

# Sorites

Σωρίτης

A Digital Journal of Analytical Philosophy  
Indexed and Abstracted in *THE PHILOSOPHER'S INDEX*  
ISSN 1135-1349

Legal Deposit Registration: M 14867-1995

**Editor: Lorenzo Peña**

(CSIC [Spanish Academy of Sciences])  
Jurisprudence-and-Logic Research Group (JuriLog)

**Regular-Mail Address:**  
Prof. Lorenzo Peña  
CSIC-CCHS — Department of Philosophy  
Albasanz 26  
E-28037 Madrid  
Spain

Fax +3491 22 986 91  
Voice Tph +3491 60 223 72

**<<http://www.sorites.org/>>**  
**<<http://www.sorites.info/>>**  
**<[sorites@sorites.org](mailto:sorites@sorites.org)>**

**Issue # 21 — June 2008**  
**[published January 2015]**

# **SORITES (Σωρίτης) ®**

ISSN 1135-1349

Issue # 21 — June 2008

[published January 2015]

Copyright © by SORITES and the authors

---

**<<http://www.sorites.org/>>**

**<<http://www.sorites.info/>>**

**<[sorites@sorites.org](mailto:sorites@sorites.org)>**

# **SORITES (Σωρίτης)**

ISSN 1135-1349

## **ROLL OF REFEREES**

---

Rainer Born . . . . . Johannes-Kepler Universitaet Linz (Austria)  
Amedeo Conte . . . . . University of Pavia (Italy)  
Newton C.A. da Costa . . . . . University of São Paulo (Brazil)  
Marcelo Dascal . . . . . University of Tel Aviv (Israel)  
Dorothy Edgington . . . . . Birbeck College (London, UK)  
Graeme Forbes . . . . . Tulane University (New Orleans, Louisiana, USA)  
Manuel García-Carpintero . . . . . University of Barcelona (Spain)  
Laurence Goldstein . . . . . University of Hong Kong (Hong Kong)  
Jorge Gracia . . . . . State University of New York, Buffalo (USA)  
Nicholas Griffin . . . . . McMaster University (Hamilton, Ontario, Canada)  
Rudolf Haller . . . . . Karl-Franzens-Universitaet Graz (Austria)  
Terence Horgan . . . . . University of Memphis (Tennessee, USA)  
Victoria Iturralde . . . . . Univ. of the Basque Country (San Sebastian, Spain)  
Tomis E. Kapitan . . . . . Northern Illinois University (USA)  
Manuel Liz . . . . . University of La Laguna (Canary Islands, Spain)  
Peter Menzies . . . . . Australian National University (Canberra, Australia)  
Carlos Moya . . . . . University of Valencia (Spain)  
Kevin Mulligan . . . . . University of Geneva (Switzerland)  
Jesús Padilla-Gálvez . . . . . Johannes-Kepler Universitaet Linz (Austria)  
Philip Pettit . . . . . Australian National University (Canberra, Australia)  
Graham Priest . . . . . University of Queensland (Brisbane, Australia)  
Eduardo Rabossi . . . . . University of Buenos Aires (Argentina)  
David-Hillel Ruben . . . . . School of Oriental and African Studies, University of London  
Mark Sainsbury . . . . . King's College (London, UK)  
Daniel Schulthess . . . . . University of Neuchâtel (Switzerland)  
Peter Simons . . . . . University of Leeds (Leeds, UK)  
Ernest Sosa . . . . . Brown University (Providence, Rhode Island, USA)  
Friedrich Stadler . . . . . Institut «Wien Kreis» (Vienna, Austria)

# SORITES (Σωρίτης)

ISSN 1135-1349

Issue # 21 — June 2008

## TABLE OF CONTENTS

---

— Abstracts of the Papers . . . . .	03
— «Evaluating Williamson’s Anti-Scepticism» by Tony Cheng . . . . .	06
— «Velleman on the Constitutive Aims of Practical and Theoretical Reasoning» by Emer O’Hagan . . . . .	12
— «Perseverance, Motivation, and Ambition’s Debt» by John Zillmer . .	31
— «Fictional Colors» by Dimitria Electra Gatzia . . . . .	37
— «Kant on Pragmatism: Kantian Notes to Brandom’s Inferentialism» by Jesús González Fisac . . . . .	52
— «What would you say then? The philosophical appeal to what one would say» by Renia Gasparatou . . . . .	63
— «An Evolutionary Argument Against Epiphenomenalism and Type-Identity Theory» by Brett Anderson . . . . .	71
— The <b>SORITES</b> Charter . . . . .	92
— Release Notice . . . . .	98

SORITES (Σοφίτης), ISSN 1135-1349  
http://www.sorites.org  
Issue # 21 — June 2008. Pp. 03-05  
Abstracts of the Papers  
Copyright © by SORITES and the authors

## ABSTRACTS OF THE PAPERS

---

### EVALUATING WILLIAMSON'S ANTI-SCEPTICISM

**Tony Cheng**

Timothy Williamson's *Knowledge and its Limits* has been highly influential since the beginning of this century. It can be read as a systematic response to scepticism. One of the most important notions in this response is the notion of «evidence,» which will be the focus of the present paper. I attempt to show primarily two things. First, the notion of evidence invoked by Williamson does not address the sceptical worry: he stipulates an objective notion of evidence, but this begs the question against his opponent. Second, his related thesis «Evidence equals Knowledge» does not sit well with his own content externalism: he promises to relate epistemology to philosophy of mind, but he fails to live up to this commitment in his crucial chapter on scepticism. Other minor problems concerning evidence will also be discussed in due course.

---

### VELLEMAN ON THE CONSTITUTIVE AIMS OF PRACTICAL AND THEORETICAL REASONING

**Emer O'Hagan**

Constitutive answers to the question 'how do rational principles bind agents?' hold that rational norms are inherently authoritative over reasoners; agents and reasons for belief and action are mutually constituted. J. David Velleman has argued that certain formulations of the constitutive approach are circular because they unpack the constitutive aims of practical and theoretical reasoning in normative terms. To avoid this purported circularity, Velleman has proposed, what I call, the 'conceptual independence thesis', the view that the constitutive aim of belief must be understood in terms which are conceptually independent of reasons for believing, and the constitutive aim of action must be understood in terms which are conceptually independent of reasons for action. I argue that the conceptual independence thesis is not adequately supported by Velleman's arguments and that it is a source of tension in his position. I argue that the normativity of reasons cannot be explained if the conceptual link between having reasons and the practice of reasoning is cut, and claim that vicious circularity can be avoided without endorsing the conceptual independence thesis. Because the conceptual independence thesis poses problems for understanding how rational norms (theoretical and practical) constrain agents, it should be abandoned. For Velleman good reasons are those which better manifest the aim of belief itself, but if the aim is mere truth, then we cannot

intelligibly unpack the metaphor of nearness to belief's constitutive aim. Indeed, the only way to unpack the metaphor is by way of the normative force of reason.

---

## **PERSEVERANCE, MOTIVATION, AND AMBITION'S DEBT**

**John Zillmer**

Ambition is often considered to be a desirable character trait. Here I analyze the moral psychology underlying ambition, and advance two main claims. One is that while the similar trait of perseverance may be desirable, it is distinct from ambition. The other is that all ambition is of an undesirable sort insofar as ambition does not serve to further an agent's ends, but rather serves to perpetuate itself.

---

## **FICTIONAL COLORS**

**Dimitria Electra Gatzia**

I argue that the rejection of color realism need not seal the fate of our ordinary color discourse. I do not argue for the claim that realism is false (partly because this is beyond the scope of this paper and partly because it is not pertinent to my proposal). I rather propose an alternative to it: an account that allows us to preserve our ordinary color discourse without having to commit to philosophically problematic properties. I then discuss some potential worries for my proposal and offer some plausible responses.

---

## **KANT ON PRAGMATISM: KANTIAN NOTES TO BRANDOM'S INFERENCEALISM**

**Jesús González Fisac**

The aim of this paper is to give kantian answers to Brandom's analyses about inferentialism. Since Brandom has dealt systematically with the kantian philosophy (one of the essential modern references of his theses) we will take the «three kantian dualisms» — just as Brandom has featured them — as the theme of this paper in order to see if Kant's work sustains them, that is, if it sustains what Brandom calls the «dualistic» reading of those three distinctions (a lecture which we can find in a good amount of the interpretations about Kant).

---

## **WHAT WOULD YOU SAY THEN?**

### **THE PHILOSOPHICAL APPEAL TO WHAT ONE WOULD SAY**

**Renia Gasparatou**

In this paper I suggest that many thought experiments, imaginary examples and counterexamples, used widely in analytic philosophy as forms of argument, rely on an old fashioned appeal to what one would say. Appealing to ordinary language pronouncements was back in the 50s suggested as the most intuitive way of coming to see the obviousness of the suggested solution. I suggest that similar rhetoric underlies many analytic philosophers'

argumentation techniques, even among philosophers who do not share the ordinary language philosophers' rationale. In this paper I will try to pick up typical species of this methodology in classic analytic writings (Kripke, Putnam, Jackson etc.) and suggest that ordinary language pronouncements can hardly prescribe answers to extraordinary quests, such as those philosophers pursue. The plea to *what one would ordinarily say* cannot prescribe answers when the context is extraordinary.<sup>1</sup>

---

## **AN EVOLUTIONARY ARGUMENT AGAINST EPIPHENOMENALISM AND TYPE-IDENTITY THEORY**

**Brett Anderson**

In this paper I reconstruct an argument against epiphenomenalism used by William James. If epiphenomenalism were true, then given that humans developed over time due to natural selection, we would expect the relationship of our mental lives to our physical actions to be random. To the contrary, our phenomenal experiences mirror our physical interactions, which strongly indicates the falsity of epiphenomenalism. Type-identity theories also fail because the identities, which are antecedent to any contingent evolutionary context, fail to predict the situation we find, namely that our mental events are of the type we would expect if mental events have causal powers.

---

<sup>1</sup>. This paper was submitted to *Sorites* in 2007 and was accepted for publication in 2008. Since then, I have explored many aspects of the philosophers' appeal to intuitions in more recent publications, which I now include in the references. Apart from updating the bibliography, though, I have made no other changes to the accepted manuscript.

SORITES (Σοφίτης), ISSN 1135-1349  
http://www.sorites.org  
Issue # 21 — June 2008. Pp. 06-11  
Evaluating Williamson's Anti-Scepticism  
Copyright © by Tony Cheng and SORITES

## EVALUATING WILLIAMSON'S ANTI-SCEPTICISM

Tony Cheng

1. Epistemologists have different attitudes toward scepticism about knowledge, but no one can sensibly deny its central place in at least some branches of philosophy. We can trace its history back to ancient Greek, and nowadays there are still sceptics around us.<sup>1</sup> It's fair to say that scepticism is a necessary ingredient of epistemology.

Timothy Williamson's *Knowledge and its Limits* can be seen as a systematic response to scepticism.<sup>2</sup> He starts with the slogan «knowledge first,» and elaborates it by arguing for the thesis that knowledge is a basic mental state: knowledge is basic in the sense that it cannot be analysed with more primitive terms (primeness), and knowledge is mental because it is a species of propositional attitudes that is externally individuated (broadness). These moves are anti-sceptic because knowledge essentially involves external conditions, *and* we cannot factorise knowledge into internal and external conditions.<sup>3</sup> The anti-sceptic character of the book becomes more prominent when Williamson attacks the idea that we are always in a position to know our own mental states (anti-luminosity), a Cartesian way of thinking that often leads to scepticism. Williamson goes on to criticise another popular (maybe not so popular today) idea in epistemology, that in order to know something, we need to know that we know it (anti-KK principle). The idea under attack also leads to scepticism, for intuitively we do not know whether we are in sceptical scenarios or not. After setting up these main points against scepticism, Williamson begins to apply them to different topics, surprise exams

---

<sup>1</sup>. For a defence of global scepticism, see Peter Unger, *Ignorance: A Case for Scepticism* (USA: Oxford University Press, 1975). If we align sense-datum theories with scepticism, then Howard Robinson, *Perception* (Routledge, 1994) and Evan Fales, *A Defense of the Given* (Rowman & Littlefield Publishers, 1996) also offer positive cases for it.

<sup>2</sup>. Timothy Williamson, *Knowledge and its Limits* (USA: Oxford University Press, 2002). Philosophers have different readings of its main purpose; see extant reviews of this book. Among them Gilbert Harman, «Reflections on Knowledge and Its Limits,» *Philosophical Review* 111 (2002): 417-28 also emphasises its an-sceptic character. This complex piece allows for more than one reading.

<sup>3</sup>. The two moves do not always go hand in hand, so the conjunction here is crucial. Some philosophers concede that mental states are essentially individuated with the help of external conditions, but nevertheless hold the separability thesis about the internal and external factors. The internal factor is dubbed «narrow content.» We can say that these philosophers endorse broadness but deny primeness. For a recent defence of this view, see G. M. A. Segal, *A Slim Book about Narrow Content* (The MIT Press, 2000). For a reply to Williamson precisely on this point, see Ralph Wedgwood, «The Internal and External Components of Cognition,» in *Contemporary Debates in Cognitive Science*, ed. Robert Stainton (Blackwell Publishing Professional, 2006, 307-25).



for example, and goes on to develop further elements in his overall project. The profoundness of the book is impressive.<sup>4</sup>

Williamson is not satisfied with the above treatments of scepticism. He is aware that some people argue for sceptical cases on the ground that normal people and their counterpart brains in vats (BIVs for short) share the same evidence. Williamson attempts to damp this line of thought by espousing an extreme thesis that evidence equals knowledge ( $E = K$ ). If that is so, our counterparts and ourselves do not share the same evidence, for we have knowledge but they do not have. Williamson hunts the sceptics until their last breaths.

As the above reconstruction shows, Williamson launches a range of arguments to repel scepticism. Some of them are centered on his conception of evidence. They will be the focuses of the present essay. I attempt to show that Williamson's notion of evidence begs the question against scepticism, that  $E = K$  is unnecessary as far as scepticism is concerned, that the argument based on  $E = K$  is incompatible with content externalism, and that «sameness of evidence» is not required by scepticism. Let me go through them in turn.

2. First of all, Williamson's stipulation of the notion of evidence begs the question. He introduces his notion of evidence in the following two passages:

That one has the same evidence in the good and bad cases in a severe constraint on the nature of evidence. It is inconsistent with the view that evidence consists of true propositions like those standardly offered as evidence for scientific theories. (*Knowledge and its Limits*, 173)

The communal case is needed: science depends on public evidence, which is neither the union nor the intersection of the evidence of each scientist. (*Knowledge and its Limits*, 185)

Here we can find that Williamson has a specific conception of evidence in mind: only public, true propositions can be evidence. I think it is not correct even in scientific contexts: to think only true propositions can be counted as evidence is to presuppose the God-eye point of view. Scientists have to start with evidence before they know whether those propositions are true or not. But let me grant that point to Williamson for the sake of argument. Now, even the concession has been made, to stick to this allegedly scientific sense of evidence and say that it is inconsistent with sceptic's notion of evidence is not to the point, though it is true: they are indeed incompatible, but to simply invoke a notion from science and say that it is incompatible with scepticism does not damp the sceptical worry at all. It's just like saying that «look, that's just how science works, so scepticism is false.» If this can be a successful case for anti-sceptics, epistemology is not so hard (and interesting) a subject.

What I am suggesting is the recognition of two different senses of evidence. The scientific sense is what Williamson has in mind (though it is not without problems), but it is irrelevant to the discussions of scepticism anyway: when a sceptic claims that a subject has the same evidence in the good and the bad cases, she is not saying that the subject has the same *public, true propositions* in the two cases. This is a nonstarter for scepticism. If we construe it this way, we cannot even make sense of it; we do not even have the slightest reason to believe it might be true. What the sceptic is driving at is that in both the normal case and the BIV scenario, the subject is not in a position to find any difference in her

---

<sup>4</sup>. This paragraph should not be seen as a comprehensive summary of Williamson's book. My selective reconstruction here is supposed to echo my main line in this essay – how Williamson deals with the notion of evidence in order to attack scepticism. I will explain this presently.

experiences. It is experientially indistinguishable for her. Subjectively, it just makes no difference. The power of this point is that in daily life we really have indistinguishable hallucinations, though not very frequently.<sup>5</sup> To stipulate the objective notion of evidence does not help here. We still want to know how to avoid scepticism and at the same time do justice to the possibility of experiential indistinguishability. Williamson takes issue with scepticism at this point by arguing against the phenomenal conception of experience, but only very old philosophical theories would hold that evidence consists in private sense-data or something like that even for science. The general problem here is that Williamson does not really confront scepticism in his arguments based on the notion of evidence.

3. Furthermore, to identify evidence with knowledge is excessive as far as scepticism is concerned. To see this, consider Williamson's claims about broadness and primeness. Williamson, like some other anti-sceptics, contends that the experiential indistinguishability does not imply that subjects in the good case and in the bad case share the same *mental state*: granted they have the same phenomenal feels, but the ontology of states should not be individuated by experiential sameness.<sup>6</sup> That's what motivates Williamson's *primeness* claim. For if one is satisfied with *broadness*, the possibility that subjects in the good case and in the bad case share the same *narrow content* has not been ruled out. With this state externalism in hand, Williamson can legitimately claim that we have knowledge whereas BIVs do not.

But then it is not clear why Williamson needs further steps. Given his state externalism about knowledge (that is, broadness plus *primeness*), scepticism based on experiential sameness has been refuted. Why do we still need to insist that they have different evidence? Williamson offers the following reason:

A natural argument is by *reductio ad absurdum*. Suppose that one has different evidence in the two cases. Then one can deduce in the bad case that one is not in the good case, because one's evidence is not what it would be if one were in the good case. But even the sceptic's opponent agrees that it is consistent with everything one knows in the bad case that one is in the good case. Therefore, one has the same evidence in the two cases. (*Knowledge and its Limits*, 169-70)

And he goes on to diagnose that «[t]he argument assumes that in the bad case one knows what one's evidence is...» (170). Williamson opposes to this, for he holds an objective notion of evidence, which consists in true propositions, therefore one does not have the same evidence in the two cases. But we have seen that to refute scepticism, state externalism about knowledge is sufficient. The motivation of objecting sameness of evidence is from his objective notion of evidence (and related thesis  $E = K$ ). So it is fair to say that as far as scepticism is concerned, opposing sameness of evidence is ill-motivated.

It is open to Williamson to reply that  $E = K$  is not especially relevant to his anti-scepticism. Scepticism has been refuted in chapter eight and some earlier chapters of the book, but  $E = K$  is argued in chapter nine. It is self-sustaining and not presupposed in his arguments against scepticism.

---

<sup>5</sup>. For a delicate reconstruction of the argument from illusion and the argument from hallucination, see A. D. Smith, *The Problem of Perception* (Harvard University Press, 2002). He argues that even the *possibility* of illusion and hallucination will do.

<sup>6</sup>. Williamson himself appeals to John McDowell, «Criteria, Defeasibility, and Knowledge,» in his *Meaning, Knowledge, and Reality* (Harvard University Press, 1998), 369-94, as an ally. See McDowell's «Knowledge and the Internal» (*ibid.*, 395-413) for further elaborations.

Unfortunately, this line of reply is blocked by Williamson's own summary of his chapter eight. I quote the second half of it:

Since sceptics have *not refuted the equation of evidence with knowledge*, they are not entitled to *assume* that we have no more evidence in ordinary cases than in their sceptical counterparts, for on the view against which they are attempting to argue we do have more knowledge in ordinary cases than in their sceptical counterparts. (*Knowledge and its Limits*, 15, my italics)

Notice that  $E = K$  is invoked *here*, but it is argued *later* in chapter nine. For one thing, it makes the above putative reply unavailable to Williamson; for another, it betrays an oddity of the arrangement of the book: a crucial thesis is invoked before it is argued. What's more, sameness of evidence is not, *pace* Williamson's accusation, an *assumption* on sceptic's part. It is *argued* on the ground of the indistinguishability of the phenomenal. We can challenge this line of thought, to be sure. But we should not regard it as an assumption and thereby attribute the burden of proof to the sceptic. Quite the contrary. The burden of proof is on someone who proposes  $E = K$ , for it is not obvious at all. Now Williamson does offer substantial arguments for the equation, but as I have argued above, what he has in mind is the objective, public, scientific notion of evidence. So the sceptic can grant the equation and still launch her challenge. Williamson says the notion invoked by the sceptic is incompatible with the one he uses, but this only shows that he does not address the real worry of scepticism.<sup>7</sup>

4. My third claim is that the argument from  $E = K$  is incompatible with content externalism, a thesis accepted by Williamson himself. Consider the passage I just quoted from Williamson, and notice the thought behind the reasoning in those remarks: because evidence is *factive* thanks to  $E = K$ , the BIV does not share the same evidence with his normal counterpart, for the former has *false, therefore fake* evidence, the latter has *true* one. But by *what standard* the putative evidence of the BIV is false? When he uses «I have hands» as evidence to support his other beliefs, the putative evidence is false *on the assumption that it is the same thought with the one entertained by his normal counterpart*. But this is wrong according to content externalism. For example, according to certain version of content externalism, the content of BIV's thought is constituted by the stimulations he receives, so if there is certain corresponding relation between the thought expressed by «I have hands» and certain patterns of brain stimulations, the thought in question may well turn out to be true. Therefore, it becomes not clear that whether content externalism sides with Williamson or not.<sup>8</sup>

There are puzzles about assigning truth-values to the BIV's thoughts, to be sure, for there seems to be no objective ground to generate truth and falsity. Williamson can then argue that since «I have hands» is true in the good case but devoid of truth value in the bad case, sameness of evidence collapse anyway. However, there are a lot more to be said in taking this

---

<sup>7</sup>. In his review, Richard Foley echoes this passage by saying that «[w]hile envatted, John [the BIV] is severely deprived of knowledge of his environment, but because on Williamson's view, one's evidence is co-extensive with what one knows, John is also thereby severely deprived of evidence.» (721)

<sup>8</sup>. There is a crucial difference between the BIV case and the hallucinatory case here. It is relatively easier for Williamson to answer the present challenge if we construe it in terms of the hallucinatory case, for in the BIV case the normal subject and the BIV do not share the same thought when both of them entertain «I have hands.» Here I am indebted to Branden Fitelson.

tack, so whether content externalism can be used to counter against scepticism remains an open question.

