic, a body which lays down decrees concerning which logical laws are, until further notice to be treated as valid or invalid. He observes:

> …we could obey these decrees; but we should lose the sense that we any longer understood what we were saying. The *rules* of the language-game would be clear enough; but its *point* would now escape us. (1991, p.207)

Presumably, what would be lacking in such a community is not merely the capacity for perception of pattern – its members could be perfectly capable of perceiving the *formal* patterns holding between the lines of a proof – but rather the capacity for perception of *semantic* patterns: for example, that the truth of the conclusion of an argument follows from the truth of its premises. The idea seems to be that in the absence of a capacity for representing such semantic relations, the subject cannot have any sense of the *point* of his inferential practices: he will be merely following the rules without understanding *why* he is following them.

It is here that an account of the rationality of deductive inference links up with an account of the thinker's understanding of logical concepts, which consists in a knowledge of their contribution to truth-conditions. After all, it is presumably in virtue of the thinker's knowledge of the contribution made to truth-conditions by logical concepts that he is capable of appreciating the semantic properties of logical forms, such as the validity of a form of argument. Dummett's claim is that what is lacking in his imaginary community is precisely the knowledge of reference in which the understanding of logical concepts consists:

> …the missing component of understanding is not to be stigmatized as a 'mere' feeling…. What he lacks is not the *feeling* of understanding, but the *knowledge* that is an essential component of understanding. It is that knowledge that we should lack if we were compelled to reason in accordance with principles that appeared to us invalid or gratuitously restricted: we could rightly confess that we no longer knew what we were saying. (1991, p.208)

7.   Meeting Boghossian's Challenge

In summary, Dummett's idea is that understanding a proof requires perceiving the semantic patterns holding between the different lines of the proof or, in other words, perceiving the validity of the proof. But how are we to understand talk of "perceiving" in this context? We have already seen compelling objections against literal interpretations of such talk, but perhaps it can be understood metaphorically.[11] It seems to me that the perceptual metaphors suggest themselves because there are clear similarities between discerning the patterns in a proof and a certain kind of perceptual phenomenon. Indeed, it is plausibly one and the same kind of phenomenon which occurs both in the context of perception and in the context of pure thought. This is the phenomenon that Wittgenstein labelled "*seeing as*" in the second part of the *Philosophical Investigations*. Just as one can be struck by certain spatial

---

[11]   Note that we use perceptual vocabulary in a wide variety of settings that do not involve sensory perception, e.g. seeing that a certain course of action is the one to take, seeing that one has made a mistake in a proof, etc.

relations between objects in the perceived array, so one can be struck by certain semantic relations between thoughts in occurrent consciousness. In some cases, this may involve a kind of conceptual structuring of the contents of conscious experience, but the basic phenomenon is pre-conceptual – it is a matter of directing one's attention or having one's attention directed in such a way that the relations in question become salient in consciousness.

This basic attentional phenomenon has a number of features which make the perceptual metaphors particularly apposite. First, it involves a characteristic kind of *passivity*, which is given expression in the idea of being *struck* by the relations in question. Second, and relatedly, being struck by such relations is not something which stands in need of reasons or justification, any more than does the passive reception of perceptual experience itself. Third, the phenomenon is belief-independent. Just as we are subject to cognitively impenetrable perceptual illusions, such as the Muller-Lyer illusion, which may persist in the phenomenology of experience even when we know about the illusion, so we are sometimes subject to illusions of validity which persist in the same way even when we know about them.[12]

These points bring out a sense in which the analogy between rational intuition and perception need not be obscure, but can actually be quite illuminating. They also defuse Boghossian's threat of the reappearance of Carrollian circularity. To the extent that the representations of logical validity that we are considering are *belief-independent*, it cannot be assumed that *inference* must be involved in bringing them to bear on the rationality or justifiability of a particular inference. On the contrary, it can be argued that the subject's conscious attention to the fact that the conclusion of an argument follows from its premises may be directly causally and rationally implicated in the fact that he makes the inferential transition from a belief in the premises to a belief in the conclusion.

So, the notion of rational intuition may have an important role to play in an account of the rationality of making a particular deductive inference. I also think it has an important role to play in an account of propositional knowledge of logical truths, including knowledge of which forms of argument are logically valid. Consider what is involved in the rationality of judging that a particular argument of a certain form is valid. On one view, which we can call the *deductive* model, this requires giving an explicit articulation of the general form of the argument in question and inferring its validity from more general principles concerning what it is for a form of argument to be valid.[13] But there is an alternative view, which we can call the *rational intuition* model, on which judgements of validity in the particular case have epistemological priority and can be rationally based directly on the deliverances of rational intuition. On this view, one's conscious awareness of the fact that the conclusion of an argument follows from its premises, which is explained by one's understanding of the logical concepts that figure in the argument, provides

---

[12] Example

[13] Thus, Peacocke's objection that negative conditions are not plausibly represented phenomenologically falls away as irrelevant.

one with a non-inferential rational basis for judging that the argument is logically valid.