5. This brings us to the last point I would like to make. The above diagnosis points to the fact that Williamson does not fulfill his promise of substantially relating epistemology to philosophy of mind in his chapter on scepticism. Although Williamson's intention to connect these two subjects can be easily detected in many places in the book, and this is indeed admirable, he nevertheless only scratches the surface in this particular chapter. At the beginning of it, he briefly mentions content externalism and says:

We assume for the sake of argument, *perhaps over-generously*, that the sceptic has some way of absorbing...implications of content externalism. (*Knowledge and its Limits*, 165, my italics)

Here he writes as if content externalism is *definitely* a bad news for scepticism, but as my third point shows, this seems to be too quick. In what follows I will give a further reason (that is, my fourth point) why I think it is too quick. The line of thought in the quotation is this: the sceptic needs sameness of evidence, beliefs or perceptions or whatever, to establish her sceptical case. But given content externalism, the subject in the good case and the one in the bad case do not share most of their beliefs and (quasi-)perceptions, for their external environments are radically different. It seems that scepticism cannot get off the ground if content externalism is true. Now Williamson grants that maybe the sceptic can find a way out of this, but still, she will face some additional problems. That's the idea. But is the dialectic situation that simple?

So the fourth point is this: given content externalism, and *given that the sceptic needs sameness of evidence*, the sceptical challenge evaporates right away. But is it true that sameness of evidence is required for the sceptic? I submit the negative answer.

It is true that the skeptic needs something common to the good case and the bad case, but that *something* should be *the support relations between evidence and hypotheses*, not the evidence itself. If the normal subject and the BIV might be *equally justified*, but only the former has knowledge, then the sceptic is home and dry.

Notice that the notion of «equally justified» is *not* interchangeable with the notion of «having the same evidence.» An analogy might help. Valid deductive arguments are equally valid, but they may well have different premises. It's *the degree* of supporting, not *the ground* of supporting, that is supposed to do the trick for scepticism. In this sense Williamson's whole argumentation against scepticism is beside the point.

Think about content externalism. Assuming the sceptic in question champions this view herself. By her light, when both the normal subject and the BIV say the sentence «I have hands,» they are in fact expressing different thoughts despite the fact that the thoughts are instantiated by the same sentence. Now she grants Williamson's arguments against sameness of evidence, but argues that there is no obvious reason why the BIV *cannot share the same degree of supporting* with the normal subject: surely the ground of supporting in the BIV case may be bad, but what prevents this bad ground from supporting the hypothesis (i.e., the belief to be justified) in question well? Granted, they have different hypotheses thanks to content externalism, and they have different evidence for those hypotheses thanks to Williamson, but

what on earth is supposed to establish the claim that the two subjects cannot have the same degree of supporting?<sup>9</sup> It seems to me that Williamson does not address this issue at all.

Let me adumbrate the points I have made before closing this short essay. First, Williamson's notion of evidence does not make any direct contact to scepticism. I am not saying that anti-sceptics need to invoke the notion of evidence adopted by the sceptics, but at least the former needs to address the latter's worry. It is not clear to me that Williamson is aware of this. Second, the role of  $E = K$  in the book is not clear. Because broadness and primeness are sufficient for rejecting scepticism, and because  $E = K$  is argued after the scepticism chapter, it seems that Williamson's anti-sceptic arguments can do without it. However, Williamson's own outline of that chapter suggests the other way around. The status of  $E = K$  is further undermined when we see that there is a tension between content externalism and it; this is the third point. And finally, the consideration of content externalism helps bring out the fact that sameness of evidence is a red herring as far as scepticism is concerned: if that were the locus of the debate, scepticism would be simply ruled out by content externalism; what should be at issue is sameness of the degree of supporting, not evidence.

6. Overall, *Knowledge and its Limits* presents not only a strong challenge to traditional epistemology, but also a solid ground for a systematic, positive project to pursue. Its treatments of scepticism, however, are not so satisfying because of the problems I discussed above. It is not that they cannot be answered by Williamson. The moral should be more positive: given the success of other parts of the book, we expect the book to do better when he takes issue with scepticism. His arguments against scepticism are as intricate as his arguments elsewhere, albeit the minor flaws I tried to point out above. I believe his case against scepticism can be more successful if he adjusts the thrusts along the lines I suggested in this critical notice.

### Acknowledgement

This paper was written when I was a visitor of U. C. Berkeley, fall 2006. I would like to express my gratitude to Brandon Fitelson and Sherrilyn Roush, for their patience and helpful comments. For technical reasons the paper did not come out until recently, but I leave the paper as it was.

**Tony Cheng**

**University of California, Berkeley**

**[h.cheng.12@ucl.ac.uk](mailto:h.cheng.12@ucl.ac.uk)**

---

<sup>9</sup>. Williamson might say that we cannot simply stipulate that they can share the same degree of supporting. He expresses his hostility to stipulations in the sceptical scenario by arguing that stipulation violates the requirement of constructing non-trivial connections between epistemic notions (169). But this begs the question against scepticism, for what it objects to is exactly substantial relations between epistemic notions. All the sceptic needs is the intuitive plausibility of conceiving sameness of supporting degree in the two cases.

SORITES (Σοφίτης), ISSN 1135-1349

<http://www.sorites.org>

Issue # 21 — June 2008. Pp. 12-30

Velleman on the Constitutive Aims of Practical and  
Theoretical Reasoning

Copyright © by Emer O'Hagan and SORITES

## VELLEMAN ON THE CONSTITUTIVE AIMS OF PRACTICAL AND THEORETICAL REASONING

Emer O'Hagan

---

Reasons for action and belief are authoritative rational norms. Just how they are authoritative, how they impose rational constraints, and what counts as good reasons and good reasoning are important and related conceptual questions. In recent years philosophers have tried to answer some of these questions, and sort out these conceptual relations, by advancing constitutive arguments. Constitutive answers to the question 'how do rational principles bind agents?' hold that rational norms are inherently authoritative over agents *qua* reasoners; agents and reasons for belief and action are mutually constituted. What might then be called a constitutive approach to answering such questions involves a general strategy for grounding the authority of rational norms in the nature of agency or autonomy.

Part of the appeal of a constitutive understanding of the nature and normativity of reasons for action and belief lies in its resources to simultaneously unify and distinguish between practical and theoretical reasoning. Because it makes the problem of the normativity of reasons a problem for reasons *per se* and not a problem for practical reason alone the constitutive approach does not foreclose on the possibility that practical norms share a ground with norms of theoretical reasoning. To many a unified account of reasoning is an attractive prospect. Practical and theoretical reasoning, on such a view, will be distinguished by the domains over which reason has authority, action or belief, and it remains an open question whether the normative constraints theoretical reasoning imposes on reasoners share a ground with the normative constraints practical reasoning imposes on agents. Accordingly such an approach does not prejudge the issue of how moral norms bind agents by supposing that theoretical reasoning is inert and that moral norms are grounded in desires. Neither does it prejudge the question of whether moral reasons are overriding.

How do moral reasons have authority over agents? Philosophers such as Christine Korsgaard hold that the question «why be moral?» can be answered by an appeal to agents as constitutively governed by reasons. Korsgaard (1996), for example, has argued that it is constitutive of agency that agents are bound by reasons for action. On her view, the authority of reasons, or rational principles, is internal to the practice of reasoning itself. To be an agent at all just is to be reason-bound. Agents have a reflective form of consciousness which makes

action (rule-governed activity) possible and this fact about us makes opting out of the business of reasoning impossible. The question of which reasons are good reasons, of course, remains.

A parallel issue for theoretical reasoning exists: the question «why be governed by what there is good reason to believe?» can be answered by an appeal to agents as constitutively governed by reasons. Lewis Carroll's (1895) famous parable of Achilles and the Tortoise gives the example of a deductively valid argument in which the conclusion is entailed by the premises and challenges us to consider what to make of Tortoise's refusal to draw that conclusion. Tortoise affirms P, and affirms that if P then Q, but refuses to draw Q as a conclusion.<sup>1</sup> What sort of error, if any, is made by an agent who refuses to draw such a conclusion, or to be guided by what there is most reason to believe? Carroll's philosophical puzzle has been taken by some (for example, Brandom, p. 24) to indicate that the correctness of even logical claims must be secured by their normative pragmatic significance. Recognizing the roles norms play in reasoning itself lends itself to the constitutive approach.

Yet even those sympathetic to a constitutive approach may worry that it cannot vindicate the authority of reasons, as it may seem that the constitutivist is getting something for nothing. Peter Railton (1997), for example, levels some important objections to the constitutive approach, challenging its claim to explain the authority of reasons over reasoners. Railton grants that an implicit commitment to the authority of reasons is evident in the practice of reasoning, but denies that this serves as a sufficient demonstration of the categorical authority of standards of reasoning. He argues that something stronger than the practice of reasoning itself is required to vindicate rational norms. His main objection is that the constitutive view, with its reliance on a conception of agents as constitutively norm-governed, fails in the defence of the categorical authority of reason over agents because the rational standards appealed to are not independent standards, but hypothetical rational imperatives. Although they are rational norms it is nonetheless neither impossible nor irrational to abandon them, and hence they cannot be invoked as intrinsically normative.

Railton's line of objection is important because it forces us to consider how far a constitutive view can go before normative authority bottoms out in descriptive fact. The worry is that the so-called authority of rational norms just amounts to a fact about how we do things. But this objection to the constitutive approach is not successful because it presupposes a dubious foundationalism about justification and a mistaken view of what it means to hold that agents are constitutively norm-governed, so I will not discuss it here<sup>2</sup>. It is, of course, legitimate and important to worry about the relation between facts and norms in any of these accounts. In this paper I want to raise some worries of my own about an attempt to derive rational norms from purported facts about the aims of action and belief. Here I want to tackle a different objection from a critic who is nonetheless sympathetic to the constitutive view, the charge that the sort of constitutive approach outlined above is viciously circular, and I want to challenge his proposal for resolving the problem he sets by focusing on his attempt to construct norms out of facts.

J. David Velleman has argued that certain formulations of the constitutive approach are circular and has advanced in their place an account which characterizes reasons for belief and

---

<sup>1</sup>. Pascal Engel, laying out the problem in this way, describes the Tortoise as an inferential akratic, p. 22.

<sup>2</sup>. I have argued for this in «Belief, normativity and the constitution of agency,» (2005).

action as constituted by the mechanisms which produce them. Where other proponents of constitutive views take norms, reasons or rationality to be constitutive of agency, Velleman argues that if the products of individual episodes of reasoning manifest the constitutive aim of theoretical or practical reasoning, they are beliefs or actions respectively. According to Velleman, the constitutive aim of a behavior defines that at which a behavior must aim in order to count as being a behavior of that kind; beliefs manifest the constitutive aim of theoretical reasoning, whereas actions manifest the constitutive aim of practical reasoning. He identifies these aims in the following way: beliefs are cognitive states which are regarded as true with the aim of thereby accepting a truth, actions are activities in which one knows what one is doing.

Velleman's position on practical and theoretical reasoning is largely a consequence of his view that in order to avoid vicious circularity the constitutive aims of practical and theoretical reasoning must be formulated in purely non-normative terms. Velleman argues for (what I will call) the conceptual independence thesis, the view that the constitutive aim of belief must be understood in terms which are conceptually independent of reasons for believing on pain of circularity, and the constitutive aim of action must, for the same reason, be understood in terms which are conceptually independent of reasons for action. Although we may be tempted to think that reasoning aims at having good reasons for belief and action, reasoning cannot, on pain of circularity, be constituted by such an aim.

I will argue that the conceptual independence thesis, that the constitutive aim of belief and action must be characterized in terms which are conceptually independent of reasons for belief and action respectively, is not adequately supported by Velleman's arguments. I will further argue that it is actually a source of the tension within his position. I hope to raise worries about the value of the conceptual independence thesis itself, arguing that the normativity of reasons cannot be explained if the conceptual link between having reasons and the practice of reasoning is cut. We will see evidence of the need to maintain this link in Velleman's own appeal to agency as essentially reason-governed, an appeal which undercuts the conceptual independence thesis. I will discuss the circularity problem and then evaluate the consequences of holding the conceptual independence thesis for the question of how rational principles bind agents first in the case of practical reason and then in the case of theoretical reason.

### **The Circularity Problem**

Velleman's approach to the authority of rational norms is constitutive in the sense that it seeks to identify the products of reasoning according to whether they bear a particular object or aim which is internal to reasoning itself. But his advancement of the conceptual independence thesis is in serious tension with this goal. Any philosopher with proclivities toward naturalism will feel the need to offer an account of norms which is at some level purely descriptive, but doing so in a way which does not undo the authority of reasons one wishes to ground is not an easy task. I argue that Velleman's conceptual independence thesis does not help this task.

For Velleman, in order to figure out what counts as success in reasoning we first need to determine what it is that reasoning is supposed to accomplish; before we can make claims about reasons which have normative authority we need to figure out what, as a matter of fact, the aim of reasoning is. In this way we can establish a non-controversial norm of correctness in reasoning, non-controversial because internal to the very nature of believing or acting:



If there were something at which action constitutively aimed, then there would be a norm of correctness internal to the nature of action. There would be something about behavior that constituted its correctness *as* an action, in the same way as the truth of a propositional attitude constitutes its correctness as a belief. This standard would not be open to question: actions meeting the standard would be correct on their own terms, so to speak, by virtue of their nature as actions, just as beliefs are correct by virtue of their nature as beliefs. And this norm of correctness for action would in turn determine what counts as a reason for acting. (Velleman, 2000a, p.16)

Reasoning cannot, on pain of circularity, be constituted by the aim of having good reasons for belief and action. We see this, Velleman argues, when we consider the difference between formal and substantive objects. The goal of an enterprise described in terms of the very concept of the object of that enterprise (or in terms that depend on it) is its formal aim (Velleman, 2000c, p.176.) For example, winning is the formal object of a competitive game. The substantive object of an enterprise, for example, winning a game by scoring the most points, expresses a goal stated in terms other than being the object of that enterprise. According to Velleman, the first step in avoiding the circularity problem is to identify the aim of reasoning in substantive, not purely formal terms. A purely formal conception of the aim of practical or theoretical reasoning ties it to the very concept of reasoning itself, and so fails to explain what counts as success in reasoning.

Purely formal accounts fail to explain reasons for belief and action because they rely exclusively on those concepts in their explanation. If having a reason (to believe or act) is just being in a state that would be produced by competent reasoning then the explanation is circular. As Velleman (2000c, p.177) puts it:

A mode of reasoning whose goal was specified solely as 'figuring out what to do' would be like a search whose object was specified solely as 'figuring out where to look,' or a question whose object was specified solely as 'figuring out how to reply'.

On the other hand, the aim of reasoning cannot be understood as expressing a particular rational standard without thereby invoking a value judgement which itself requires justification. So conceptions of the aim of reasoning which are purely formal are either viciously circular or incomplete. An explanation of practical or theoretical reasons will be circular if it requires that an agent must have a concept of a reason in order to act at all. It will be incomplete if it invokes a standard of rational justification which itself requires further defense. Velleman concludes that the most promising way for a constitutive approach to proceed is by identifying the substantive and non-normative aim of reasoning, an aim which must be conceptually independent of reasons for believing or acting. This procedure should ensure a non-circular account by identifying a condition for success in reasoning which is norm-free and by identifying the products of reasoning without reference to the results that good reasoning ought to produce.

## **Practical Reasoning**

I want now to examine whether Velleman's conception of the aim of practical reasoning has the resources to explain how rational principles bind agents. If we grant the necessity of identifying a substantive, not purely formal aim of reasoning do we thereby give up on the possibility that practical reasoning aims at discovering something as general as 'what there is most reason to do'? Velleman thinks so, and argues that the aim of practical reasoning cannot be the good, or what there is most reason to do. According to Velleman, the constitutive aim of action provides an internal criterion of success for action. An ideal norm of correctness for

action would not leave any room for questions about whether one should act in accord with that norm; if reasons for acting in accordance with the norm need to be supplied then those reasons will determine what counts as a reason for action and the explanation will be circular (Velleman 2000a, p. 16). Because reasons for action qualify as reasons through their relation to this constitutive aim, avoiding circularity requires first identifying the function of practical reason in non-controversial terms and then evaluating particular reasons in terms of how well they meet this function. To hold that one reason is better than another is to hold that it better satisfies the function of practical reason.

This leads Velleman to assert that reasoning's constitutive aim must be sub-agential, not something which an agent consciously seeks in acting, lest the circularity problem reappear. Noting rightly that action theory has, for the most part, failed to deal with the diversity of cases that fall outside of fully intentional actions, Velleman devises a position which holds that behaviors become actions as they make evident the aim of practical reasoning; action is behavior executed under conscious control (Velleman, 2000c, p. 192). A slip of the tongue, for example, is a case of behavior not fully intentional activity, in which the speaker fails to inhibit his verbal activity and does not know what he is saying. Just as our verbal behavior is governed by the aim of knowing what we are saying, only albeit rarely as a conscious aim, rather as a sub-agential aim, action itself can be seen as being regulated by the general aim of knowing what one is doing. His analysis leads him to the conclusion that the aim of practical reasoning is to know what one is doing.

Reasons for action, considerations in light of which an action would make sense to an agent performing it, qualify as reasons by providing the agent with an understanding of how her behaviors are her own by providing a form of self-knowledge: «reasons for doing something are considerations that would provide the subject with an explanatory grasp of the behavior for which they are reasons.»<sup>3</sup> When I know what I am doing I conceive of my movements under concepts that have explanatory power, given the context of motives and circumstances. When we act in a fully intentional way we know what we are doing. We have self-knowledge in the sense that our behavior is brought into the realm of reasons so that awareness of what we are doing becomes a belief about our own behavior; we accept the truth of what it is we are doing in such a way that it is made true (Velleman, 2000c, p. 195). Our reasons for action tell us what we would be doing if we took them to be true.

What then makes one reason for action better than another? According to Velleman, some reasons are better than others by virtue of their nearness relation to the constitutive aim of action, the source of normativity for reasons: «the better the reason will be the one that provides the better rationale – the better potential grasp of what we are doing» (Velleman, 2000a, 29). Better reasons are those which more successfully evidence the constitutive aim of knowing what one is doing. Freudian slips, unlike full-blooded intentional speech acts, are examples of utterances which do not manifest the speaker's grasp of what she is doing; in such cases we act without good reason because we fail to fully understand what we are doing. In this way Velleman attempts to fix the goodness of reasons without appeal to the product of competent or good practical reasoning, and without circularity.

---

<sup>3</sup>. Velleman, 2000a, p. 26. In an earlier formulation of his position Velleman has called the constitutive aim of action «autonomy», he now favors «knowing what one is doing».

Part of the appeal of Velleman's position is its attention to significant and under-discussed dimensions of action. He is surely right that our reasons for acting are, in one respect, better when we are aware of what we are saying and are not moved entirely by sub-conscious forces. But this is a special sense of 'better reasons for action' and one with limited normative consequences. Better here implies a minimal kind of self-awareness. This minimal sense applies, for example, when I cut my finger chopping vegetables, seeing just prior to doing so that I am about to cut my finger, and accepting that as a fact. It is minimal precisely because there is no normative sense in which I have a *reason* to cut my finger in this case. Indeed, even the value of self-knowledge in this case is slight, and although it may have been true that I knew what I was doing, as an agent I aspire to be guided by my knowledge, hence by good reasons, not simply to see what I do as I do it. In this minimal sense I can know very clearly what I am doing and why, without being guided by reasons at all.

It might be objected that although this example involves self-knowledge, it is not a problem for Velleman's view because it is not a case of acting on a consideration that contributes to self-knowledge. According to Velleman, reasons for action are considerations that make an action make sense to the actor by giving her an explanatory grasp of the behavior for which they are reasons. Reasons explain to me my behaviors; self-knowledge exists as a form of beliefs about my behavior. My reasons for action are good reasons when they provide me with a good grasp of what I'm doing. Cutting one's finger (even though it is something that one does to oneself) is at best an activity and not an autonomous action. It is peculiar only because I am the victim of my own lack of skill, but there is no good reason to cut one's finger because there is no sense in which 'cutting my finger' is ever in such circumstances thought of as justified from the agent's point of view.

However, if reasons are produced by a mechanism designed to regulate action, and their conceptual analysis requires that we first identify the mechanism and only then begin to evaluate reasons as good or bad in reference to how well they satisfy this regulatory function of action, then the rationalizing force of the reason (the part of the cognitive process which includes being thought of as justified from the agent's point of view) can not itself appeal to justificatory norms. Justificatory norms depend for Velleman on the mechanism, not the other way round. So even though Velleman insists that the concept of reasons must include a sense of acting on considerations that contribute to self-knowledge, he isn't entitled to a use of this knowledge which would contribute justificatory reasons.

Consider a case of practical reasoning aimed at answering the question «should I go to the protest march?» In such a case I will deliberate over my obligations, the likely consequences of my action, and my motives. If I am torn (perhaps I believe that I ought to participate but dislike the social tension at such events) then I will struggle to make a choice and self-knowledge could be very useful help. If I come to see, for example, that my reluctance to participate in the protest march is really an attempt to avoid a potentially hostile and uncomfortable situation, I gain a measure of moral control and if I am honest I will confront the conflict between what I am doing and what I say I am doing. Is this a case of having an explanatory grasp of what I am doing and hence a reason for action? If I am feeling low I may skip the march, rebuking myself for failing to do what I ought to have done, and if I have self-knowledge I will see in my choosing that personal comfort has won out over political principle. In this case it seems that I clearly see what I am doing, it also seems that I accept the truth of what I am doing in such a way that it is made true; it includes a

consideration that contributes to self-knowledge and so fits Velleman's account of an autonomous action and a good reason.

Since the second case fits Velleman's account in exactly the same way as the first, but the two cases differ exactly in respect of offering good reasons, Velleman's account fails to explain to explain how some reasons are better than others. Self-awareness is potentially a moral and an epistemic virtue because it gives one information about one's motives and this is relevant to reasoning well, but mere self-awareness of first person facts is not the same as reasoning well. Having espoused the conceptual independence thesis and the need to identify a non-controversial aim of reasoning, Velleman cannot, by the constraints of his own argument, take self-awareness to have a significant normative dimension. It cannot, for example, simply replace the guise of the good<sup>4</sup>, nor can it require guidance by the weight of reasons without smuggling in an essential normative dimension. But lacking a significant normative dimension the authority that reasons have over agents goes unexplained. If I want to avoid injury, I have more reason to chop vegetables slowly than quickly and I ought to slow down. If I want to have integrity I should live in accord with my political beliefs. But if the better reason were just the one that provides a better grasp of what I'm doing, then what I have most reason to do will just indicate whatever I am, as a matter of fact, doing whether that is being cautious or incautious, strong or weak-willed, and it cannot promote or endorse one over the other. A more robust sense of self-awareness, one which included a notion of the self as constitutively reason-governed, for example, would be able to overcome this problem but would, in doing so, violate the conceptual independence thesis by making reason's aim depend on normative standards that are part of the practice of reasoning. It appears that Velleman faces a dilemma: if you try to explain how reasons bind agents in terms of a goal of reasoning but you specify the goal of reasoning without reference to a substantive idea of reason your explanation will either be too minimal to account for the force of reason or will violate the conceptual independence condition. This suggests that the conceptual independence condition is the source of the problem.

Can the aim of practical reasoning be expressed in substantive terms which nonetheless allow for an account of how agents can be guided and constrained by rational norms, without vicious circularity? Philip Clark (2001) has argued that we can grant that the aim of practical reasoning cannot be expressed in purely formal terms (in terms of the concept of being the aim of that enterprise) without giving up on a conception of the aim of practical reasoning that is itself norm-governed.<sup>5</sup> We can avoid the circularity problem by adopting a substantive conception of the aim of practical reasoning but we can do so without adverting to a non-normative mechanism of action. This strategy has the advantage of maintaining the conceptual link between reasons and rational agency and so has resources for explaining how agents are governed by good reasons. The aim of practical reasoning can be expressed in substantive, but generic terms without vicious circularity and without incurring a commitment to any particular standard of practical rationality which requires justification. Clark points out that it cannot be the mark of the substantive that it is a real, natural phenomenon, that it exists. The edge of the Earth, he notes, expresses the substantive object of a search even though the edge of the

---

<sup>4</sup>. In «The Guise of the Good,» Velleman (1992) argues that practical reason should not be understood as a cognitive power for pursuing value.

<sup>5</sup>. Clark defends a form of externalism about practical reason against arguments advanced by Velleman, specifically, he defends the position that reasons for action can be conceived of in terms of the good for agents.

Earth doesn't exist. It is substantive because it is not defined formally in terms of being the object of a search; the mark of the substantive is that it has «a semantic life of its own, apart from the notion of being the object» (2001, p. 587).

'What there is most reason to do' is similarly substantive in the sense that it describes an aim in terms other than being the object of the enterprise of practical reasoning itself. Its standard has its own semantic life, picking out something other than a purely conceptual description of practical reasoning which would be either trivially true or circular. 'What there is most reason to do' identifies a generic standard for success which may be achieved by competent practical reasoning (a truth about any proposed standard of practical reasoning) but isn't defined in terms of competent practical reasoning. In short, the notion of a standard or ideal does not require a specification of how it is to be met. Standards are, in general, the kind of thing which transcend any particular conception or implementation of them. Accordingly, to say that *X* reflects a standard of superior performance is not to say what makes a particular performance superior. Velleman's claim that a substantive, generic expression of the aim of practical reasoning incurs a commitment to a particular standard which is itself in need of justification is thus groundless and we need not accept his assertion that such explanations are incomplete or circular. The aim of practical reasoning can be substantive without expressing a specific standard of success in reasoning. This does not, of course, supply an answer to the question 'what makes a consideration count as a good reason' it merely defends the view that practical reasoning is aimed at the good against Velleman's objections.

Consider a case of practical deliberation. When one is wondering whether to keep one's promise on a particular occasion, 'what there is most reason to do' can provide a substantive goal for deliberation without committing one to a specific view of what there is most reason to do. One's aim is to figure out what one should do. The practice of reasoning about what to do is entirely compatible with uncertainty about whether there is some (one) thing that ought to be done. The search for a right answer to the question «what should I do?» does not itself impose a justificatory burden; a justificatory burden arises only once a specific standard is in place (Clark, 2001, p. 589).

In particular, having a view about the best reasons doesn't commit me to holding that my view is right. Consequently, just because 'what there is most reason to do' is a substantive object it need not express a specific conception (mine) of what there is most reason to do; it can be both generic and substantive. Velleman hasn't shown that a non-formal and non-specific description of the aim of practical reasoning must be guilty of either vicious circularity or appeal to an undefended substantial standard that requires justification. His claim that the very possibility of reasons for action depends upon correctly identifying a substantive, non-generic aim of practical reasoning is thus unfounded; we can make sense of reasons for action without first identifying a non-normative mechanism of action. The fact that the aim of practical reasoning can be characterized in terms which are not conceptually independent of reasons for action without vicious circularity shows that the conceptual independence thesis is false.

It might be objected that the substantive, generic aim of practical reasoning provides only a technical victory - the standard itself is incomplete, reasoning about what to do doesn't amount to anything until there is a particular conception of reason in place. Demonstrating that the aim of practical reasoning can be expressed in substantive but generic terms doesn't demonstrate what ought to be done, and so provides no substantive guidance. However, there is no reason to think that a conceptual analysis of the aim of reasoning should give specific

guidance. The point is not to argue for a particular version of good action (which would incur a justificatory debt) but merely to show that a normative account of the aim of practical reasoning need not be incomplete. As a standard, ‘what there is most reason to do’ picks out a privileged class of actions but not by smuggling in a judgement about what makes reasons good. ‘What there is most reason to do’ expresses a substantive object or aim of practical reasoning without expressing a specific conception of which reasons are good.

Clark provides varied examples to bring home this point. If I tell you to bring me the best gumshoe in town I do not thereby commit to a evaluative view on quality detectives - I may well be unsure and so seek your assistance. If I tell my child to live a good life I do not mean that she should live the life I think is good, even though I have specific views about what makes a life good. ‘Good life’ refers to an evaluative standard, not a description of a state of affairs. Prohibitions against cruel and unusual punishments also function as substantive but generic standards in this way; when written into a constitution they are intended to rule out a privileged class of actions that are identifiably cruel and unusual, not whatever the authors of the constitution believe them to be (if that were so the authors could merely list them). Generic standards do not require justification because they do not advance particular value judgements, but this fact does not make them incomplete.

It might further be objected that a substantive generic aim of practical reasoning such as ‘what there is most reason to do’ still seems circular because it characterizes reasons and what there is most reason to do simultaneously. The conceptual independence thesis demands that we keep the aim of practical reasoning and reasons for action conceptually independent to avoid circularity. But because the substantive, generic formulation of the aim of practical reasoning does not rely exclusively on the concepts concerning what counts as success in practical reasoning, the claim isn’t viciously circular. ‘What there is most reason to do’ has a semantic life of its own and as such its conceptual connection to reasons for action is crucially indirect. It will, on this view, turn out that the process of sorting out what there is most reason to do is what successful practical reasoning (under optimal conditions) would arrive at. But what counts as a reason for action is not defined by what will be produced by successful practical reasoning. For there to be a vicious circle ‘what there is most reason to do’ would have to pick out a purely descriptive claim about the outcome of a process of rational deliberation, and it does not, instead it offers a generic evaluative claim.

The circularity problem is avoided, as Clark has shown, by expressing the constitutive aim of practical reasoning in substantive, generic terms. Hence, Velleman’s justification for the claim that we must look to identify the constitutive aim of belief and action according to the mechanisms which produce them is undermined; he would need to provide some further reason for this strategy. The objector might go on to claim that whether or not Velleman is successful in doing so, a norm-laden constitutive aim of practical reasoning seems unable to explain what makes something a good reason by appealing to ‘what there is most reason to do’, and doesn’t show itself as a positive alternative.

This sort of objection does nothing to strengthen Velleman’s position, of course, but more importantly goes wrong in assuming that a conceptual account of reasoning and reasons should tell us what reasons are good, as a matter of fact. Rational discourse itself, we hope, will work out what counts as good reasons, surely a meta-theory about what sort of phenomenon practical reasoning is will not. Consider how well Velleman’s own account of the constitutive aim of practical reasoning fairs on this score. If knowing what you are doing is the aim of action then I have good reasons when I clearly know what I am doing. So when

I, out of frustration, speak harshly to my Mom I do so with good reason when I am aware of doing it as I do it - when my rationality has been engaged in such a way that I, at that moment, see it as justified. When I speak to her harshly out of frustration and without full awareness I will not have acted autonomously and will not have had a good reason. This is not a promising result, at least if we hope to get beyond motivating reasons to normative reasons. What makes one reason a better reason for action than another is not merely its clear descriptive accuracy. We have seen that norms need not be excluded from the constitutive aim in the first place, so the burden Velleman puts on himself is unnecessary. Moreover, given that practical reasoning governs both prudential action and moral action, it seems implausible to suppose that any single non-normative aim or mechanism of action exists that can explain how they are interrelated.

We have seen that we need not advocate the conceptual independence thesis in order to explain reasons for action, and indeed, that there are independent reasons for rejecting it. The claim that the constitutive aim of practical reasoning must be conceptually independent of reasons for acting threatens to displace reasons from the practice of reasoning, and in so doing to undermine the normative authority of reasons. The conceptual independence thesis, hiving off reasons from the normative commitments involved in the practice of reasoning, leads to a position which cannot explain how some reasons are better than others and how better reasons impose rational constraints on agents. If instead the aim of practical reasoning is expressed in substantive but generic terms, something like 'what there is most reason to do', then we can make sense of the claim that practical reasoning guides and constrains agents, by its very nature. But because Velleman insists on characterizing the goal of reasoning without reference to the notion of a reason, he has no resources for so doing. Success in practical reasoning no longer implies success in meeting rational standards and so cannot guide or constrain agents.

The reasons for abandoning the conceptual independence thesis become clearer once we consider how Velleman might explain the normative authority of reasons for action. Partly in response to Clark, Velleman modified his position to allow that in reasoning about action agents may not be explicitly guided by the aim of practical reasoning; instead the aim may only be operative sub-agentially in the «psychological mechanisms in which that aim is implicit.»<sup>6</sup> So the constitutive aim of action need not be explicit in the agent's conception of what she is doing. It is not, as Velleman had earlier proposed, a desire to be actuated by reasons. He made this change to accommodate the fact that one may reason about what to do without thereby being committed to a particular conception of the good. Reason's aim and a particular agent's act of reasoning come apart. Velleman concluded that practical reason's aim may operate independently of the subject's desires and sometimes be only implicit in the agent's psychological mechanisms. The mechanisms productive of reasons will only sometimes be felt as a desire to know what we are doing. In effect, in order to maintain the conceptual independence thesis, Velleman drove reason's aim further underground. Even if it were true that practical reasoners are moved by sub-agential desires for self-knowledge we can still balk at the claim that that bit of mental machination is constitutive of the process that produces reasons.

---

<sup>6</sup>. Velleman, 2000a, p. 20. Various footnotes and parenthetical remarks in this introduction and in papers in this collection suggest that Velleman is aware of some of the normative problems that arise for his view. If my diagnosis is correct, these problems cannot be solved without abandoning the conceptual independence thesis and making space for a normative dimension in reason's aim.

Why do what there is most reason to do? Indeed, why do what will produce self-knowledge? As we have already seen, there are different senses of self-awareness or self-knowledge. I have argued that Velleman's conceptual independence thesis restricts him to a non-normative conception of self-knowledge. However, this restriction is violated when Velleman adds to his sub-agential account of agency, a necessarily imprecisely described «desire to be actuated by the best reasons.»<sup>7</sup> Now it must be said that if our tendency to be moved by better reasons is a desire, it is a desire of a non-standard sort. It is added to his account in order to explain how we are directed toward what there is most reason to do. This mechanism is essential to agency and it does the work of completing practical reason's normative aim, Velleman thinks, in a non-controversial, non-normative way. We can see that Velleman has come a long way in the attempt to satisfy the demands of the conceptual independence thesis. It is clear that something like a tendency or inclination to be moved by the best reasons belongs in an account of practical reasons, but Velleman's insistence that it be a mechanism, necessarily imprecisely described, merely demonstrates his loyalty to the conceptual independence thesis and is not a real solution to the normative problem.

If 'what there is most reason to do' were a mechanism describing a cognitive tendency then it could not explain the normative authority of better reasons over weaker reasons. A failure to be moved by better reasons is still only a cognitive misfiring without normative significance if success in reasoning is not success in meeting rational standards. The metaphor of proximity to practical reason's constitutive aim must be unpacked so as to include a normative appeal to reasons for action, but the conceptual independence thesis precludes exactly this sort of unpacking. As we have seen, vicious circularity can be avoided with a substantive, normative conception of reason's aim. Velleman is wrong to suppose that correctness in episodes of practical reasoning requires that the aim of practical reasoning first be established in a descriptive way. The norms governing practical reasoning are internal to the practice of reasoning and cannot be derived from, or fully explained by, a descriptive account of the causal mechanisms underlying the processes of reasoning.

### **Theoretical Reasoning**

Let us now consider Velleman's conception of the aim of theoretical reasoning and whether it has the resources to explain how rational principles bind doxastic agents. Velleman's approach to theoretical reasoning is analogous to his approach to practical reasoning. Here too he appeals to the conceptual independence thesis: in theoretical reasoning the constitutive aim of belief must be understood in terms that are conceptually independent of reasons for believing. Velleman takes the constitutive aim of belief to be fixed by the regulative mechanisms productive of beliefs, the aim itself being a sub-doxastic attitude of acceptance. A cognitive attitude is a belief only if it bears belief's constitutive aim. Again Velleman's strategy is to find that thing at which belief aims and then to assess correctness as a proximity relation to that aim: «Belief aims at the truth in the normative sense only because it aims at the truth descriptively, in the sense that it is constitutively regulated by mechanisms designed to ensure that it is true» (Velleman, 2000a, p. 17). While other cognitive states, such as fantasy and assumption, take their propositional objects as true, belief is unique in taking true with the aim of thereby accepting a truth. Belief is an attitude of accepting as

---

<sup>7</sup> Velleman, 2000a, p.14. He explains that the desire or mechanism must be imprecisely described lest his account require that agents must have a concept of a reason in order to act at all. It is not clear how insisting on ambiguity avoids that problem.



true with the further aim of thereby coming to have a true belief, and thus is inherently truth-tracking (Velleman, 2000b, p. 251).

Velleman takes himself to have established a non-controversial norm of correctness for belief which is internal to the nature of belief itself. Indeed, he claims that belief aims at the truth in the same sense that circulation of the blood aims to supply bodily tissues with nutrients and oxygen (2000a, p. 17). Correctness in believing is simply believing true, as a matter of fact, so that «even an unjustified true belief is correct or right» (Velleman, 200b, p. 277). Velleman's split with the normative is thus very sharp and this is not a trivial split, since belief arguably aims at something more substantial than truth, such as knowledge or the guidance of action. Given the possibility of an ethics of belief the comparison between the aim of belief and the aim of the circulation of the blood is overstated. If belief aims at something like knowledge then it must aim at something normative (substantive but generic) such as what there is most reason to believe. Velleman grants that this view, if true, would subsume his view, but notes that it would fail to account for the intuition that unjustified beliefs may nonetheless be correct beliefs because true. Of course this is an intuition which is not universally shared. In defence of his own position and against the claim that belief aims at something more normatively substantial, Velleman states: «I take it to be a conceptual truth that beliefs are correct when true and incorrect when false: false beliefs are necessarily faulty or mistaken» (2000b, p. 277). What is wrong with false beliefs is not that they may have negative practical consequences, but that they are inherently faulty because they fail to be regulated in the way they were meant to be regulated. False beliefs are «the ones whose regulation has not succeeded in producing the kind of cognitions that it was designed to produce» (2000b, p. 278).

These are controversial claims and do not provide a non-controversial, descriptive foundation on which to base norms of theoretical reasoning. Much hangs on the strength of the intuition that unjustified beliefs may be correct simply because true, that truth is sufficient for correctness. While we may agree that false beliefs are necessarily faulty, we may disagree that true beliefs are necessarily correct. A false belief doesn't represent reality and so it is clearly not getting things right. But whether all true beliefs are correct is an open question because 'correct' is ambiguous between the minimal sense of corresponding with what is as a matter of fact true and the more robust sense of 'correct' as true and appropriately rule-governed. A lucky guess (where taken as true with the aim of thereby accepting a truth) is correct by Velleman's standard even while it lacks an adequate basis in reason. But neither false beliefs nor unjustified true beliefs fail to meet the aim of theoretical reasoning in the same way that bad circulation fails to supply the tissues with oxygen. What is wrong with many such beliefs is that they are not appropriately rule-governed. In such cases something is going wrong, and in order to identify what is going wrong we must vacate the agent's doxastic perspective and introduce a larger perspective from which we can see both that the agent's belief is false and that epistemic standards (evidence, consistency, and so on) speak against it, that is, she has good reason not to believe it. The appeal to the truth or falsity of belief necessarily involves an external analysis (except perhaps with regard to an agent's beliefs about her own mental states).

Velleman's defence of the conceptual independence thesis leaves him with few resources to explain the normative constraints theoretical reasoning imposes on reasoners. It is true, as a matter of fact, that believers tend to give up or revise their beliefs when they are found wanting. If I believe that my keys are in my pocket and my hand fails to detect them, I quite

naturally lose the belief that my keys are in my pocket. As a believer I am (mostly) inclined in this manner. But it is also true that, *ceteris paribus*, I ought to stop believing that my keys are in my pocket if there continues to be no evidence for that proposition. Or, more significantly, I ought to stop believing that the rate of violent crime is rising if sound statistical evidence suggests that it is not, even while sensationalistic journalism leaves me inclined to suppose that violent crime is rampant. It is this latter sort of case which poses problems for the adherent of the conceptual independence thesis. To the question ‘why believe what there is most reason to believe?’ Velleman cannot, he has argued, respond that theoretical reasoning has as its aim believing what there is most reason to believe without violating the conceptual independence thesis.

If the argument of the last section was successful it showed that, at least in the case of practical reasoning, vicious circularity can be avoided without embracing the conceptual independence thesis. In the case of reasoning about belief we find Velleman’s commitment to the conceptual independence thesis leading to equally problematic results. By insisting that the constitutive aim of belief be conceptually independent of reasons for believing he creates important gaps between the processes that make particular cognitive states beliefs, reasons for belief, and epistemic norms governing the practices of believers. In this section I will argue that Velleman’s account of the aim of belief, in order to be plausible, must be interpreted as a norm-laden conception (ie., as substantive and generic) and will suggest that his troubled account of the normative constraints governing believers is again due to his adherence to the conceptual independence thesis. By refusing to hold that reasoning about belief aims at, or is governed by, what there is most reason to believe, Velleman makes beliefs produced by processes which must be insensitive to rational epistemic norms.

Are believers constrained to believe what there is good (or better) reason to believe if, as Velleman suggests, the constitution of belief is via mechanisms designed to ensure its truth? How can epistemic norms arise out facts about cognitive mechanisms? If I don’t care about the truth of a particular topic why not believe whatever I want about it? Recognizing that these normative issues need to be addressed, Velleman distinguishes between epistemic norms as they govern agents in general and norms as they govern believers on a particular topic. We are governed by good or better reasons for belief, whether we like it or not, because we are believers, but not because belief aims at good reasons for believing. Velleman explains that what there is reason to believe ultimately depends upon the inclination that makes one a believer; theoretical reasoning constrains at the level of agency as believer.

Believers, he says, are constituted in part by an inclination to be moved by the constitutive aim of belief. Because reasons for belief depend on our nature as believers, not a particular agent’s inclinations as a believer, the normative authority of reasons is unconditional – it applies whether or not those believers accept the epistemic norms in question. So although indicators of truth such as authoritative historical sources won’t, as a matter of fact, incline some believers to believe that, say, the Holocaust happened, these facts continue to count as reasons for such persons to believe, even in the absence of an inclination making them susceptible to their influence. According to Velleman (2000c, p. 187), the inclination on which theoretical reasons depend is constitutive of belief itself and to that extent indicators of truth, like historical facts, are reasons simply for belief rather than for any person to believe. This appeal to the constitution of agency allows Velleman to claim (2000a, p. 18) that if there is reason to believe  $\phi$  in circumstance C then relevantly situated agents ought to believe  $\phi$  in C:

Indicators of truth count as reason for belief because truth is the aim of belief; but truth's being the aim of belief just consists in the way that the mechanisms of belief are designed to regulate it – which they do by responding to indicators of truth. Ultimately, then, indicators of truth count as reasons for belief because they are the considerations in response to which belief is designed to be regulated. The aim of belief and reasons for belief are fixed simultaneously, the one being determined by the way in which belief is constitutively regulated in response to the other.