The comparison with the epistemology of moral and epistemic judgements is instructive. If we consider what is involved in the rationality of judging that a particular case exemplifies knowledge, or goodness, then our options are structurally similar. On the deductive model, it requires giving an explicit articulation of the conditions that are necessary and sufficient for the exemplification of knowledge, or goodness, and showing that these conditions are satisfied in the particular case in question.[14] On the rational intuition model, by contrast, judgements about the exemplification of knowledge, or goodness, in particular cases have epistemological priority and can be rationally based directly on the deliverances of rational intuition. On this view, one's conscious awareness of the fact that a particular case exemplifies knowledge, which is explained by one's understanding of the concept of knowledge, provides one with a non-inferential rational basis for judging that the case exemplifies knowledge.

The cases of moral and epistemic judgements bring out particularly clearly the intuitive appeal of the rational intuition model, as compared with the deductive model. But more can be said to *explain* its intuitive appeal. The problem with the explicit inference view is not *just* that it makes the acquisition of moral, epistemic and logical knowledge an extremely difficult and arcane matter; it also makes it viciously circular. In order to know that a particular case exemplifies knowledge, I already need to know a great deal about knowledge – indeed, I need to have explicit knowledge of which conditions are necessary and sufficient for knowledge. But how am I to acquire this knowledge? At this point, it is hard to deny that we acquire knowledge of necessary and sufficient conditions for knowledge on the basis of an attempt to systematize and explain the judgements that we are pretheoretically inclined to make in particular cases, and which we are (defeasibly) entitled to presume are in good epistemic standing. Certainly, this has been the methodology which is characteristic of the mainstream literature in epistemology concerning the conceptual analysis of knowledge. But the explicit inference view makes it viciously circular.

For similar reasons, it is quite implausible to require that I need to have articulated, or even to be capable of articulating, the general form of an argument in order to be rational in judging that a particular argument of that form is valid. Consider the following "Days of the Week" argument:

(1)   Today is Monday.

(2)   If today is Monday, then tomorrow is Tuesday.
        Therefore,

(3)   Tomorrow is Tuesday.

I can just *see* that this argument is valid, and be rational in judging that it is, without being able to articulate the general form of the argument. It is not just

---

[14]   See McDowell, "Virtue and Reason". Whether we need to appeal to tacit knowledge of such principles is another matter. See Peacocke (2003). Question: what role is there, if any, for tacit knowledge in an account of the epistemology of moral, epistemic or logical judgements?

that the capacity to articulate the general form of the argument is somewhat arcane and technical; it also depends on a more fundamental capacity to evaluate particular arguments as valid or invalid. Suppose I form the belief that the Days of the Week argument is valid and go on to try to justify my belief on the basis of the claim that the argument instantiates the following form which I also claim to be logically valid:

(1)   P

(2)   If P, then Q

Therefore,

(3)   Q

Then I am open to apparent counterexamples, in the style of Van McGee, involving instances of this form which have conditionals embedded in the antecedent of the conditional premise. Now, I might try to explain away the intuitive plausibility of these apparent counterexamples and, who knows, I might even be successful, but their initial plausibility cannot be denied. The crucial point is that my judgement that the Days of the Week argument is valid (as opposed to my attempt to give an explicit justification of that judgement) should not depend on the outcome of this debate. Rather, it is plain that the debate itself is a debate about how best to systematize and explain the judgements we are pretheoretically inclined to make in particular cases, which we are defeasibly entitled to presume are in good epistemic standing.

A purely deductive model of our acquisition of logical knowledge is unacceptable; it needs to be recognized that the acquisition of logical knowledge depends both on rational intuition and on a broadly inductive process of inference to the best explanation of judgements based on rational intuition.