This passage includes a subtle equivocation on reasons for belief, a non-normative and a normative version. Because the aim of belief and reasons for belief are fixed simultaneously, and the aim of belief is the non-normative regulative process of taking true with the aim of thereby accepting a truth, what counts as a reason for belief arguably refers to whatever has caused the belief to be generated. On this non-normative reading, indicators of truth amount to reasons for belief because they are appropriately causally related to the truth that (purportedly) non-controversially describes belief's aim. The mechanisms of belief are designed to regulate it by producing true beliefs, so the bits of reality to which these mechanisms respond could be called reasons for belief in virtue of this relation and nothing else. This is a stipulative notion of reason for belief which carries none of the ordinary normative consequences for believers. The appeal of this picture comes from reflecting on perceptual cases where it is plausible to hold that belief is caused – seeing the cat on the mat is what caused me to believe that the cat is on the mat – but it is limited because it seems to have all the problems of the early sense datum view of knowledge and can not support inference. Clearly he cannot have this in mind, but instead a richer inference-guiding notion of what makes something a reason for belief that includes things not present to the senses.

On the normative reading, indicators of truth count as reasons for belief by serving as the kinds of considerations which, investigation reveals, rationally support belief. When Velleman describes reasons for belief as «considerations in response to which belief is designed to be regulated» the concept has a semantic life of its own, one that includes the normative governance of belief and not merely an appeal to a causal link. This is a substantive, generic conception, where 'reasons for belief' picks out a privileged class of cognitive states, a normative class. This more plausible reading, however, does not derive norms by appealing to a non-normative fact about the function of belief. Reasons for belief cannot credibly be thought to be conceptually fixed by mechanisms regulating belief because mechanisms cannot guarantee a connection to truth. In this sense an indicator of truth isn't merely causally connected to the truth of something, and reasons for belief are not non-normative.

Velleman's thin account of belief's function is not robust enough to itself supply an account of good reasons and the authority of good reasons over believers. The constitutive aim of an individual belief is determined by the regulative mechanism (aiming at the truth in order to thereby accept a truth) which produces it, with or without an agent's intentional cooperation. Velleman argues that normative constraints on belief arise at the level of agency and not at the level of belief but we can wonder whether the introduction of normative constraints at this level presupposes a normative account of the aim of belief. To say that indicators of truth like historical facts are reasons for belief is not to make a normative judgement because reasons apply to all those «who are constitutively truth inclined» (Velleman, 2000c, p. 186). An agent will aim at the truth on some particular topic and in doing so be subject to reasons for belief concerning it, and not aim at the truth on another.

Insofar as belief and reasons for belief are fixed simultaneously, belief cannot be conceived of as non-normative. For one who does not adhere to the conceptual independence

thesis, this isn't a problem. In fact, allowing normative considerations at the level of belief itself has important consequences for explaining how believers are bound by epistemic norms generally. Velleman's problematic attempts to explain this introduces norms at the level of agency and can be avoided once a normative relation between beliefs and reasons for belief is maintained.

Consider an example in order to see how Velleman runs into trouble in accounting for the normative authority of reasons for belief. The truth about the age of the Earth fixes what there is reason to believe on this topic. The truth is fixed by the fact, and goodness in belief is getting the facts right. However, indicators of truth about the age of the Earth, of which there are many, are all contingent upon the availability of technological resources, acquaintance with scientific theories, ordinary background knowledge, and so on. What it will be reasonable to believe is then not fixed merely by the truth but by the availability of a variety of forms of expertise to which an agent may be reasonably presumed to defer. Why believe what there is most reason to believe about the age of the Earth rather than believing what suits one's religious convictions?

If reasons for belief are normative and constituted by the substantive, generic aim of what there is good reason to believe, then believing is implicitly governed by norms of good belief. This helps us to see that the problem with the religious fundamentalist's belief that the Earth is about six thousand years old is not primarily that it is false, but that it is based in reasons which don't offer plausible support. Because no one has direct access to the truth of the matter, everyone's beliefs are in the same evidential boat. The only thing that can determine what one ought to believe in this case is the weight of reasons. This is why it makes sense to say that the primary problem with the belief is not that it is false but that it is not grounded in good reasons. The belief that the Earth is ten thousand years old would be nearer to the truth (if the aim of belief is bare truth), but no better as a belief. Because Velleman casts the aim of belief in these minimal terms he cannot make sense of this difference. The belief is not grounded on good evidence and it is unjustified. My claim is not that Velleman has nothing to say about these cases of epistemic failure, but rather that his appeal to governance by the weight of good reasons relies on a normative conception of the aim of belief, from the start.

It might be objected that Velleman can simply introduce epistemic constraints on believers by holding that theoretical reasoning aims at something more than mere correctness, such as knowledge or what there is good reason to believe. Believers are governed by epistemic constraints because belief aims at the truth, and believing what there is most reason to believe is the best way to have true beliefs so that, insofar as reasons for belief are regulated by the mechanism of belief, then agents have good reason to believe whatever there is good reason to believe. So, the objection goes, the norm can be derived from the descriptive fact about the nature of belief. However, the equivocation problem is relevant here. On the non-normative reading, reasons for belief are causal links between truth and the mechanism of belief and so the move from what there is reason to believe to what there is most reason to believe in order to maximize true beliefs is invalid. The sub-doxastic mechanism constitutive of belief is not the same as the practice of being guided by a normative policy of believing what is well-justified in order to maximize true beliefs. The non-normative reading does not support the prudential epistemic policy. The normative reading, arguably does, since if epistemic norms are part of belief itself then epistemic constraints are implicitly constitutive of belief. But if my argument is correct, the normative reading undermines the conceptual

independence thesis and there is no need for the further argument. If the aim of theoretical reasoning were conceived of as tied to what there is most reason to believe then the connection to the normative would come more easily. The conceptual independence thesis demands that the conceptual link between beliefs and what there is reason to believe be severed, and without this connection episodes of believing are set free of their rational moorings.

According to Velleman, if a person is indifferent to reasons for belief about a topic then she is not a believer about that topic. She may instead be a fantasizer, in which case her cognitive attitudes aren't irrational; she doesn't have a failed or faulty belief because she doesn't have a belief. Velleman takes it to be an advantage of his account that indifference to reasons for belief about a topic doesn't imply irrationality, although how this is advantageous isn't spelled out. But in our shared social life indifference to reasons for belief often presents itself as irrationality.

It is clear that a Holocaust denier, for example, is making an epistemic error by believing what goes against the weight of reasons for belief, not by having cognitive attitudes which fail to manifest belief's real constitutive aim. She cannot avoid the normative force of better reasons for believing by having a rich fantasy life and associating only with bigots. Her belief that the Holocaust never happened is both a commitment to the truth of that proposition and a commitment to the propriety of that judgement; in believing she also takes her own doxastic commitments as sufficient for belief. So her belief implicitly sanctions epistemic standards and is thus irreducibly normative. Its being a belief is, in part, constituted by these implicit epistemic norms. Believing is constitutively norm-governed; to be a believer is to be reasons-responsive even in the first person; in believing one takes one's own reasons to be adequate grounds on which to make a judgement. Again, my claim is not that Velleman has nothing to say about the epistemic failures of holocaust deniers, but rather that his appeal to governance by the weight of good reasons can only work on the normative reading of the function of belief.

As we found with practical reasoning, to include the normative dimension in the aim of theoretical reasoning need not lead us to circularity or incompleteness. 'What there is most reason to believe' is substantive in the sense that it describes an aim in terms other than being the object of the enterprise of theoretical reasoning itself. As a standard it has its own semantic life and is not a purely conceptual description of theoretical reasoning which would be either trivially true or circular. 'What there is most reason to believe' identifies a generic standard for success which may be achieved by competent theoretical reasoning but isn't defined in terms of competent theoretical reasoning. The notion of a robust standard or ideal does not require a specification of how it is to be met. Indeed standards of theoretical reasoning are generally such that they surpass any particular conception or implementation of them.

If the aim of belief were non-normative in the way that the circulation of the blood is non-normative then beliefs could be perfectly adequate in their own constitutive terms without satisfying *any* epistemic standards. Of course Velleman is right that in order to be beliefs, beliefs need not be well-formed, nor true. But in order to explain how good reasons ought to be believed over bad reasons (how they are rationally authoritative) a normative conception of the constitutive aim of theoretical reason is required. Construed merely as an inclination to be governed by good reasons, epistemic norms lose their connection to rationality. If reasons for belief themselves depended on a mere inclination which makes us believers, then

a failure to be so inclined would be a statistical anomaly rather than an epistemic error. To see this, consider an albeit unusual case where the epistemic standards and the products of theoretical reasoning come apart.

Velleman offers the example of Ronald Reagan who, it is said, tended to mistake old movie plots for historical fact regularly enough that these cognitive attitudes became mixed in with his beliefs, to the extent that it seems he in fact believed them. (It should be noted that it was only later that it became known that Ronald Reagan had been diagnosed with Alzheimer's; the example had no malicious intent and helped to illuminate a distinction.) Even here Velleman is cautious with charges of irrationality, responding that Reagan believed that he had beliefs about history but really they were fantasies – fantasies and assumptions can be misclassified in with our beliefs. Reagan may have believed that he believed what he was saying but, according to Velleman, he didn't believe what he was saying, he fantasized it. As a fantasy, it does not fall under epistemic norms. Velleman grants that *if* an attitude's being treated as a belief were sufficient for its being a belief then misclassified fantasies would tend to fall under epistemic norms, and this would further imply that they should be discarded or revised so as to conform to the facts. But this is not the case. All that is required is that they be reclassified (2000b, p. 280). That an attitude is treated as a belief, on Velleman's account, is not sufficient for it being a belief. This raises the question of whether in our own doxastic self-regulation we can ever know whether particular states are beliefs and thus subject to regulation.

Velleman's willingness to create an elaborate taxonomy of cognitive states is perhaps a result of his commitment to the claim that the aim of theoretical reasoning must be characterized in terms which are conceptually independent of reasons for belief. But the taxonomy fails to satisfy the entirely plausible thought that our concept of belief is at least partly a functional notion. If a cognition functions as a belief then it is one, and is governed by rational norms for believing. It seems confused to continue to insist that whether or not a propositional attitude is a belief can be decided strictly by appeal to the sub-agential psychological mechanisms which produced it, and not only because we rarely know what the relevant mechanisms are. When an agent claims that *X*, and *X* is false and is nonetheless presented as a matter of fact we will not be satisfied to respond 'you think that you believe *X*, but in fact you fantasize *X*, and only believe that you believe *X*'. If the propositional attitude functions as a belief, it bears the epistemic burdens that beliefs bear, and we believers are entitled to hold it up to scrutiny just as we would hold other beliefs up to scrutiny. The expressed claim is a commitment to the truth of a propositional content and so it is not just a psychological byproduct. Part of what makes an attitude a belief is that it makes a commitment to truth, a commitment which requires us to hold it to epistemic standards. We hold the believer accountable for her truth-taking, not for her poor cognitive housekeeping. In part the regulation of belief is social.

If my cognitive attitude does the work of a belief (it allows me and others to make inferences based on it, incurs commitments and entitlements, and so on) then it should not be denied status as a belief because the mechanisms which produced it aren't belief-producing mechanisms. Perhaps anticipating some of these issues, Velleman writes, in a footnote (2000b, p. 279), that he is considering the view that «an attitude qualifies as a belief partly by virtue of being *treated as* truth-directed.» But this proviso cannot simply be tacked on to his otherwise mechanistic strategy for it would amount to an admission that belief is not just the product of sub-agential mechanisms representing belief's constitutive aim. The problem

Velleman's account raises for the conceptual independence thesis should be clear. If the aim of theoretical reasoning is unpacked in terms conceptually independent of reasons for belief then the normative connection with reasons for belief is severed and an adequate explanation of how rational norms of believing bind agents cannot be supplied. Velleman tries to maintain the conceptual separation between the aim of theoretical reasoning and reasons for belief by asserting that believers are constitutively truth inclined, but the inclination to truth in the minimal sense of correctness, without governance by rules, leaves no room for normative constraints, and an inclination to truth in the more robust sense which includes a notion of appropriate governance by rules is at odds with Velleman's purported hope to establish a non-controversial correctness condition for belief, internal to the nature of belief itself.

### Conclusion

I have examined Velleman's attempt to preserve the attractions of a constitutive account of reasons while avoiding the normative burdens it places on agency. We have seen that Velleman's claim that any account of epistemic norms must meet the independence thesis implicitly relies on the normative character of agency. This reliance in turn violates the demands of the independence thesis. In the case of practical reasoning a substantive but generic normative conception avoids the circularity problem while leaving room for a plausible account of the normative authority of practical reasons. In the case of theoretical reasoning a substantive but generic normative conception leaves room for an account of the normative authority of theoretical reasons. Because the conceptual independence thesis poses problems for understanding how rational norms constrain agents, and circularity can be avoided without endorsing it, it should be abandoned.

I have argued that constitutive views can avoid circularity without adopting the conceptual independence thesis. But this is not an argument against the conceptual independence thesis per se. We have also seen independent reasons for rejecting it. We have seen reasons for thinking that it is a mistake to sever the connection between the practice of reasoning and its aim. Goodness in reasoning cannot adequately be explained by a nearness relation to the constitutive aim of reasoning. For beliefs that are not directly evident, there is simply no way to measure nearness other than by the weight of reasons for belief. The nearness relation is just a metaphor for the aim of good reasons. For Velleman good reasons are those which better manifest the aim of belief itself, but if the aim is mere truth, then we cannot intelligibly unpack the metaphor of nearness to belief's constitutive aim. Indeed, the only way to unpack the metaphor is by way of the normative force of reason. It is not better to believe that the Earth is much older than six thousand years because that belief is nearer to the truth; it is better to believe it because it is more reasonable.

### References

- Brandom, Robert. (1994). *Making it explicit*. (Harvard, MA: Harvard University Press.)
- Carroll, Lewis. (1895). What the tortoise said to Achilles. *Mind*, 4, 278-80.
- Clark, Philip. (2001). Velleman's autonomism. *Ethics*, 111 (3), 580-593.
- Engel, Pascal. (2005). Logical reasons. *Philosophical explorations*, 8(1), 21-38.
- Korsgaard, Christine. (1996). *The sources of normativity*. (Cambridge: Cambridge University Press.)

- O'Hagan, Emer. (2005). Belief, normativity and the constitution of agency. *Philosophical Explorations* 8(1), 39-52.
- Railton, Peter. (1997). On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action. (In G. Cullitty & B. Gaut (Eds.), *Ethics and practical reason* (pp. 53-80). Oxford: Oxford University Press.)
- Velleman, J. David. (1992). The guise of the good. *Nous*, 26, 3-26.
- Velleman, J. David. (2000a). Introduction. (In his *The possibility of practical reason* (pp. 1-31). Oxford: Clarendon Press.)
- Velleman, J. David. (2000b). On the aim of belief. (In his *The possibility of practical reason*, (pp. 244-281). Oxford: Clarendon Press.)
- Velleman, J. David. (2000c). The possibility of practical reason. (In his *The possibility of practical reason* (pp. 170-199). Oxford: Clarendon Press, 2000.)

**Emer O'Hagan**

**Department of Philosophy**

**University of Saskatchewan**

**Saskatoon, Saskatchewan, Canada**

**S7N 5A5**

**[emer.ohagan@usask.ca](mailto:emer.ohagan@usask.ca)**



SORITES (Σωρίτης), ISSN 1135-1349

<http://www.sorites.org>

Issue # 21 — June 2008. Pp. 31-36

Perseverance, Motivation, and Ambition's Debt

Copyright © John Zillmer by and SORITES

## PERSEVERANCE, MOTIVATION, AND AMBITION'S DEBT

John Zillmer

---

It is sometimes thought that ambition is, if not quite a virtue, at least a trait of character the lack of which is a loss; to call someone *unambitious* is to criticize. Determining if ambition is a virtue, or if lack of ambition is a loss, is the aim of this essay.

There are a variety of different methods by which a concept may be analyzed. Let's begin by narrowing the analysis: let us cast aside the sense in which Caesar was said by Brutus to be ambitious. Caesar's death is said to have paid the debt to the sort of ambition that is simply ambition for ambition's sake, raw power disjoined from everything but itself. We can be agreed, I trust, that the sort of ambition that might have some value in a good human life will not be of this sort.

Perhaps a valuable sort of ambition might actually be similar to that of Caesar's in the limited sense of its being good at all is so because it is good in and of itself. Caesar's 'ambition for ambition's sake' is of this type. On this conception, having ambition would be valuable in the same sort of way in which it is valuable to act from duty, or out of concern for the welfare of sentient beings, or for reasons relating to aesthetic value. All these are, in certain manifestations, intrinsically rather than instrumentally good. Thomas Nagel, for one, wrote of the irreducible value of being the sort of person who finishes what she starts.<sup>1</sup> This sort of steadfastness, or perseverance, is of course instrumentally valuable in helping one realize her aims. It may be intrinsically valuable as well; Nagel's laudable words are not only of finishing what one starts but of being the sort of person who does so. Whatever the value of the attainment of the goal, there is value in the character of the person who attains it. Surely Caesar, too, had this quality, and had it to the last; moments before his death, on receiving the plea to repeal the banishment of Publius Cimber, Caesar says that among all men «I do know but one/ that unassailable holds on his rank,/ unshak'd of motion: and that I am he,/ Let me a little show it, even in this,/ That I was constant Cimber should be banish'd,/ And constant do remain to keep him so.» And thereafter Caesar, as was noted, paid his debt.

But this is not clearly ambition at work. Whatever the trait of character underlying the initiation of the order of banishment, Caesar's refusal to be moved by Metellus' plea was

---

<sup>1</sup>. Nagel (1979), pp. 129-130.

grounded in perseverance rather than ambition. While ambition can motivate one to begin a project, perseverance is relevant only to its continuance.

So we've come to see that ambition can in this regard be distinguished from perseverance. But the two concepts are also related: one way in which they are related is that having ambition can be seen to entail the possession of perseverance. A way in which ambition entails perseverance is insofar as one who starts projects without finishing them – that is, insofar as one lacks perseverance – she is also generally said to lack ambition. One recalls A.A. Milne's character Tigger, who is forever going off in search of a new adventure, but quickly gives up after finding that some other escapade tugs on his desires a bit more strongly. Tigger lacks ambition *because* he lacks perseverance, and so ambition demands one be perseverant as well.

To continue to probe this connection, it is notable that ambition, as a trait of character, is a diachronic characteristic of a person (that is, it persists over time). One who is ambitious is not ambitious just for a moment. She may, of course, have been ambitious *then* but is not ambitious *now*; this change in character occurs often enough. Such a change, though, is typically marked by one of at least two features. One can lose her ambition gradually, in much the same way as other long-standing things (such as bridges and friendships) crumble from lack of attention. One might, alternately, lose her ambition by way of some impetus external to ambition itself, an impetus as radical as a conversion ( I take it that Paul was more ambitious than Saul), as ordinary as a 'change of heart' or as sadly common as a depressive episode. It is, though, not in the nature of ambition to be a fleeting trait, and if it does come to end, this end is due to time's ravages or a force stronger than itself. It is the nature of ambition to persist; ambition, one may say, itself perseveres.

Another reason why ambition persists is because it is not satisfiable by the attainment of a goal. In a way, neither is perseverance thus satisfiable. Perseverance is only self-satisfiable – that is, the only way to satisfy one's perseverance is to persevere. But perseverance becomes absurd when the goal toward which one perseveres has been attained. This is not because the attainment of the goal satisfies one's perseverance. If the attainment of a goal toward which one perseveres satisfied one's perseverance, then every goal would be a dual goal; for example, if Brenda had a desire to run a marathon, and so she ran one, persevering through the grueling 26.2 miles, her full answer to the question 'Why are you doing this' would have to be 'In order to satisfy my desire to run a marathon, *and* in order to satisfy my perseverance.' But this is odd. It is odd because perseverance is the means, not the end; perseverance is one of the vehicles with which one carries herself through a marathon. So, although perseverance is not *satisfiable* by persevering, it can be *dismissed* at the end of the task, even if it is dismissed from work without severance pay, so to speak<sup>2</sup>.

If perseverance, then, can be likened to a piece-worker, ambition can be likened to a taskmaster. This dual labor-structure analogy is apt since perseverance helps one achieve goals toward which she is already progressing, or at the least, a goal she has set for herself as something more than a mere hope. Ambition, though, can be the force that motivates one to begin a project; over and above the mere having of a goal (that is, as a 'mere hope'), ambition can motivate a person to make her way toward the having of the goal itself.

---

<sup>2</sup>. Interestingly, 'persevere' and 'severance' are of etymologically distinct heritage; the former being rooted in the words for 'through' and 'harsh', the latter in 'sever' - that is, 'to cut'.

Suppose there is a need for a difficult project to be completed; suppose Andrew is to deliver a package of vaccine to a village over the mountain. Once he is underway in the delivery of the vaccine, perseverance may get him through the project if other motivations are to falter. «I'm tired,» he may say to himself, «and hungry and perhaps I'll be bitten by a venomous snake up ahead and perish here on the mountain, and then no one will benefit from this load of vaccine, and I'll lose my life as an extra bonus. Perhaps I ought just to turn back before things get more difficult.» But if he has perseverance, these thoughts may not be sufficiently forceful to persuade him to turn back. He may, that is, persevere.