8.   On Brewer's Non-Reflective Access Requirement

For these reasons, then, it seems to me that rational intuition is both a genuine phenomenon and one that has an important role to play in the epistemology of logic. However, it is a further question whether it can be invoked in a plausible defence of Brewer's claim that the rationality of deductive inference depends on the fact that the inference is made on the basis of a conscious appreciation of its validity. Brewer's proposal seems to be that a subject's belief in the premises of a deductively valid argument provides him with a reason for believing its conclusion in virtue of embodying an appreciation of the logical validity of the form of argument in question. According to this line, the appreciation of logical validity is actually embodied in the very beliefs which serve to provide the premises of the inference. But this is not Dummett's conception of the matter. According to him:

It is not enough merely to grasp the thought expressed by each line of the proof; in addition, one must perceive patterns common to those thoughts and others, patterns which are not given with the thoughts as a condition for grasping them but which require a further insight to apprehend. (1991, p.198)

Indeed, he takes this point to be crucial in the explanation of how it is that deductive inference can be informative, in the sense of providing a source of new knowledge:

Deductive reasoning is thus in no way a mechanical process, though it may be set out so as to be checkable mechanically: it has a creative component, involving the apprehension of patterns within the thoughts expressed, and relating them to one another, that are not required for or given with a grasp of those thoughts themselves. Since it has this creative component, a knowledge of the premises of an inferential step does not entail a knowledge of the conclusion, even when we attend to them simultaneously; and so deductive knowledge can yield new knowledge. Since the relevant patterns need to be discerned, such reasoning is fruitful; but since they are there to be discerned, its validity is not called into question. (FPM, p.42)

If believing the premises of a valid argument were sufficient to provide you with an appreciation of how the truth of the conclusion follows from the premises, then deductive inference would not be informative. On Dummett's account, by contrast, it requires a creative insight to see how the conclusion of an argument follows from its premises, and so the informativeness of inference is preserved.

So, on Dummett's account, it is possible to understand the premises of an argument without thereby having any conscious appreciation of the fact that the conclusion of the argument follows. We therefore have to consider the case in which a subject believes the premises of an argument and makes the transition in thought to believing its conclusion, but *without* being guided by any conscious appreciation of the fact that the conclusion follows from the premises. According to Brewer, such a transition cannot be rational. But this seems to me to be setting the hurdle too high. Brewer and Dummett make a convincing case that following a proof with a genuine sense of understanding requires an appreciation of its validity. But following a proof is a relatively sophisticated activity. There are much more mundane cases of deductive reasoning in which it is not so plausible to demand an appreciation of validity, or even the capacity for appreciation of validity, as a requirement for rationality.

9. A More Permissive Proposal

I want to make a more permissive proposal according to which a deductive inference is rational, in the most basic cases, just in case the form of the inference is a valid form, and the thinker makes the inference *because* it is of that form.

What is it to make an inference *because* it is of a certain form? My answer is that it is just a matter of exercising a *disposition* to make inferences of that form, though this must be a disposition in the "full-blooded" sense, i.e. one that is grounded at the subpersonal level by some state that is a common cause of the various exercises of the inferential disposition, and hence which embodies tacit knowledge of the form of inference in question.[15] But there is no requirement on representation of the form of inference at the personal level: in particular, there is no requirement that the subject should be guided

---

[15] See Peacocke (1992) Ch.7.1 for a more detailed proposal. Note that tacit knowledge of a form does not require explicit representation of the form: it can be embodied by a processor. See Fodor 1987, Ch.1.

in making the inference by any conscious appreciation of the fact that the form of inference in question is a valid form.

But if this all there is to be said, then what, if anything, is left out on Fodor's account of the rationality of deductive inference? The suggestion we have been considering is that deductive inference requires understanding, which requires consciousness; but Fodor leaves out any mention of consciousness in his account.[16] The basic idea behind this suggestion seems to me to be correct, but it can be deployed in different ways. The Dummett/Brewer/Peacocke line is that making a deductive inference in a way which involves genuine understanding requires a conscious appreciation of the validity of the form of inference. My objection was that, while this requirement may be appropriate in relatively sophisticated cases, such as following a proof, it is inappropriate in more mundane cases of deductive inference. But I think there is a different way to make the connection between inference, understanding and consciousness.

A deductive inference is certain kind of (non-deviant) causal transition between *beliefs*. And in order to have a belief with a certain content, it is necessary to *understand* that content, which requires understanding its component concepts. Now, the crucial question is this: what is required for understanding a logical concept? And the crucial claim is that it is not sufficient to have tacit knowledge of what are in fact valid logical forms. Tacit knowledge of logical forms is just a matter of having a full-blooded disposition to make *transitions between intentional states* whose contents instantiate those forms. There is no requirement for understanding the contents of the states involved in the relevant transitions. (Consider the intentional contents involved in low-level computations in the visual or grammar modules.) Understanding a logical concept, on the other hand, is a matter of having a full-blooded disposition to make *deductive inferences between beliefs* whose contents instantiate certain logical forms. Since we are now in the domain of belief and inference, possession of the relevant disposition *does* require understanding the contents of the beliefs involved in the inference.

But now we seem to be moving in a circle which is too small to be illuminating. How can we get a grip on what is distinctive about inferences between beliefs as compared with merely inference-like transitions between merely belief-like intentional states, without simply appealing to a notion of understanding which is taken as primitive? How can we get some independent grip on the claim that understanding is involved in the one kind of case, but not the other?