Consider, alternately, Thomas, who considered not beginning the journey over the mountain at all. «You want vaccine,» Thomas may have said, «carried over *that* mountain...I don't know; it seems quite a big mountain, and that package looks heavy. Besides, I don't know any of these people in the village, and they might make it through this without the vaccine anyway. Perhaps I'm just not the right person for the job.» But an ambitious person, call him Peter, would not be so likely to be dissuaded by difficulty and the chance of the project's turning out futile. Whether Peter has any predisposition to traversing mountains or helping the sick matters little, since if he is ambitious, he will be likely – more likely than Thomas was, anyway – to agree to undertake the project.

It is not that Peter would necessarily have seen the situation differently, weighing the difficulty less and the chance of fruitfulness more. Though he indeed may do this – and perhaps it is true that ambitious people happen to be more optimistic about the chances of their projects succeeding – optimism is not integral to ambition. We can imagine one who is ambitious but not entirely optimistic about his chances of success; avalanche rescue workers are indeed ambitious, but not, I'll guess, entirely optimistic. But since I am not an expert on the psychology of avalanche rescue workers, and am not ambitious enough to conduct a survey, I'll offer a less empirically-based example. Consider Martha, who is an accomplished mathematician hoping to be awarded the Abel Prize (the equivalent of a Nobel Prize, but in mathematics) for her research on probability, a subject on which she is an expert. If we can assume that since Martha is an expert she must be ambitious (since the unambitious are very unlikely to become experts in mathematics, especially), we can also assume that since she is an expert on probability, that she realizes that the odds that she (among all the mathematicians in the world) be awarded the Abel Prize are not very good. So if she is reasonable, she won't be optimistic about her chances of being awarded the prize (though of course she could still be hopeful of her winning). Ambition, then, need not motivate by helping one to be more optimistic about her chances of success.<sup>3</sup>

Perhaps ambition is a motive that is a sort of catalyst; in chemistry, certain chemical reactions proceed only, or proceed more quickly or more strongly, with the addition of some chemical that is not itself a reactant but helps the reactants themselves to react. The catalytic converter on your car's exhaust system uses the metal platinum as a catalyst to cause nasty polluting exhaust gases to chemically change into less nasty, less polluting carbon dioxide and water; the platinum isn't used up, though, because it isn't itself a reactant in the chemical reaction the exhaust gases undergo. If ambition works analogously, an ambitious person may

---

<sup>3</sup>. There are possible conceptions of ambition similar to the foregoing: perhaps being ambitious enables one to see more clearly the reasons there are for acting, or (these following are likely equivalent in practice) strengthens one's motives or makes her more sensitive to motives as they stand.

have the same reasons as a non-ambitious person, but the ambitious person may act due to the catalytic action of her ambition.

The closest theoretic fit with the catalyst conception might be found in Jonathan Dancy's notion of 'contributory reasons.'<sup>4</sup> «A contributory reason for action,» Dancy writes, « is a feature whose presence makes something of a case for acting, but in such a way that the overall case for doing that action can be improved or strengthened by the addition of a second feature playing a similar role.»<sup>5</sup> So ambition as a contributory motive is an enlightening conception insofar as ambition 'improves or strengthens' other motives, since I may have an existing but insufficient motive to, say, carry the vaccine over the mountain, a motive which may be sufficiently strengthened by the presence of ambition.

There are further intriguing aspects of contributory reasons that inform on the workings of ambition. A feature that supports the performing of a certain action in one context may undermine its performance in another; in Dancy's terms, a strengthening consideration may be an attenuating consideration, or not a consideration at all, in different circumstances. The fact that my passport is valuable to me when I'm traveling is a reason to go to great lengths to avoid its loss; but when my wife falls over the rail of the cruise ship, my motive to save her because she is my wife is not clearly strengthened by the motive to save her because she happens to have my passport in her pocket; the latter motive seems to weaken the force of the former.<sup>6</sup> Ambition has, often, the same effect: a given motive is not always strengthened by one's ambition. Imagine me, to my wife, back on the deck of the ship, «*Of course* I would brave the ocean to save you, but don't think our marriage or our love is enough to motivate me; these things are almost enough, but it was my ambitious nature that ultimately got me into the water.» So ambition does not always contribute in a helpful way.

This is the point in the analysis at which the catalyst understanding of ambition runs aground. Contributory reasons also serve to motivate actions on their own, and so are not analogous to catalysts, which do nothing on their own (qua catalysts, at least). This claim about contributory reasons in Dancy's terms runs like this: a consideration that in one situation is a strengthener (that is, makes other existing reasons stronger reasons than they otherwise would have been) can figure in some other situation as a favoring reason on its own (that is, the reason does not strengthen some other reason but rather functions as the sole reason). My rescuing of my wife could be motivated by the thought that she is my wife, or by the thought that she is my wife along with – strengthened by – the thought that she happens to have my passport in her pocket and boy I don't want to lose that, or (the third option) the rescue could be motivated merely by the possible loss of the passport. Different contexts make each of these motivational stories comprehensible, if not morally laudable.

---

<sup>4</sup>. In, especially, *Ethics Without Principles*, chapters 2 and 3 (Dancy, 2004). A schematic of the account: what we often consider reasons proper Dancy terms *favourers* if they are reasons to  $\phi$  and *disfavourers* if they are reasons not to  $\phi$ . Considerations that make a favourer or disfavourer stronger (that is, make a reason to  $\phi$  more of a reason to  $\phi$ ) are *strengtheners*; *attenuators* do precisely the opposite; neither of these are reasons. Lastly, and also not themselves reasons, *enablers* are features of a situation that make the favourer able to be a favourer; *disablers* effect the disappearance of a favourer.

<sup>5</sup>. P. 15.

<sup>6</sup>. This example is written with Bernard Williams' (1981) 'one thought too many' argument in mind.

Similarly, ambition itself can be a motive for beginning some project. It can be a motive in the same way as the thought of my passport's loss can be a motive to rescue my wife. It can be a motive since in the example of the vaccine-delivery, the ambitious person, Peter, saw the world in the same way as did Thomas (so on these grounds the two had the same reasons for beginning the project), but had an additional reason for taking on the project. This is the reason that actually motivated him: he is ambitious.

\* \* \*

I'd like to return shortly to two central issues we've been addressing. One, it is the nature of ambition to persist in a similar but distinct way in that perseverance persists. The second issue we will return to is the interrelated observations that ambition might well be an independent motive, and also that it might be non-instrumental.

There is one more matter to consider, though, before returning to these issues. It was noted earlier that there are a variety of methods by which to analyze a term. Etymological scrutiny is one of these ways, and while not always philosophically fruitful, it is often curiously enlightening. In accordance with Caesar as one example of the ambitious, it was the Latins who gave English the precursor of ambition: *ambire*. Literally, the word means 'to walk around' in the sense of walking around from person to person, especially to solicit votes (hence the political connotation of 'ambition'). The Latin word 'ambire' also means simply 'circle' or 'to complete a circle', as in walking around simply for the sake of walking around; from sense of the word this comes the uncommonly used term 'ambit'.

At first glance, etymological analysis seems not to enlighten. The former sense of the term, the political sense, refers to the Caesarian sense of ambition that this paper is not directed toward. The latter sense seems irrelevant to the discussion of ambition as a motivator or even as some sort of virtue. If ambition moves one toward a goal, the concept of a circle is quite inadequate to elaborate the account of ambition. A continuing circle is the opposite of progress toward an end state.

This was, we said, one of the distinctions between ambition and perseverance. Perseverance does not itself direct one toward an end state (though it can help one reach an end state that she is already progressing toward). Perseverance, we discovered, aims not at a goal but at itself; perseverance aims to persevere. This was in contrast to ambition, which can motivate one to begin on a path toward a goal that she does not already have. Ambition can even be a primary motivator, so to speak, motivating a person to begin a project that she wouldn't, without ambition, do.

But if this is the case – that ambition can motivate one to begin a project she wouldn't have otherwise begun – what exactly is it that ambition aims at? What, that is, is the goal of someone who acts because of her ambition?

The ambitious person's goal need not be the goal that lies at the end of the project that ambition motivated her to embark upon. Recall the person carrying the vaccine over the mountain, Peter: he, in his ambition, didn't begin the project because he thought it would benefit the people in the village – though he may have indeed thought this, the person who for lack of ambition opted out of the delivery, Thomas, thought this as well, but it didn't motivate him to cross the mountain. Ex hypothesi (since we can suppose that Peter and Thomas are similar in every characteristic except the possession of ambition), the thought of benefit was not motivating, but ambition was.

If the ambitious person's goal, then, need not be the goal that lies at the end of the project, then it seems ambition is more closely allied to perseverance than was initially apparent. Ambition aims at itself in much the same fashion as perseverance aims at persevering.

But recall that perseverance can be dismissed at the end of the project. Perhaps perseverance will be of use in attaining one's next goal, but perseverance will not manifest itself until the next project is begun (since perseverance doesn't motivate the commencement of projects, but only their continuance). Not so with ambition. Ambition cannot be dismissed at the end of the project, for at least two reasons. First, since ambition need not have aimed at the goal of the project itself (in our example, the goal of the project was to get the vaccine over the mountain), ambition will not be satisfied by the attaining of the goal in the way that hunger is satisfied by eating or loneliness by companionship. Since ambition is not thus satisfied, it will, all other things being the same, remain unsatisfied and remain present.

This fact, that ambition is not satisfied by the accomplishment of a goal, has a number of implications. It implies that ambition will tend to persist in a person, to become a character trait; this was mentioned early on, and is readily apparent without the foregoing argumentation. Another implication is that if ambition is not satisfied, and if a person is ambitious, she will pursue projects motivated by ambition. But since these projects can well be instigated because of ambition alone, ambition can be seen, as was perseverance seen, to pursue itself.

This conclusion, that ambition pursues itself, for and of itself, illuminates the notion of ambition as completing a circuit, a circuit that starts and ends with projects motivated by ambition. Another place this circuit brings us to is back to Caesar, back to ambition disjoined from everything but itself, the power that motivates actions not for their own sake but merely for the sake of ambition itself. And if this is what ambition is, it will bring not virtue but only debts, as Caesar's, to be paid.

### Works Cited

Dancy, Jonathan (2004). *Ethics Without Principles*, Clarendon Press.

Nagel, Thomas (1979). *Mortal Questions*, Cambridge UP.

Williams, Bernard (1981) «Persons, Character, and Morality», *Moral Luck*, Cambridge UP, pp. 1-19.

**John Zillmer**

**[zillmerj@gmail.com](mailto:zillmerj@gmail.com)**

SORITES (Σορίτης), ISSN 1135-1349

<http://www.sorites.org>

Issue # 21 — June 2008. Pp. 37-51

Fictional Colors

Copyright © by Dimitria Electra Gatzia and SORITES

## FICTIONAL COLORS

Dimitria Electra Gatzia

---

### 1. Introduction

Things typically have the colors they appear to have. This thesis, known as color realism, seems *prima facie* true. It also seems consistent with the fact that we ordinarily conceptualize colors as perceiver-independent, non-dispositional, intrinsic, qualitative features of physical objects with which (normal) perceivers (in a neutral state of adaptation) are directly acquainted.<sup>1</sup> Nevertheless, color science invariably informs us that this view is incongruous with what we know about the physical world. Although physical surfaces, gases, and volumes may appear colored, vision scientists maintain that no such properties are instantiated (Herring, 1964; Hardin 1993; Nassau, 1983, 1997). Recent empirical evidence further suggests that there are widespread intrasubjective and intersubjective color variations among *normal* subjects—that is, subjects who do not suffer from any color deficiencies (Hardin, 2004). These variations stand in the way of drawing a non-arbitrary distinction between veridical and non-veridical color experiences. For if colors are indeed mind-independent, physical properties, as color realists contend,<sup>2</sup> then not all (phenomenologically) different color experiences of a single object had by normal subjects (under the same viewing conditions in a neutral state of adaptation) can be veridical. But how can we determine which of these color experiences are indeed veridical? Since any consideration that could arise in support of the veridicality of the color experience of one subject could be matched by considerations in favor of the color experience of the other, there are no non-arbitrary ways of determining which of the color experiences of normal subjects are veridical.<sup>3</sup>

---

<sup>1</sup>. 'Acquainted' is to be understood in terms of having awareness, not being revealed, since many realists deny that the nature of the colors is revealed in color experience; the most the realist is committed to is that colors are present in color experience.

<sup>2</sup>. See, for example, Tye (2000) or Hilbert & Byrne (2003).

<sup>3</sup>. I argue for this claim in an unpublished paper, which I presented to the Pacific APA in 2008. See also Tye (2006), and (2006a). For replies to Tye see Cohen *et al.* (2006) and (2007). I argue elsewhere that the individual variability problem threatens both realism and subjectivism about color, see Gatzia (2007).

Color realists have typically attempted to solve this (and other related) problem(s) by attempting to revise our ordinary conception of color in order for color concepts to successfully denote physical properties, which can then be identified with particular colors.<sup>4</sup> However, if there is no error-free revisionist account, color realism cannot be true.<sup>5</sup> It follows that the discourse that commits us to these properties—and which is both logically coherent and useful in the sense that it plays an important role in many aspects of our epistemological, social, and personal lives—seems to be in jeopardy. Johnston (1997), for example, argues that rejecting (all revisionist versions of) color realism will jeopardize the entire subject matter.

In what follows, I shall argue that the rejection of color realism need not seal the fate of our ordinary color discourse. I will not argue for the claim that realism is false<sup>6</sup> (partly because this is beyond the scope of this paper and partly because it is not pertinent to my proposal).<sup>7</sup> I will rather propose an alternative to it: an account that allows us to preserve our ordinary color discourse without having to commit to philosophically problematic properties (§ 2-4). I will then discuss some potential worries for my proposal and offer some plausible responses (§ 5).

## 2. Color Fictionalism: color ascription in the absence of colors

Color fictionalism is a species of error theory. An error theory about color can be understood as a conjunction of a *conceptual* claim and an *ontological* claim.<sup>8</sup> According to the conceptual claim, ordinary color discourse genuinely purports to describe properties that have certain features. As I mentioned earlier, they are perceiver-independent, intrinsic, non-dispositional properties of physical objects with which (normal) perceivers (in a neutral state of adaptation) are directly acquainted. According to the ontological claim, there are no properties like *that*. Color fictionalists acknowledge that color statements are apt for truth or falsity, but maintain that they are systematically false when taken at face value. However, unlike eliminativists,<sup>9</sup> who hold that our ordinary color discourse and its peculiar commitments should be abandoned<sup>10</sup> (perhaps because they think that eventually it will be

---

<sup>4</sup>. Johnston (1997), McLaughlin (2003), and Cohen (2000) are among those who defend a variety of distinct revisionist accounts.

<sup>5</sup>. For various arguments for the claim that revisionist accounts are flawed see Barry Maund (1995), (2006), and (2006a), Matthen (1999), Arstila (2005), and Gatzia (2007).

<sup>6</sup>. I argue for this claim in Gatzia (2007).

<sup>7</sup>. My account will be useful even if it turns out that either color properties are instantiated (in the actual world) but are radically different from what we thought they were (and it seems this much might be true) or they were the kinds of properties our ordinary color concepts denote but failed to be instantiated in the actual world. My proposal allows us to continue to talk *as if* there are colors in both cases.

<sup>8</sup>. Here I am following Mackie (1977), Van Fraassen (1980), and Maund (1995).

<sup>9</sup>. As I shall use it, the term 'eliminativist' denotes the elimination of a target discourse upon discovering that that the properties in question that give rise to it do not exist; not merely to denote the elimination of such properties as the term has often been used.

<sup>10</sup>. I think everyone would agree that eliminativism is inadequate since it proposes that we should jettison a discourse that is both coherent and useful. Joyce (2006) offers an argument against moral eliminativism, which, I think, can also be applied to color eliminativism. The argument in a nutshell is that if the eliminativist is at a



shown to be incompatible with scientific findings<sup>11</sup>), color fictionalists recommend that we continue using the existing discourse even though it is flawed.

The version of color fictionalism I propose is prescriptive in nature.<sup>12</sup> Prescriptive fictionalists hold that the discourse «entails or embodies a theory that is false», but prescribe that we should «carry on employing the discourse, at least in many contexts, as if this were not the case» (Joyce, 2001: 185). Carrying on employing the discourse as if it were true requires replacing assertion with the act of make-believe, which amounts to pretending to assert that things are colored (see also § 2). By revising our practices in this way, we can continue using our ordinary color discourse as we have thus far not because it is true, but because it can be derivable from truths about the qualitative aspect of color experience together with truths about the external world and the false, non-negotiable<sup>13</sup> propositions.<sup>14</sup> More specifically, the proposed account consists of three components: the *base discourse*, the *fiction*, and the *bridge laws*<sup>15</sup> that connect the former to the latter (c.f. Nolan et al., 2005).

---

disadvantage regardless of whether she chooses to preserve or discard color discourse: if she chooses to preserve it, she will be asking us to continue believing untruths; if she chooses to discard it, she will be depriving us of a useful discourse.

<sup>11</sup>. Other reasons for deciding to jettison color (or any other flawed) discourse may involve taking non-referring terms to be meaningless or thinking that it has little, if any, utility.

<sup>12</sup>. Descriptive fictionalists, by comparison, hold that the target discourse is already treated as a fiction, and as such it is not strictly speaking false. According to them, although ordinary people seem to be expressing propositions that commit them to the existence of color properties when they employ the discourse in question, in actuality they engage in some kind of pretense. To see this let us suppose that an ordinary person utters the following statement S:

S: «Lemons are yellow»

The descriptive color fictionalist would say that when the ordinary person utters statements like S, he or she is saying something true, but not because there *are* colors. Rather, because S is short for P\*:

P\*: «In the fiction, lemons are yellow»

So, although our ordinary color discourse is assertoric, ordinary people do not have false metaphysical beliefs about the colors because they are merely pretending that objects are colored (as the hidden operator 'in the fiction' indicates). The descriptive fictionalist is thus merely describing our practices; she is not proposing that we change them. See also Gatzia (2007).

<sup>13</sup>. I am using Joyce's (2001) conception of non-negotiable propositions here. Non-negotiable are those propositions that must be satisfied if something is to be a color property. By contrast, propositions that we could reject without having to deny that something is a color property are called 'negotiable'. Joyce (2001) uses what he calls the 'translation test' to determine which propositions in a give discourse are negotiable and which are not: whether a proposition is negotiable or non-negotiable depends on whether it plays a determinate role in deciding whether the translation goes through. To see this, consider a community of non-English speakers who seem to have a concept that appears rather like our concept of color, call it 'chroma'. Suppose further that we find that they do not endorse one, or perhaps more, of the propositions we accept. If this is sufficient for rejecting that «chroma» should be translated into «color», then these propositions are *non-negotiable*. But if the fact that they do not endorse one, or more, of those propositions is insufficient for rejecting that the translation goes through, then they are *negotiable*. Johnston (1997) makes a similar distinction when he talks about 'core' and 'peripheral' beliefs.

<sup>14</sup>. Assuming that the fiction is incomplete, the rest of the ordinary color discourse will be true-according-to-the-fiction, where *according to the fiction* is to be understood in terms of derivability in the above way. I am grateful to Mark Heller for valuable comments on this issue.

<sup>15</sup>. The term 'bridge laws' often brings to mind philosophy of science, and in particular Carnap's reductive models. However, my proposal is not to be understood as a reduction. All I mean by 'bridge laws' is that there is a way to connect the false utterances contained in the fiction with the base discourse in order to determine which color

The *base discourse* is the unproblematic and literally interpreted part of our discourse—that is, it must not contain any positive claims about the colors although it could contain some color terminology. I suggest that it should include the *Dual-Process theory*<sup>16</sup> of color vision, which consists of two stages. The first is the *trichromatic-process* stage, which pertains to the activity of the three types of photoreceptors, i.e., cones, each of which have different peak sensitivities. The second is the *opponent-process* stage, which pertains to the photoreceptors linked together to form three opposing channels: blue/yellow, red/green, and black/white. Both stages are essential in explaining what is known about human color vision. The trichromatic-process stage allows us to explain phenomena at the photoreceptor level while the opponent-process stage allows us to explain phenomena that result from the neural interconnection among photoreceptor outputs (Hardin, 1993; Palmer, 1999). The dual-process theory thus allows us to explain a variety of phenomena pertaining to visual perception, including the fact that the entire visual spectrum can be produced by combining only three wavelengths, the appearance of after-images, and so forth.<sup>17</sup> The base discourse must also include color experiences and color categories. Our color categories are far less precise than our individual color experiences, which accounts for the fact that we are able to effectively communicate with others despite the fact that our individual color experiences can be, and often are, significantly different. Lastly, we must allow that the base discourse contains some color claims since we want to be able to say, for example, that *literally speaking* «No red objects exist.» A restriction to positive statements is, of course, needed since in most vocabularies it will be possible to formulate both a sentence and its negation, and the fictionalist need not suggest that both of these are false (Nolan et al., 2005). Claims that cannot be part of the base discourse because they are literally false are of the form ‘a is red’, ‘b is green’, and so on.

The *fiction* will contain the false theory of color we ordinarily accept (namely, that physical objects are colored) as well as all the positive color claims. The fiction will be linked to the base discourse via the *bridge laws*, which allow us to «go from information about the world and the objects in it to conclusions about the same subject matter, taking a detour through the fiction» (Nolan et al., 2005: 313). The following bridge laws are proposed as the first approximation of complex laws that will connect the fictional discourse to the base discourse and back:<sup>18</sup>

(BL1) (In the fiction) bananas are yellow if and only if (in typical conditions) bananas would cause the B/Y channel of an (human) observer to be in the state of excitation producing experiences that ordinary people would classify under the color category ‘yellow’.

---

utterances are fictionally true (or appropriate) and which are not. Since reduction requires that we go from truths to truths, my proposal cannot be construed as a reductive one.

<sup>16</sup>. The term is taken from Palmer (1999), p. 110. Notice that even color realists seem to recognize that if we are to solve the problem of color, we must appeal to this theory. See, for example, Hilbert and Byrne (2003) and McLaughlin (2003).

<sup>17</sup>. Other future findings that help explain the various idiosyncrasies of our visual system can also be included in the base discourse.

<sup>18</sup>. These laws are not necessary but rather contingent since things could be different in other possible worlds.

(BL2) (In the fiction) sapphire is blue if and only if (in typical conditions) sapphire would cause the B/Y channel of an (human) observer to be in a state of inhibition producing experiences that ordinary people would classify under the color category ‘blue’.

(BL3) (In the fiction) tomatoes are red if and only if (in typical conditions) tomatoes would cause the R/G channel of an (human) observer to be in a state of excitation producing experiences that ordinary people would classify under the color category ‘red’.

(BL4) (In the fiction) grass is green if and only if grass would cause (in typical conditions) the R/G channel of an (human) observer to be in a state of inhibition producing experiences that ordinary people would classify under the color category ‘green’.

It is tacitly assumed here that the observers are in a neutral state of adaptation. Typical conditions are to be understood as situational in the sense that they depend on our purposes and interests.<sup>19</sup> This is consistent with the fact that assigning colors to objects requires that we specify our «particular interest and purpose» (Hardin, 1993: 81). For, as Hardin (1993: 81) cautions,

[g]iven a particular observer in a particular adaptation state and a particular standard condition, a color can be assigned to an object as precisely as the observer’s perceptual condition warrants, but we cannot expect the assignment to remain the same when the set of conditions or the observer’s adaptational state is changed.

This is particularly troublesome for dispositional accounts of color that rely on ‘standard observers’ and ‘standard conditions’ (c.f. McLaughlin, 2003). The only way to avoid this problem is to relativize colors to observers and circumstances (c.f. Cohen, 2003 and 2004). This move, however, fails since, among other things, it entails that *all* color experiences are veridical (Gatzia, 2007).

Color fictionalism avoids such problems since it maintains that there are pragmatic conditions (which go beyond the semantics of color discourse) that allow a range of normal observers and typical conditions depending on the purpose of the procedures. For example, the fact that the appearance of contrast colors depends on their immediate surroundings presents serious difficulties for many realists since they must incorporate the objects’ immediate surroundings in their analysis.<sup>20</sup> Yet, from a purely phenomenological point of view, contrast colors do not differ from the rest of the colors. Color fictionalism honors this intuition since the dual-process theory can treat color experiences uniformly by reference to mechanisms operating across space. The bridge laws are thus in place to allow us to go from literal truths to fictional truths and back—they are not intended as an analysis of color

---

<sup>19</sup>. Assumptions about adaptation, specifications as to what counts as typical conditions, etc., could be easily incorporated into the bridge laws.

<sup>20</sup>. Contrast colors, i.e., brown, olive, black, white, navy blue, etc., differ from other colors, like red, blue, etc., in that their appearance is depended on the background. Realists like Hilbert and Byrne who want to identify colors with surface spectral reflectances (SSR) cannot provide a unified account because the appearance of the contrast colors depends not only on the object’s SSR but also on its surroundings. The only reason that a surface looks, say, brown rather than orange is that the light that reaches the eye is a function of the surface reflectance of the object, its surrounding surfaces, and various other variables like illumination, etc.

properties. These laws also allow us to determine which utterances are *appropriate* and which are *inappropriate*. (I discuss these notions further in § 4.)

Talk of *real* and *apparent* colors, as used in ordinary speech, can also find its place in typical circumstances. For example, we could pretend to distinguish between the apparent colors of an object, say, a banana looking blue through a filter, and its real color, i.e., the colors it seems to have in typical circumstances. However, since the bridge laws express the sloppiness of our ordinary color perception, it is to be expected that they can, and often will, be sloppy. In addition, it is to be expected that they can, and occasionally will, break down, so long as they do not do so most of the time.

To be sure, the bridge laws are not proposed as an analysis of the nature of the colors. They cannot be thought of as reductions since reductions can only take us from truths to truths. Therefore, color fictionalism must not be confused with dispositional accounts of color. Color fictionalism is an approach towards preserving our ordinary color discourse upon determining that it is flawed. This explains why, unlike color realism, color fictionalism can tolerate the sloppiness or the occasional breaking down of the bridge laws. It also explains why, unlike color realism, color fictionalism can maintain that ‘normal observers’ and ‘typical circumstances’ are relative to our particular interests and purposes.

### 3. What does the pretense amount to?

I said earlier that fictive claims should be understood as an act of make-believe.<sup>21</sup> Claims made as an act of make-believe have the same content whether they are «used as part of a fairy tale or to foolishly assert something false» (Joyce, 2005: 293). When we utter such sentences as an act of make-believe, it is not the content of the target utterance that changes but rather the *force* with which it is uttered. This is the point Peter Geach makes in his famous paper titled «Assertion» when he writes that a «thought may have just the same content whether you assent to its truth or not; a proposition may occur in discourse now asserted, now unasserted, and yet be recognizably the same proposition» (Geach, 1965: 449). Take, for example, the proposition «Apples are red.» Whether it is uttered with or without assertoric force, its meaning remains unchanged; what changes is the *force* with which it is uttered.<sup>22</sup> Since the presence or absence of assertoric force doesn't affect its content,<sup>23</sup> fictionalists can maintain that there is no relevant difference between uttering propositions as an act of make-believe or asserting them. This ensures that valid arguments remain valid when premises are uttered as an act of make-believe. To see this consider the following example:

- (P1): Blueberries are purple
- (P2): My sorbet is made of blueberries
- (C): Therefore, my sorbet is purple

---

<sup>21</sup>. Here I have in mind the conception of fiction discussed in Currie (1990) and Joyce (2001).

<sup>22</sup>. See also Currie (1986) and Joyce (2005).

<sup>23</sup>. Geach (1965), also points out that prefixing statements with «It is true that...» or «There exists an A» will not help to change the meaning of the proposition they express although it may give them assertoric *force*. It is not clear to me whether in fictional discourse prefixing statements with «It is true that...» or «There exists an A» or even «Yes, I am asserting *p*» would give them assertoric force since it might turn out that this is just another part of the fiction.

We have no difficulty saying that this argument is valid when (P1), (P2), and (C) are asserted. But for color fictionalists, who take them to be uttered as an act of make-believe, (P1), and (C) are *not* assertions. Does this matter? Not at all. When one asserts (P1), one presents it as something that one believes. But when one utters it as an act of make-believe, one does not present it as something one believes, nor is one saying that others should believe it; one simply pretends to assert that some things are colored knowing fully well that nothing is colored. As Joyce (2001) rightly argues, the act of make-believe differs from self-deception. When one utters *p* as an act of make-believe, one knows that *p* is false but pretends that it is true.<sup>24</sup> Since the content of the propositions expressed by (P1), (P2), and (C) remain the same whether they are asserted or unasserted, validity is preserved.

#### 4. What is it for something to be fictionally true?

It seems intuitively true that although some color attributions are correct others are not. Any theory of color must be able to preserve this intuition. But since color fictionalism says that *all* color attributions are false, the only way it can satisfy this condition is to distinguish between color attributions that are fictionally true and those that are fictionally false (although they are all literally false). Take, for example, the statement «Elephants are pink». Intuitively, color fictionalists should be able to say that this statement is incorrect while the statement «Elephants are gray» is correct. Color fictionalists can satisfy this requirement by replacing the notion of *veridicality*, which is typically used by the color realist, with the notion of *appropriateness*. Accordingly, color attributions are false but can be appropriately uttered in ordinary contexts because they are useful. This can be done with the help of the bridge laws (as described above).

Since the bridge laws connect the fiction to the base discourse and *vice versa*, color fictionalists can use them to determine which utterances are appropriate and which are not. Accordingly, utterances that are consistent with the bridge laws can be said to be appropriate while utterances that violate them can be said to be inappropriate. So, although there are no yellow objects, it is appropriate to utter «Elephants are gray» or «Bananas are yellow» but inappropriate to utter «Elephants are pink» or «Bananas are blue.»<sup>25</sup> This is due to the fact that the bridge laws connect the fictional discourse with the base discourse, which we take to be literally true. To see this consider once again (BL1):

(BL1) (In the fiction) bananas are yellow if and only if (in typical conditions) bananas cause the B/Y channel to be in the state of excitation producing an experience that ordinary people would classify under the color category ‘yellow’.

According to (BL1) in order for the statement «Bananas are yellow» to be an appropriate utterance certain conditions in the world must obtain. Assuming that the subject’s visual system has the appropriate types of cones, the B/Y channel of her visual system must be in the state of excitation producing an experience that ordinary people would classify under the color category ‘yellow’. Thus, if the right hand side of (BL1) obtains, it is inappropriate to say that bananas are blue. The same can be done with the rest of the utterances. In general, it can be said that if the right hand side of (BL1) obtains, it is appropriate to say that objects

---

<sup>24</sup>. See also Currie (1990).

<sup>25</sup>. It is also appropriate to say that some bananas are red since some bananas would cause the R/G channel to be in a state of excitation producing experiences that we would classify as ‘red’.

are yellow. In addition, the color fictionalist can determine what is appropriate for subjects that have less than two (e.g., dichromats) and more than three types of cones (e.g., tetrachromats) by creating bridge laws that reflect their distinct constitution. Thus, although color fictionalists deny that there are colored objects, they can preserve our ordinary color discourse because they believe that the appropriateness of an utterance comes apart from the truth of its literal interpretation.

## 5. General Worries

Thus far I have outlined a theory that will allow us to continue using our ordinary color discourse even if it turns out that nothing is colored. In what follows, I discuss some potential worries for this proposal and offer some plausible responses.

### 5.1. If there are no colors, how do we acquire color concepts?

Jonathan Ellis (2005) argues that an error theory about color is inadequate because the error theorist cannot explain how we acquire color concepts. More precisely, he argues that «[o]n no plausible account of propositional content can an error theorist explain how we acquire colour concepts» (Ellis, 2005: 55). Although there are many views of propositional content that account for the «normativity of concept possession», Ellis argues, none is available to the error theorist. To see this let us consider the following sentence:

S: «The banana is yellow»

Let us assume that S expresses the following Russellian proposition:<sup>26</sup>

P: <Banana, yellowness>

Ellis argues that since the error theorists claim that there are no colors, nothing instantiates yellowness. Hence, the error theorist must say that S expresses P\* below rather than P above:

P\*: <Banana,  $\emptyset$ >

Notice that in P\* the placeholder for property has an empty extension, i.e.,  $\emptyset$ , which indicates that ‘yellow’ fails to denote. Ellis thus argues that the error theorist cannot explain how we come to acquire color concepts using *any account of propositional content*. The proviso (marked in italics) is very important in understanding Ellis’ argument. His claim is *not* that the error theorist cannot provide an explanation concerning color concept formation, but rather that she cannot provide an explanation using some account of propositional content. It is this proviso that allows him to conclude that error theory about color is unattainable.

However, Ellis’ inference is problematic since the error theorist could deny that sentences express propositions (or that color contents are necessarily conceptual). More importantly, Ellis’ argument is not only applicable to error theory about color but to *every* error theory. The error theorist about morality, for example, would face the same problem were she attempting to explain how we acquire moral concepts like ‘good’ using some account of propositional content. Similarly, the error theorist about Euclidean space would find it difficult to explain how we acquire concepts like ‘parallel lines.’ The same can be said about a host of other concepts, including ‘bitter’, ‘solid’, and so on. It thus appears that explaining how we acquire concepts that have empty extensions is not a problem unique to the error theorist about color. This gives rise to the following question: if no account of propositional

---

<sup>26</sup>. The analysis will be similar for any other type of proposition.

content can accommodate concepts that have empty extensions, why think that the problem lies with the error theory about a subject matter? Why not think instead that the problem is the result of the limitations of these propositional accounts?

Ellis attempts to deny that his argument could be applied to *every* error theory, although he admits that it could be generalized to error theories «concerning solidity...or goodness» (Ellis, 2005: 70). He argues that a «similar argument would not be effective, for instance, against error theories concerning judgments in which the concept such as *witch* is applied to someone» (ibid). This is because «an error theorist about witches has an explanation available to her that the error theorist about colour (solidity or morality) does not. In the case of the concept *witch*, it is natural to suppose that we form the concept through concatenation, i.e., by constructing it from other concepts that we already possess» (Ellis, 2005: 69). However, this is problematic since Ellis' explanation of how we come to form concepts such as *witch* does not tell us how we come to form concepts that cannot presumably be formed through concatenation.<sup>27</sup> In addition, it is unclear why Ellis thinks that unless error theorists can provide an explanation that is consistent with some plausible account of propositional content, error theory is false, especially when other explanations can be available to them. An error theorist about color, for example, could argue that color concepts are acquired from color appearances. We acquire the concept *red*, for example, due to the fact that certain objects *look red to us* even if no objects actually have the properties they appear to have. This explanation is consistent with the fact that we learn about the colors of things by first pointing to objects and then uttering the appropriate color terms. This suggests that having the relevant color experience is not only prior but also essential to the formation of color concepts. Color properties need not feature in our explanations since color appearances are explainable in purely subjective terms. (At the same time, since color appearances suggest that colors are objective, intrinsic, non-relational properties of the surfaces of objects, it makes sense to ask whether anything has these properties and to conduct experiments to provide answers to this question.) But error theorists could also argue that certain concepts are not acquired—this is not to say that they are innate but rather they are the result of some evolutionary process. Joyce (2001), for example, argues that moral concepts like «requirement» and «forbidden» are not acquired but are rather the product of natural selection. Perhaps the same can be said about color concepts, especially since we now know that the visual system of our ancestors differed dramatically from ours in that they had fewer than three types of cones and hence fewer channels.<sup>28</sup> To reject such explanations on the basis that they are independent of any account of propositional content seems gratuitous.<sup>29</sup> What Ellis ought to say is that all plausible accounts of propositional content are limited in that they cannot account for all our concept formation, not that it is impossible to acquire concepts that have empty extensions. If this is right, it follows that this is not a problem for error theorists but rather for any account of propositional content which fails to explain how we succeed in communicating about, say, *redness*, *bitterness*, *goodness*, and a whole host of other concepts whose acquisition cannot be explained by that account of propositional content.

---

<sup>27</sup>. Examples can be drawn from particles of theoretical physics such as the Higgs Boson.

<sup>28</sup>. On the basis of genetic evidence, scientists also believe that some human females have four types of cones.

<sup>29</sup>. It seems to me that Ellis rejects this because he assumes a strong notion of externalism about content that one might be inclined to reject. See also Tyler Burge's «Individualism and the Mental», *Midwest Studies in Philosophy* 4: 73-121, 1979.

## 5.2. Is color fictionalism preferable to color realism?

Color fictionalism is preferable to color realism for a variety of reasons. Firstly, since color realists take colors to be physical properties of objects, they must be able to distinguish between color experiences that are veridical and those that are illusory. However, drawing a meaningful distinction between veridical and non-veridical color experiences is not as easy as it might seem. The underlying assumption behind color realism is that the primary function of color vision is to detect colors. If a realist denies this, she «eliminates motivation for thinking that whatever objective property we have identified with color is color» (Hilbert, 1992: 9). However, evidence from psychophysics, physiology, and perception support the view that the primary function of color vision is to discriminate forms and objects, not to detect the colors (Werner and Webster, 2002; Gouras and Zrenner, 1981). What is important in discriminating forms is essentially color differences, i.e., their relations in a given scene. The absolute values of the colors do not play any role in making such discriminations. It thus seems unreasonable to let the distinction between illusory and non-illusory colors turn on whether the colors are veridically represented. This is a distinction that has to be made within the domain of perception, not by reference to the representation of physical properties. How it is to be made is thus a pragmatic question to which color fictionalists can provide an answer.

Secondly, color fictionalism is consistent with our color language. One important point that sometimes goes unnoticed in discussions about illusory color experiences is that utterances describing such experiences have the same semantic structure as utterances describing what color realists take to be veridical color representations. To see this consider the following examples:

- (1) «I see a red apple»
- (2) «I see a pink elephant»
- (3) «I see a blue circle on the white wall»

(1) is typically uttered when one has an experience of a physical object, in this case an apple; (2) could be uttered by someone hallucinating an object; and (3) might be used by someone having an after-image of an object. A color fictionalist can include utterances like (1)-(3) in her discourse since she can explain illusions and after-images by reference to the dual-process theory. She can accept that in ordinary contexts we can talk about hallucinatory objects or afterimages as having colors (c.f. Maund, 1995). For example, she can say that it is appropriate to utter «I see a pink elephant» or «I have a blue after-image that moves with my eyes» and so on. Or we can ask questions such as «What color is the after-image you are seeing?» or «What color is the elephant you are hallucinating?» and so on. After all, as (2) and (3) suggest, we do talk about them in this way. Color realists have to explain why our ordinary color discourse fails to capture the differences between what they take to be veridical color experiences and hallucinatory color experiences or after-images.<sup>30</sup> Unlike color realism, color fictionalism has the open-question advantage. Namely, it is an open question whether utterances regarding hallucinatory or illusory color experiences should be distinguished from utterances pertaining to correct color experiences.

---

<sup>30</sup>. Harman attempts to explain the difference by arguing that the use of the term 'see' is ambiguous. See «The Intrinsic Quality of Experience», *Philosophical Perspectives*, 1990, vol. 4, pp. 31-52.



Thirdly, color realists have to give up the idea of color property *simpliciter* since the appearance of some colors depends on their immediate surroundings. These colors are known as *contrast colors* and include black, white, brown, navy blue, and olive. A surface that looks brown in a given surrounding, for example, can look orange or yellow upon eliminating its immediate surrounding, i.e., under the aperture mode of viewing. Since contrast colors arise only by the contrast of the surrounding colors, it follows that removing the contrast will deprive us of such colors. This presents a serious problem for color realists since they cannot identify brownness (or any other contrast color) with a perceiver-independent physical property of objects; the reason these objects appear brown and not orange is that the light that reaches the eye is a function of, among other things, the surface reflectance of the object, its surrounding surfaces, and various other variables like illumination. Thus, color realists are forced to reject the claim that contrast colors are non-illusory.<sup>31</sup> But this comes at a high price since they have to reject a whole array of colors. It could be argued that color fictionalists are in a worse position since they have to reject the *entire* array of colors. However, this is not a problem for color fictionalists because (a) their theory is consistent with the facts about the world and (b) they are able to preserve our color discourse without having to inflate their ontology.

### 5.3. The Standard Challenge: is it a threat to color fictionalism?

The standard challenge (which is a general version of the Quine-Putnam indispensability argument and targets primarily mathematical fictionalism) is a point about explanation.<sup>32</sup> It purports to establish the truth of the claim that there are Fs on the basis that the F-theory is explanatory indispensable. In general, we tend to think that if things look just as they would look if they were colored, then one explanation is that they are colored. Prescriptive fictionalists, however, deny this explanation while continuing to rely heavily on the color theory that they say is false (Szabó, 2001). Thus, they have no satisfying answer to the standard objection.

There are at least two possible responses to this objection. The first is to reject the cogency of the argument. Many philosophers have done just that.<sup>33</sup> The second is to deny that the target theory is indispensable. Field (1980), for example, argues that contrary to appearances, mathematical theories do not have to be true in order to be useful in applications;

---

<sup>31</sup>. Tye (2000), for example, claims that «our ordinary experiences of color place (many) object colors on the surfaces of objects independently of what is going on elsewhere in the surrounding...We experience the redness of a ripe tomato as not involving anything away from the facing surface of the tomato as being a *local* feature of that surface...To take a relational view of color is to repudiate this common sense fact.» (153) But this gets him into trouble since he goes on to falsely claim that contrast colors *can* be experienced without their immediate surroundings.

<sup>32</sup>. See Szabó (2001). This argument is widely known as the 'indispensability argument'. Note that Szabó is not targeting *color* fictionalism *per se*, but his argument could be applied to some versions of color fictionalism, albeit not mine, or so I argue.

<sup>33</sup>. Philip Kitcher (1984) argues that the argument does not show why mathematics is indispensable (see *The Nature of Mathematical Knowledge*, New York: Oxford Press); Penelope Maddy (1992) denies that we ought to have ontological commitments to *all* entities that are indispensable to scientific theories (see «Indispensability and Practice», *Journal of Philosophy*, 89:6: 275-289); and Elliot Sober (1993) argues that mathematics does not receive confirmation from empirical evidence since it is employed by every scientific theory (see «Mathematics and Indispensability», *Philosophical Review*, 102:1: 35-57).

they simply need to be *conservative*.<sup>34</sup> They are useful because they simplify calculations. And since their utility is merely pragmatic,<sup>35</sup> Field argues that it does *not* follow from the fact that mathematical theories are useful that they are either true or indispensable. The same argument can be made, perhaps with greater plausibility, about the colors. It can be argued that the (false) color theory fictionalists employ is dispensable because every phenomenon that can be explained by reference to it can also be explained without reference to it. Nonetheless, color fictionalists recommend that we continue using it because it is pragmatically useful. Thus, the standard challenge poses no threat to color fictionalism.