The answer I propose is that the notions of belief, inference and understanding are *normatively* individuated: that is to say, they are individuated by their *rational* role. We can make sense of the distinction between causally isomorphic systems of inference-like transitions between belief-like intentional states on the one hand, and inferences between beliefs on the other, in virtue of the fact that the latter exhibit rational relations, whereas the former do not. But we have still not yet reached a satisfactory stopping point. There must be some property of beliefs in virtue of which they

---

[16] Note: tacit knowledge of the logical forms of inferences does not suffice for tacit knowledge of the semantic value of a logical concept.

are capable of standing in rational relations to one another, a property which might be lacked by a causally isomorphic system of belief-like states. That property, I claim, is availability to consciousness.[17]

In summary, understanding a logical concept is a matter of being disposed to make distinctively *rational* inferences between beliefs, where the rationality of these inferences is grounded, in part, by the fact that the beliefs involved are available to consciousness. So, as promised, there is an alternative to the Dummett/Brewer/Peacocke account of the connection between consciousness, understanding and the rationality of deductive inference.

This alternative is more permissive than the Dummett/Brewer/Peacocke account insofar as it denies that a rational deductive inference must be made on the basis of a conscious appreciation of the validity of the form of inference in question. I argued that this requirement might be plausible for relatively sophisticated cases, such as following a proof, but not in more mundane cases of deductive reasoning. But I also think that the permissive alternative provides the materials for explaining how conscious appreciation of validity is possible in more sophisticated cases. According to my proposal, the following are requirements for the rationality of a deductive inference:

(1) The thinker must understand the contents of the beliefs which serve as the premises and conclusion of the inference; in other words, the thinker must know what it would be for them to be true;

(2) The beliefs involved in the inference must be available to consciousness.

The fact that there are conditions under which the beliefs involved in the inference are occurrent in consciousness grounds the possibility that the thinker may become consciously aware of certain relations between their contents. And given his knowledge of what it would be for them to be true, it grounds the possibility that he may become aware of certain *semantic* relations between their contents: in particular, that the truth of the conclusion follows from the truth of the premises. In other words, it grounds the possibility of conscious awareness of the validity of the logical form of the inference.

10. Access Internalism

Brewer's non-reflective access requirement is motivated by the need to accommodate the internalist intuitions that are prompted by standard counterexamples to pure reliabilism without raising the problems faced by traditional forms of access internalism. According to Brewer's diagnosis, these problems stem from the following pair of commitments:

(1) to the thesis that a person's reasons are essentially recognizable by her as such; (2) to the idea that this recognition can only be a matter of her second-order knowledge that the mental state providing the reason in question is appropriately related to that for which it is a reason, where

---

[17] See "Rationality and the Subject's Point of View" for a defence of the claim that reason-giving states must be available to phenomenal consciousness.

this is independent of the first-order state itself in that she might have been in just that state yet not had the second-order knowledge required for its status as a reason for her. (1999, p.164)

The pure reliabilist response is to reject (1) altogether, whereas Brewer's suggestion is that we can endorse (1) while rejecting (2). Effectively, then, his proposal is that we should replace the traditional *reflective* access requirement by means of the *unreflective* access requirement.

However, there is an another way to characterize the elusive middle ground that we should be seeking. According to this suggestion, what is wrong with traditional forms of internalism is not the reflective access requirement *per se*, but rather the traditional interpretations of that requirement. On the traditional interpretations, the possibility of access (i.e. recognition of reasons) is grounded in such a way as to impose substantive constraints on the reflective capacities of the rational subject in question. This interpretation of the requirement is basically taken over by Brewer, except that the constraints are imposed on unreflective, as opposed to reflective, capacities of the rational subject. On the alternative interpretation, by contrast, the possibility of access is grounded in the nature of the reason-giving states themselves. Thus, even if the subject is not capable of recognizing the rationality of her belief, it may be rational all the same, so long as it is formed in such a way as to ground the possibility (in principle) of recognition.

In giving an account of the requirements for rationality, we always have to be sensitive to the fact that there are many different degrees of sophistication in rational thought and action. We should not be so impressed by the more sophisticated kinds of rationality that we end up denying that more mundane kinds qualify as rational at all. But neither should we be so impressed by the more mundane kinds that we rob ourselves of the resources for explaining the more sophisticated kinds. I have tried to give an account that is as permissive as possible, but which provides the resources for explaining how it is that the most mundane kinds of rational thinking make possible the more sophisticated kinds.

**Consciousness and Self-Consciousness
Research Centre**

www.warwick.ac.uk/go/consciousnesscentre