#### 5.4. Can the color fictionalist avoid global fictionalism?

The color fictionalist proposes that we do away with colors but embrace them as a fiction. However, if we choose to be fictionalists about colors what is to stop us from embracing global fictionalism? To put it another way, can one maintain fictionalism about color properties while avoiding fictionalism about all other properties? The answer is affirmative for the following reasons.

Firstly, there are important scientific differences between colors and physical properties such as being square or having a negative charge.<sup>36</sup> For example, the latter are explanatory—they can explain a variety of physical phenomena. By contrast, any phenomenon that can be explained by appealing to the colors can also be explained without appealing to them. Secondly, physical properties are basic to the causality of the world in the way colors are not. They can be observed and measured by interacting with the object in multiple ways. As such, they are properties that have a role in the physical world in the way colors do not—particularly because their existence *is* verifiable through experimentation. More specifically, there are methods independent of our vision for measuring the amount of negative charge in a body or an object's shape. But there are no such independent methods in the case of colors. They have to be defined exclusively in terms of (human) experiences. These are some of the reasons the colors are not scientifically respectable properties—a fact that even color realists are willing to admit.<sup>37</sup> Color fictionalists maintain that although colors are not instantiated, our color discourse is nonetheless pragmatically useful. It is in this respect that color discourse differs from other discourses that are also flawed, e.g., the phlogiston discourse. Color fictionalists can thus consistently maintain that we should be fictionalists

---

<sup>34</sup>. This roughly means that no consequences that follow from mathematics would fail to follow from a nominalistic scientific theory.

<sup>35</sup>. Maddy (1997) also aims to undermine the plausibility of the first premise of the Quine-Putnam argument by showing that confirmational holism should be rejected. See «Indispensability and Practice», *Journal of Philosophy* 89:6, 1992; and *Naturalism in Mathematics*, Oxford: Clarendon Press.

<sup>36</sup>. Whether all physical properties are intrinsic is another question. In the case of the electron, the consensus seems to be that negative charge is intrinsic. But things get more controversial when it comes to other physical properties. Some, for example, hold that curvature is intrinsic to physical space. Since an object's shape depends on the curvature of the space in which it is embedded, whether shape is intrinsic depends on whether curvature is intrinsic. But others deny that shape is intrinsic since it is always relative to a reference frame thereby denying that curvature is intrinsic. See Graham Nerlich's «Is Curvature Intrinsic to Physical space?» *Philosophy of Science* 46: 65-72.

<sup>37</sup>. See, for example, Hilbert (1992) and Hilbert and Byrne (2003).

about the colors but not about theoretical entities such as phlogiston or having a negative charge.

## 6. Conclusion

I have outlined an account of color that would allow us to continue using our ordinary color discourse as we have thus far despite the fact that it is flawed. To remedy the seeming oddness of this suggestion, it was prescribed that we stop asserting propositions attributing color properties to objects and instead begin uttering them as an act of make-believe. Further, by incorporating the dual-theory of color vision in our base discourse, we can explain a wide range of phenomena without having to postulate the existence of color properties, ordinarily conceived. As presented, the bridge laws have a dual function: they allow us to go from the base discourse to the fiction and back, but they also allow us to distinguish between color utterances that are appropriate and color utterances that are not. Lastly, I have presented a number of worries and have argued that none presents a serious threat to my proposal. There is thus little reason to shy away from it.<sup>38</sup>

## REFERENCES

- Arstila, Valtteri. (2005) *The Paradox of Colors* (dissertation), Finland: University of Turku.
- Boghossian, Paul A. & Velleman, David J. (1991) «Physicalist Theories of Color», *Philosophical Review*, 100:1: 67-106.
- Carnap, Rudolf. (1967) *The Logical Structure of the World*. London: Routledge & Kegan Paul. Translated by Rolf George.
- Cohen, Jonathan. (2003) «Color: A Functionalist Proposal», *Philosophical Studies*, 113(1): 1-42.
- \_\_\_\_\_. (2004) «Color Properties and Color Ascription: A Relationalist Manifesto», *The Philosophical Review*, 101:3: 551-588.
- Cohen, J., Hardin, C. L., McLaughlin, B. P. (2006) «True Colors», *Analysis*, 66: 292: 335-340.
- \_\_\_\_\_. (2007) «The Truth about «The Truth about True Blue», *Analysis* 67: 294: 162-166.
- Currie, Gregory. (1990) *The Nature of Fiction*. Cambridge: Cambridge University Press.
- \_\_\_\_\_. (1986) «Fictional Truth», *Philosophical Studies*, 50: 195-212.
- Ellis, Jonathan. (2005) «Colour Irrealism and the Formation of Colour Concepts», *Australian Journal of Philosophy*, 83:1: 53-73.
- Field, Hartry W. (1980) *Science Without Numbers: A Defense of Nominalism*. Princeton: Princeton University Press.
- Gatzia, Dimitria Electra. (2007) *Color Fictionalism: Color Discourse Without Colors* (dissertation). New York: Syracuse University.

---

<sup>38</sup>. I am grateful to Barry Maund, Andre' Gallois, Mark Kalderon, and especially C. L. Hardin for helpful comments and suggestions.

- Geach, Peter T. (1965) «Assertion», *Philosophical Review*, 74: 4: 449-65.
- Gouras, P. and Zrenner E. (1981) «Color vision: a review from a neurophysiological perspective», *Progress in Sensory Physiology* 1: 139-179, (Eds.) Autrum, H., Ottoson, D., Perl, E.R. and Schmidt, R.F. Springer-Verlag, Berlin Heidelberg New York.
- Hardin, C. L. (2004) «A Green Thought in a Green Shade», *The Harvard Review of Philosophy*, XII: 30-39.
- \_\_\_\_\_. (1998) «Basic Color Terms and Basic Color Categories», *Color Vision: Perspectives from Different Disciplines*. Berlin: Walter de Gruyter & Co.
- \_\_\_\_\_. (1997) «Reinverting the Spectrum», *Readings on Color: The philosophy of color*, vol. 1 (Ed.) Hilbert, D & Byrne, A. Massachusetts: The MIT Press.
- \_\_\_\_\_. (1993) *Colors For Philosophers: Unweaving the Rainbow*. Indianapolis: Hackett.
- \_\_\_\_\_. (1990) «The Virtues of Illusion», *Perceiving, Measuring and Using Color, The Society of Photo-optical Instrument Engineers*, (Ed.) Brill, M.
- \_\_\_\_\_. (1984) «Are 'Scientific' Objects Coloured?» *Mind*, 93:372, 491-500.
- Hering, Ewald. (1964) *Outlines of a Theory of the Light Sense*. Hurvich, L. M., & Jameson, D. (trans.) Massachusetts: Harvard University Press.
- Hilbert, David R. (1992) «What is Color Vision?», *Philosophical Studies* 68: 351-370.
- \_\_\_\_\_. (1987) *Color and Color Perception: A study in Anthropocentric Realism*. Stanford, California: CSLI.
- Hilbert, David & Byrne, Alex. (2003) «Color Perception and Color Science», *Behavioral and Brain Sciences*, 26: 1: 3-21.
- Jackson, Frank. (1996) «The Primary Quality View of Colour», *Philosophical Perspectives*, 10: 199-219.
- Johnston, Mark. (1997) «How to Speak of the Colors.» *Readings on Color: The Philosophy of Color*: vol. 1 (ed.) Hilbert & Byrne. Massachusetts: The MIT Press.
- Joyce, Richard. (2001) *The Myth of Morality*. Cambridge: Cambridge University Press.
- Kalderon, Mark Eli. (2005) *Moral Fictionalism*. Oxford: Clarendon Press.
- \_\_\_\_\_. (2007) «Color Pluralism», *The Philosophical Review*, 116: 4: 563-601.
- Levine, Joseph (2006) «Colors and Color Experience: Colors as Ways of Appearing», 60: 3: 269-282.
- Matthen, Mohan. (1999) «The Disunity of Color,» *The Philosophical Review*, 108:1: 47-84.
- Maund, Barry. (2006) «The Illusory Theory of Colour: An Anti-Realist Theory», *Dialectica*, 60: 3: 245-268.
- \_\_\_\_\_. (2006a) «Color» in the *Stanford Encyclopedia of Philosophy*.
- \_\_\_\_\_. (1995) *Colours: Their Nature and Representation*. New York: Cambridge University Press.

- McLaughlin, Brian P. (2003) «The Place of Colour in Nature», *Colour Perception: Mind and the Physical World*. (Ed) Mausfeld, R. & Heyer, D. Oxford: Oxford University Press, 475-502.
- Nassau, K. (1997) «The Causes of Color», *Readings on Color: The Philosophy of Color*: vol. 2 (ed.) Hilbert & Byrne. Massachusetts: The MIT Press.
- \_\_\_\_\_. (1983) *The Physics and Chemistry of Color: The fifteen Causes of Color*. New York: John Wiley & Sons.
- Nolan, D., Restall, G., & West, C. (2005) «Moral Fictionalism versus The Rest», *Australasian Journal of Philosophy*, 83:3: 307-329.
- Palmer, Stephen E. (1999) *Color Vision: Photons to Phenomenology*. Massachusetts: The MIT Press.
- Rosen, Gideon. (1990) «Moral Fictionalism», *Mind* 99: 327-54.
- Searle, John R. (1969) *Speech Acts*. Cambridge: Cambridge University Press.
- Stroud, Barry. (2000) *The Quest for Reality: Subjectivism and the Metaphysical of Color*. New York: Oxford University Press.
- Szabó-Gendler, Zoltan. (2001) «Fictionalism and Moore's Paradox», *Canadian Journal of Philosophy*, 31: 293-308.
- Tye, Michael. (2006) «The Puzzle of True Blue», *Analysis* 66: 173–78.
- \_\_\_\_\_. (2006a) «The Truth about True Blue», *Analysis* 66: 292: 340.
- Van Fraassen, Bas C. (1980) *The Scientific Image*. Oxford: Clarendon Press.
- Werner, John S. & Webster, Michael A. (2002) «Color Vision is Form and Object Vision», *9<sup>th</sup> Congress of the International Colour Association*, Proceedings of SPIE 4421: 10-15.

**Dimitria Electra Gatzia**

**The University of Akron**

**dg29@uakron.edu**

SORITES (Σωρίτης), ISSN 1135-1349

<http://www.sorites.org>

Issue # 21 — June 2008. Pp. 52-62

Kant on Pragmatism: Kantian Notes to Brandom's  
Inferentialism

Copyright © by Jesús González Fisac and SORITES

# KANT ON PRAGMATISM: KANTIAN NOTES TO BRANDOM'S INFERENCEALISM

Jesús González Fisac

## 1. The kantian dualism: a pseudoproblem? Outlines of the pragmatic solution.

A *dualism* is a distinction whose «components are distinguished in terms that makes their characteristic relations to one another ultimately unintelligible».<sup>1</sup> In Brandom's opinion this is what happens with the kantian distinction between concept and intuition, since schematism is not a good solution. However, as Brandom points out, it can also be considered that Kant maintains a *pragmatism* in his theory of cognitive activity<sup>2</sup> as well as in his schematism theory. But what is pragmatism? The aim of our paper is to support the following assumption: the starting point of Kant's philosophy (including the theoretical philosophy as well as the practical philosophy and the philosophy of history) is that *experience* is not primarily sense data, which are different from conceptual contents (we will go back to this point when we examine the first dualism). That is, experience is not only the reserve of contents.<sup>3</sup> As Kant says in the third Critique, experience is the only possible ground or territory (*Boden, territorium*)<sup>4</sup> of every knowledge (this is probably the more precise and at the same time the broader meaning of the word). Now, knowledge is not an intellectual output in which such a distinction between sensory and conceptual contents would be plausible.

---

<sup>1</sup>. Brandom, 1998, p. 615; see also Brandom, 2001, p. 167.

<sup>2</sup>. Brandom, 1998, p. 80.

<sup>3</sup>. Kant says that the representations of the external senses constitute the only material or stuff, *Stoff*, with which the mind is occupied with (*KrV*, B 67). But he does not say the same about external experience. Experience in general is a «*bathos*» (IV, 373) or an «inexhaustible» reserve of new information (*KrV*, A 1).

<sup>4</sup>. *Kritik der Urteilskraft* [KU], V, 174 (*Critique of the Power of Judgment*. Trans. Paul Guyer, New York: Cambridge University Press, p. 61-2). Passages of Kant's works are quoted as usual according to the academical edition *Kants Gesammelte Schriften* (volume, in roman numbers, and pages, in arabic numbers). The passages of the *Critique of Pure Reason* will be quoted according to the original first (A) and second (B) edition.

Knowledge belongs also to the life and activity of the faculties of the mind. Furthermore, as we will see in what follows, *experience is a pragmatic game*.

As is well known, Kant makes a distinction between two different points of view, the physiological and the pragmatic one, in order to build a doctrine of the human being (VII, 119). Notice that there are not two kinds of knowledge, which would demand another difference, that is, the difference between two realms (*ditio*).<sup>5</sup> The difference between the two legislations which reason can yield is not at stake; neither is the critical knowledge. It is rather a distinction which depends only on the position of the subject in relation to itself. Of course it is not the self-positing subject supported by idealism (Fichte), since in this case there is no relationship at all (for Kant the logical selfawareness is tautological);<sup>6</sup> on the contrary, this position goes together with the subject's recognition of its position in the world. This activity, as well as its sensible effect on itself called feeling, can be assigned to the *Urteilkraft* (the Feeling of Pleasure and Displeasure is also the Faculty of reflection and of feeling). Notice that the position of the subject is a position related to others and also to himself. That is, world is not a simple correlatum but a milieu. It is a similar phenomenon to that of orientation where subject feels himself occupying a place (VIII, 139-140). Now, the pragmatic knowledge can be located in the same range. Since the pragmatic point of view is focused on «what he [the human being] as a free-acting being makes of himself, or can and should make of himself»,<sup>7</sup> it requires that man becomes aware of himself as another, that is, it requires that he takes himself as a possible object of his activity (even if this activity is thinking, or more precisely, reflecting or apperceiving *–infra*). That is, human being is able to become object of a making, *facere* (cfr. *KU*, V, 303). This strong link between man and the world has one of the first approaches in the doctrine of empirical apperception and the inner sense of the first Critique but it is rooted in every output of the human faculties of reflection and feeling.<sup>8</sup> Human beings are able either to act on themselves or to find themselves because they are, first of all, beings-in-the-world. In other words, for Kant solipsism is nonsense (it can only be a malady of the head). The human being has a world and is not only satisfied with contemplating it. This is the very meaning of pragmatics in a broader but genuinely kantian sense. The human being is aware of himself inasmuch as *he is taking part in the game of the world*.<sup>9</sup>

Let's take the case of theoretical knowledge. Experience is the very knowledge. But experience is also the beginning of every knowledge. Experience is not only

---

<sup>5</sup>. *KU*, V, 174.

<sup>6</sup>. See *KrV*, B 409. See below, note 8.

<sup>7</sup>. *Anthropologie*, VII, 119. See also Kant, *Anthropology from a Pragmatic Point of View*. Trans. R. Louden. Cambridge (UK), Cambridge University Press, 2006, p. 3.

<sup>8</sup>. The pragmatic point of view shed an unexpected light on the phaenomenon of self-affection. It is common not to consider the example of «attention» which Kant introduces to explain this phaenomenon (*KrV*, B 156-7, note). But notice that attention is a genuine anthropological activity (see *Anthropologie*, § 3). Attention involves a *force* which is produced by the subject himself, and since this force is directed to himself (Kant uses the word «*Bestreben*», intentness), attention must overcome some resistance. Thus, the possibility for the subject to be active and passive is not an oxymoron. Knowledge must only be considered as such a wordly game.

<sup>9</sup>. *Anthropologie*, VII, 120.

phenomenologically but also cronologically the first *datum* (B 1). This ambiguity shows us that we can not consider what theoretical knowledge is unless we take into account our relation to time. In fact, the key to the first Critique, schematism, is the attempt to explain such a relation. Now, Judgment<sup>10</sup>, *Urteilstkraft*, is the very mediatory faculty (in the sense of a «common root», not of a derived faculty). But Judgment is the faculty that must discern which concept, that is, which rule is to be applied to the case (B 171-2). This is why Judgment is also *the* faculty of experience (in fact, Judgment is not a «faculty» which gives rules for the sake of knowledge but a «talent» to find them; cfr. B 172). Judgment is the faculty which works in our empirical knowledges, either to subsume them under general laws or to find a law for a given fact (cfr. *KrV*, A 646/B 674 ff.). Notice that these knowledges are always only the knowledges of the case, because experience, the instance or the case as such, is the first (see B 1-3). Theoretical knowledge belongs also to experience. It always takes place where there are rules and in the application –in the very functioning– of those rules as well. This means that theoretical knowledge is not only a derived application from transcendental principles in a higher level than that of experience; it is also *the fact that we can know as beings whose faculties are in the experience too*. This is the very meaning of «transcendental»: its function as condition of possibility (not as mere concept but as determination for an existence). So, together with the transcendental performances of those faculties, transcendental apprehension, transcendental association and transcendental apperception, there must also be their corresponding empirical events: empirical apprehension, empirical association and empirical apperception (cfr. A 98-110; see also B 164). Faculties are settled in the territory of experience. But, this is the point, they are not settled as possible objects of an empirical knowledge. Since Faculties are part of life, the constitutive dimension of faculties involves a very particular implication of experience. In the case of theroretical knowledge, transcendental knowledge (which is the constitutive dimension of the Faculty of Knowledge) deals with a kind of *retraction* of the experience in the very constitution of its possibility. But, in general, every critical knowledge involves the constitution of its realm as a kind of experience (this is confirmed by the presence of sensibility, that is, of an aesthetic part, in every critique).<sup>11</sup> To this experience belongs every phenomenon of reason, every output of the different games of forces in the *Gemüt*. In this sense experience is always a game where there are rules, since every faculty belongs to Nature, as well as a resistance to them, *and* since reason (in a broader sense) is a free force. In other words, in experience reason recognizes itself as a part of a game and as a player as well. The game of the

<sup>10</sup>. In german, and specially in Kant's uses of language (due either to his own thinking or to his Age), there are three terms which are conceptually and etimologically conected: *Urteilstkraft*, *Beurteilung* and *Urteil*. Since they have a common root, it is important to notice the difference between them as well as their ambiguity. *Urteilstkraft* is the Faculty which performs either *Urteilen* or *Beurteilungen*. *Beurteilung* is, in a strict sense, the name of every execution of the *Urteilstkraft*. Although Kant's use of *Beurteilung* is mainly found in asthetical texts, this term means the very performance of the *Vermögen zu urteilen* in general (B 94). The ambiguity arises from the fact that Taste, the appreciation of beauty, has its own judgments, *Urteilungen*, which are non-objective, and therefore un-interested (V, 205), neither in a theoretical nor in a practical sense. This fact reveals the subjective economy of *urteilen* as a game of powers of faculties. *Urteilstkraft* has thus deep roots, roots behind our relation to objects, in an «harmonious accordance». Anyway, in what follows we make such a distinction, together with its ambiguity, translating «Judgment» or «power of judgment» for *Urteilstkraft*, «judgment» for *Urteil*, and «judgement» for *Beurtheilung*.

<sup>11</sup>. In the first Critique, the «Transcendental Aesthetic» (but also the second edition, speciallu the chapters of the Transcendental Deduction and of the Paralogisms); in the second, the chapter of the *Triebfeder*, 'driving forces' (see L. Beck, *A Commentary of Kant's Critique of Practical Reason*, Chicago/London, The University of Chicago Press, 1963, p. 209 ff.); and, of course, the third Critique itself.



theoretical and practical knowledge (including Analytics and Dialectics), of the methodological («the secure path of a science») and of the historical knowledge, of the *sensus communis* (communicability), of the Enlightenment, etc., every phenomenon on which critical and metaphysical works have been focused are phenomena of this type.

Actually Kant explains the work of the faculties as a ruled one. That is the case of Understanding. Brandom has made the distinction between «according to rules» and «according to the representation of rules» (Brandom, 1998, 31-2). This distinction is necessary to elucidate the feature of normativity of the faculties. Everything which occurs in nature obeys rules, says Kant (*Logik*, IX, 11). We find this fact in the inferior or empirical faculties as well as in the superior faculties, even in the game of the *a priori* knowledges. The problem in the human case is that rules belong to both Nature and Reason. Human beings are gifted by *Gemütskräfte*, forces of *Gemüt*, and they are not only natural beings, that is, mechanical and physiological bodies which are blindly ruled by natural laws. Hence human beings act not only conscious of following either natural laws or their own rules but also aware of their power to act over themselves. In both cases they are conscious of such laws, so that they are always able to decide to follow one of both laws or not, but they are also able to modify the activity of their *Gemüt's* faculties which is ruled by nature. That is the middle position which make human beings able to act pragmatically, since they can make representations of the laws of faculties (subjective and objective genitive). We think it is not correct to interpret this difference in terms of «implicit» and «explicit» because pragmatism must understand what is the case in other terms rather than that of consciousness.

Brandom takes the distinction «according to rules» and «according to the representation of rules» from the practical philosophy. The mistake maybe rests on the consideration of the normativity in terms of actions of the will (Brandom, 1998, pp. 31-32). This means that it is a «normative» difference, not a factual one (Brandom, 1998, p. 48). However, in order to avoid the consciousness in the interpretation of normativity, it is necessary to consider the action, *Handlung*, rather as an exercise, *Übung*, than as an application. Since morality rests on the universality of the maxim of the action, Kant has sustained his fundamentation of morality through the logic of the subsumption. Judgment, however, is not a mere logical faculty. After all, it is the faculty of reflexion, a logical operation which involves the other faculties. Reflexion is the mundane operation of the logic, which moves the *Gemüt* in both directions, from the rule to the case and from the given case (even if it is given to us through sensibility) to the rule. In theoretical knowledge this movement performs a positive knowledge or a knowledge of experience. In practical knowledge, sensibility is not a positive part of the operation of knowledge (decision). Here intuition is only a resistant instance which must be overcome by reason as well as the product of this resistance (the feeling of respect). But in both cases, in the theoretical and in the practical knowledge, the operations of the faculties of *Gemüt* are movements towards a positive result. Thus, in our opinion the analysis of any activity should take the point of view of *intentionality*. Intentionality and pragmatism (and not psychologism, epistemology or logicism) are the two interpretative keys to Kant's philosophy. Notice that if there is one and only one game of the world, it could be said that practice is previous to any explicitation, and that the game is also before the practice. Intentionality means in this context the zero logic of every game, i.e. the logic of the *very working of the game*. Hence, normativity is not an ideal of regulation but the concrete play of rules and attempts.

Actually, normativity is a finite normativity. That means that there can be mistakes, both in practice and in judgement, *Beurtheilung*, of norms (Brandom, 1998, p. 52). Notice that in both cases there is the same assumption: practice and normative status can not coincide. This is the starting point of the pragmatic point of view. What it ought to be and what it is in fact are not the same thing (ibid, p. 11). Thus, taking into account that there is such a separation, and since judgement is also a practice which belongs to the order of life, judgement is also exposed to mistake. We agree that if we want to gain a pragmatic version of Kant, the idealism which considers that, since the explicit norms cannot be denied by any fact, only these norms can be principles in *sensu stricto*, must be left out. We can say, with Kant, that transcendental dualism must be avoided. This dualism can be considered as a pseudo-difference, since it only sustains one of the components, either things or representations, at the expense of the other, that is, at the expense of making it something absolutely inaccessible or «transcendental» (in the sense of *KrV*, B 81). In the case of normative idealism (which can be considered as a kind of material idealism),<sup>12</sup> such dualism deprives facts of every meaning. The other possibility is the dualism of practices, which takes them for the one and only reality, independently of their normative adjustment. Notice that this dualism cannot be avoided by considering normative status as implicit. In our opinion, there must be a practice which can be distinguished from the normative status but, and this is the crucial problem in our opinion, which is not absolutely independent of it.

Thus, the mistake does not lie only in the separation of normative status and practices, but also in the assumption that such a separation affects the *consciousness* of the norms. Rules in themselves are not norms; it is the fact of taking the rules as means to obtain outputs, in this case the output of knowledge. In other words, acting in accordance to rules is not a practice since there is no intentionality involved. It is in the field of this intentionality, in the worldly fact of judgement, where freedom takes its place and plays a factual role. Both if we speak about practical knowledge as such and if we speak about the theoretical one.

We will turn back to this point when we deal with the difference between spontaneity and receptivity, but here we just want to point out what we consider as a mistake in the approach to the question, at least in kantian terms (unless we will see that Brandom gives us a suitable distinction to solve this difficulty). Brandom recognizes that certain practices have a normative condition; the normativity of these practices is implicit and our normative attitude towards them is what makes them explicit. What we understand, with Kant, is that the rules according to which whatever natural event takes place, rules which we can not suppress nor elude in anyway, are not a practice. There is only something like practice in so far as we act with intentionality, which is the consensus noun we can put to judgement. But, what about the mistakes? The mistakes have its origin in the violation of the normative attitude by the very normative status; in kantian terms, when the experience in which we find and which we recognize in our practices is taken as the validity and the normativity itself. Mistakes are always mistakes of application and therefore they belong to the own Judgment understood as talent. But let's go on with the distinctions.

## **2. The dualism form/matter and the need of sensibility's conditions.**

Brandom interprets the dualism form/matter in terms of the difficulty of the form, which is intellectual, and acts over the matter, which is sensorial (the dualism is thus interpreted in

---

<sup>12</sup>. For this idealism, understood in its original gnosological terms, see *KrV*, B 274

intellectualistic terms). Form and matter are interpreted in this way since both are considered incompatible, so that it seems that one can only be maintained at the expense of the other; or, at least, as a distortion, if not an occultation, of one by the other. This is the idealism of consciousness. The fact is that we can only have clear representations if the sensorial matter has been formally depurated by consciousness; that is, if the matter is elided from the cognitive representation (*KrV*, B 59-60). But this is not Kant's assessment. As he points out, «matter» and «form» are actually only «concepts of reflection», concepts destined to judge the difference between representations depending on the faculty of knowledge from which they proceed, either from sensibility or from understanding (B 317-8). As we have said, this is a kind of orientation of the *Gemüt* between its forces and faculties (called «transcendental reflexion» when the faculties concerned are the faculties of the possibility of theoretical knowledge, understanding and sensibility). Kant's problem are both intellectualist positions, as those of Leibniz, who considers the understanding as an isolate and autonomous faculty of knowledge, and empirist positions, as those of Berkeley's material idealism, for which sensibility is the source of phenomenal appearances. Only if we recognize the understanding and its operations as merely logical (when understanding works alone) we could avoid the intellectualistic difficulties, as that of the position of bodies in space (the principle of identity of indiscernibles). Through reflexion the *Gemüt* becomes aware of the forces and their specific outputs, as well as of the disorders and confusions, in their concrete uses, that is, when faculties are concerned with each other in different ways. But reflexion is not an activity promoted by nature but a talent which must be exercised. Thus, reflection is an obligation, *Pflicht* (*KrV*, A 263/B 319), whose lack has brought philosophers to unilateral interpretations of knowledge.

But let's approach one of Brandom's thesis. Brandom tries to overcome this kind of dualism through his theory of inferentialism. Since knowledge is a discursive practice, its articulation in complex terms can be thought of in inferential terms which are normative relationships between contents considered materially; contents determinate themselves reciprocally through rules which constitute practices. We know as we make statements (knowledge is basically assertive), which can only be understood as inferentially connected to other statements and propositional contents. There is however a «dogma of formalism» (Brandom, 1998, p. 97 ff.) which considers that the inferential articulation is logical. The assertion that «Granada is to the east of Cádiz» leads to the inference that «Cádiz is to the west of Granada»; here the connection is material and based on what Kant called in a precritical work «directions of space». Thus, it is not a logical connection. Or at least it has not the logically inferential form of the conditional. Since this if-then form cannot give reason of material inferences, «form» must be understood as a normative status of non-logical contents.

We find a double confirmation of this. First because the connections that we can consider as materially significant in a pragmatic and linguistic sense, the categories, are concepts which are found in material assertive commitments; this is, categories are found in commitments to the conditions of the possibility of experience. Notice that the problem of Kant in the second edition of the *Critique of Pure Reason*, the problem of the Refutation of idealism, consists of carrying the commitment of the pure concepts of understanding also to space (see B 288 and B 155, note). Schematism expresses primarily the commitment of the understanding to the inner sense and to its form, time. The unexpected emphasis on space as a necessary correlate of categories reveals a specific normative status where the understanding is also connected with the world.

Second, we can also find in Kant the problem of a similar formalism if we consider that the form of judgments is based only on the principle of non-contradiction and that, therefore, it is based on a concept of the relationships supported by the logical exclusion of the non-identical notes, this is, supported by the form of the identity of predication. Thus, rationalist dogmatism takes the concept rather as a set of predicates (quantity), as affirmation of the predicate (quality), as the condition of the predicate (relation) and, finally, as the mere position of the predicate (modality). Kant, however, shows that the formal logic has an inalienable material background and that the synthesis, which is the commitment of the form of judgment with the form of space and time (the form or essence of Judgment), rules this form (*KrV*, B 110-111). This means that at first there is the totality, which is the unity of the manifold, or singularity (quantity), the limitation, which is the negation of reality (quality), etc.

Let us continue with the next dualism.

### **3. The dualism general/particular and the ontological extent of the form of judgment**

For Brandom Kant's understanding of judgment seems to belong to a dogmatic tradition where concepts are general representations, the representations of a class which subsumes individuals, which are precisely the non-conceptual. According to this «classificatory model» (Brandom, 1998, p. 86) judgements must not be understood in terms of material inferences. This is an intellectualised version of the form of judgment, a version whose basic assumption is to consider concepts as *explicitly* given in mind (*idem*). However, as Brandom accepts, a pragmatic version of Kant is possible. Kant agrees that the minimum unity of knowledge is judgment and not noun. So judgments are part of other logical relationships of a higher level, that is, of arguments and inferences. In fact, behind understanding is reason; and also Judgment. It is the same as we have said before about the pragmatic primacy of this faculty, as the faculty of the application of rules to instances. Because Judgment is the faculty which makes the distinction general/particular possible.

Inferences belong to a faculty of the mind. As a particular inner force, every activity can be considered as a particular making of the subject onto itself. Knowledge is a discursive practice in which we make assertions only because we are able to give reasons for these assertions. There is a kind of inferential background, a holistical background, which sustains the practices. But, is there anything like that in Kant? Let's go back to our main argument. In Kant we can not separate the experience which we know from the knowledge of the (which is itself) experience (*KrV*, B 1). This means that all knowledge always constitutes a certain product of our faculties. Because knowledge takes places whenever we make a statement such as «x is y» (let's agree with Brandom that the form of knowledge is the form of assertion), this is, whenever it always takes place as a particular ontological knowledge. But this particular knowledge is always within inferentials nets. The «regulative principles» are the rules which constitute this net. In these principles we find the faculty of reason in its broadest sense, that is, as another name for Judgment. These principles are probably the clearest confirmation of the intentionality of *Gemüt* in its activity of knowing. The regulative principles are actually principles of orientation of the faculty of knowledge through experience (they rule how to find new empirical knowledges).

The principles of homogeneity, species and continuity (B 685) constitute the practico-discursive whole of knowledge which connects every assertion to any other one. This whole is the system, the articulated whole of knowledges about nature, the whole of knowledges of physics, if we consider «physics» as the assertions about the particular laws of experience.

The regulative principles displace the predicates upwards and downwards; « $y$  is  $z$ » means that  $y$  is subsumed under a superior genus, « $w$  is  $x$ » means that  $x$  subsumes a inferior genus. Notice that this displacement, which is done thanks to the classificatory form of judgment and concept (from species to genus and from genus to species), is not merely classificatory since, as it is implied in the third of the principles –that of continuity–, knowledge consists of recognizing all the possible «distinctive grades» of a thing. The regulative principles are principles of determination, principles which put something in the net of material inferences (where the net is the assumption which is working in our cognitive activity, and so the regulative use is a «hypothetical use» –cfr. B 678). Regulative principles are *search* principles which look for the very place of a knowledge; they are principles of inferences searching. Above all, they are principles of knowledge considered as an activity and as a practice. So they belong to Judgment and so they are principles of purposes. The aim of these principles, the sense of system, is to place the things which we know as singular (the things related to their particular existence, considered as «this» or «that» thing) within the net of experience. The notion of purpose is the right inferential and material version (also de subversion) of the notion of concept. «... the purpose is the object of a concept, in so far as the concept is regarded as the cause of the object (the real ground of its possibility)» (*KU*, V, 220). The purpose is the representation whose meaning is not derived from the thing; the purpose constitutes the thing itself in its singularity. It is the causal, material and inferential representation of the concept. «... and the causality of a *concept* in respect of its *Object* is its purposiveness (*forma finalis*)» (*idem*). The purpose is the intentional representation *par excellence*.

We find the same inferential work in practical knowledge in the rules of prudence. In the strict sense of the term, prudence is merely «the skill in the choice of means to his own greatest well-being» (*Grundlegung*, IV, 416). It is an hypothetical knowledge. The counsels of prudence do not prescribe the action categorically; they prescribe it as a mean. However, since the purpose of these rules, happiness, is given as an essential purpose of man as a finite being, since it is a material but not a particular purpose, prudence is not only a skill but it is related to wisdom. Happiness belongs to the supreme good, which is the object of the rational will and consists of the conjunction of happiness and morality. Thus, prudence is a kind of practical knowledge, but not a constitutive one, which is the very morality; prudence is a regulative knowledge (*KU*, V, 457). It is the knowledge to reach the supreme good. In this sense prudence is related to wisdom. It is an inferential and (in kantian terms) quasi-practical knowledge. This is, a pragmatic knowledge or an intentionally practical knowledge.

Granted that there is a knowledge that we can consider as typically pragmatic, like the knowledge implicit in practices, the problem then is if this knowledge includes that unity between rules and representation of the rules which we have pointed out at the beginning. So let's go to the third dualism.

#### **4. The dualism spontaneity/receptivity or what is a normative status?**

We have already said that knowledge, since it is an intentional attitude, also implies a normative attitude. Thus, theoretical knowledge consists of the acknowledgement and use of the laws of nature while practical knowledge consists of the acknowledgement and use of the laws of freedom. This acknowledgement occurs in a special game called «regulative use» (or «prudential use» in the case of practical knowledge). Knowledge only takes place when we act according to the representation of the laws themselves. When these laws happen in nature we must not speak about knowledge. This is the event of representational phenomena like

empirical association or reproduction or of phenomena of behaviour like our actions through instinct. Knowledge occurs when and in so far as we are able to distance ourselves from the events of nature which we belong to. This can be a formulation of the difference between spontaneity and receptivity, as the difference which there would be between concepts qua representations and causal order (Brandom, 1998, p. 616). A difference which could be expressed in linguistic terms as the one between definite description and indexicals. However, if we want to consider that there are always only normative attitudes and that these are always only propositional, discursive practice must be understood in inferential terms; so spontaneity and receptivity have not to be separated, just as it is in Kant..

Here we want to show at least two things: first, that the difference between laws and the representation of laws is an intentional one and understandable in terms of normative attitude. Second, that normative status is basically political (or social, but in a very particular way).

Let's take into account the two following considerations. Every normative attitude implies that we take responsibility for whatever we do since we act according to the representation of laws. Brandom points out that this responsibility can be a responsibility for whatever we do, but also responsibility for whatever we have done (*ibid.*, pp. 862-3). The first responsibility would be the expression of spontaneity and the second the expression of receptivity. «Responsibility for» would be the base of our commitments, like the commitment to give reasons for our assertions and to justify them inferentially, which definite descriptions consist of (a inferentially satisfactory meaning). The second would constitute a normative attitude in its limitation (in the double sense of «limited» and «limitant») through the causes. Brandom thinks that the first responsibility is the one which takes place in the actions of the subject, both in transcendental actions of theoretical knowledge, ruled by the transcendental principles of understanding, and in moral actions of practical knowledge, ruled by moral law. The responsibility of receptivity would be that of assuming the given as the representational content; it would be the responsibility for whatever is the instance and for whatever is imposed as intentional content. In our opinion this double bind can be understood in kantian terms as follows.

Knowing is acting according to the representation of laws. This means that we assume the responsibility, not for principles and laws whose normative status is, that is to say, «pure», as something independent of experience, but for the effects which our actions will produce in experience. This is, the responsibility, which is involved in our spontaneous acting (whatever the product was, theoretical or practical), consists of assuming the consequences of our actions according to the causal laws which are responsible of their condition of factual and empirical actions. The spontaneity consists of assuming the authority of these laws as that of a force which is not in our power. This force is the force of our cognitive faculties and of our faculty of desire. Naturally, the acknowledgement of this force can not take place out of the own facticity of those laws, this is, out of the instance. Both in theoretical and in practical knowledge the commitment of spontaneity only takes place when the instance is given: a sensation of our intuition or a demand of our feeling. It takes place in knowing and in having experience of knowing. At that moment we are responsible for such experience we have not created and whose laws we cannot modify. At that moment we feel the force of laws.

Factual knowledge, the regulation in which actually knowledge consists of, is our orientation (finite and therefore sentimental, although we can not consider this point in detail) before the given laws of experience and nature. Our commitment to these laws and to their consequences (and also to the incompatibility of consequences among themselves); our

«inferential commitment» in Brandom's terms.<sup>13</sup> In our opinion this is the committive dimension of the assertions of knowledge (also of practical knowledge, whose typical form would be also «x is y», where x is an action and y is its/their consequence/s). In the critical works, which are another and very special kind of knowledge, we do not find the development of an explicit normative status but the acknowledgment and location of such commitment to laws, of our place between laws and representations, as beings who inhabit time (pure Reason) and action (practical Reason), this is, as beings who inhabit experience (Judgment).

Finally, let's consider the second topic pointed out above. Brandom thinks that the committive dimension of assertion means a dimension of justification. This dimension goes together with the fact that assertion must be produced by a practitioner entitled to do it. Since knowledge is a discursive practice, the pragmatic terms of experience or of world also include the speakers and their capacity of producing knowledges in general. The world, the world of purposes which gave sense to philosophy, is an intersubjective world. Does this mean that we are dealing with a practice and therefore before a typical social normative status? Is telling the truth, in Brandom's words, «a social deontic attitude»? (Brandom, 1998, p. 202). It cannot be denied that Kant dealt with the question of authority; but he did it in terms not only nor strictly legal. In his opuscle *An Answer to the Question: «What is Enlightenment?»*, as well as in his texts about University and its Faculties, the question is *who* are entitled to tell the truth. The very problem is the role of publicness in the development and performance of knowledge as a work of reason (called by Kant –VIII: 37– «public use of reason»). Therefore, the point would not be if one is entitled to speak or not, which is the problem of the coercion of power, but in so far as it is necessary to speak and to do it openly, which is the problem of power as a social and political (aesthetic) force. In other words, «freedom to think» is opposed both to «civil» as to «consciousness coercion» (VIII: 144-5).

The authority which is important is the one we have as rational beings. In fact, it is in the third Critique where Kant points out that we are entitled to make statements of knowledge because and to the extent that we are beings who are related to the things as particulars, as instances, in their non-conceptual singularity. The man who enunciates a judgment of beauty is not doing an assertion; he is simply enunciating his condition as a legitimately rational being, as a member of the rational community of men. This community is not transcendental in a habermasian sense but real and factual. The community of men who speak and have discursive veritative practices. Publicness and the mere possibility to speak with frankness are essential for the happening of the community. It is not that social practice was behind the deontic commitment; it is just the opposite: it is the deontic commitment with objects and actions which makes the social practice possible. Thus, Brandom's distinction between «the activity of instituting conceptual norms» and «the activity of *applying* those norms» (Brandom, 2001, p. 156) is nonsense. For Kant, we can only believe in norms since we recognize (or even feel) the necessity of fixing meanings (which is what experience consists of, a framing of meanings). Making rules in general, both theoretical and practical, is not an *a priori* activity, not even if we think that the normative institution is social. Rules are rather found. The fact is that we recognize meanings. This is the original fact: *there is* language. This fact is not merely transcendental since it constitutes the inalienable fact per excellence, neither a priori nor a posteriori; it is a metaphysical one. In other words, we can only

---

<sup>13</sup>. See also Brandom 2003, chapter 4: II.3.

stipulate and fixe meanings since *we are just applying them in (which consists of) experience*.

#### REFERENCES

- Brandom, R. (1998), *Making it Explicit*, Cambridge (Massachusetts), Harvard University Press.
- Brandom, R. (2001), «Vocabularies of Pragmatism: Synthesizing Naturalism and Historicism», in Brandom, R. (ed.) (2001), pp. 156-182.
- Brandom, R. (ed.) (2001), *Rorty and his Critics*, Malden (Massachusetts), Blackwell.
- Brandom, R. (2003), *Articulating Reasons. An Introduction to Inferentialism*, Cambridge (Massachusetts), Harvard University Press.
- Kant, I. (1902 ff.), *Gesammelte Schriften*, Ausgabe der preussischen Akademie der Wissenschaften, Berlin, de Gruyter.
- Longuenesse, B. (2000), *Kant and the Capacity to Judge*, Princeton (New Jersey), Princeton University Press.

**Jesús González Fisac**

**Universidad de Cádiz**

**[jesus.gonzalez@uca.es](mailto:jesus.gonzalez@uca.es)**



SORITES (Σορίτης), ISSN 1135-1349

<http://www.sorites.org>

Issue # 21 — June 2008. Pp. 63-70

What would you say then?

The philosophical appeal to what one would say

Copyright © by Renia Gasparatou and SORITES

## WHAT WOULD YOU SAY THEN?

### THE PHILOSOPHICAL APPEAL TO WHAT ONE WOULD SAY

Renia Gasparatou

---

Thought experiments, imaginary examples and counterexamples are used widely in philosophy as forms of argument and counter-argument. I suggest that the force and structure of many of them is based on an appeal to intuitions about *what one would say* when confronted with the imaginary case described. Ordinary language pronouncements (our spontaneous linguistic responses) are taken as safe premises that are capable of *demonstrating* a solution to a philosophical problem. *Demonstration* or *showing* is put forward, suggesting that genuine proofs in philosophy are surveyable. Any valid solution to a philosophical problem would be *showable* and taken in easily rather than *sayable*, that is in need of long argumentation and contemplation.<sup>1</sup>

Appealing to ordinary language pronouncements was back in the 50s suggested as the most intuitive way of coming to see the obviousness of the suggested solution. Although this line of argumentation is not always explicitly practiced as an appeal to ordinary language pronouncements, I suggest that similar rhetoric underlies many analytic philosophers' argumentation techniques.<sup>2</sup> In this paper I will try to pick up typical species of this methodology in classic analytic writings and suggest that the appeal to ordinary language pronouncements can hardly prescribe answers to hypothetical, extraordinary circumstances. And philosophical quests are extraordinary.

Invoking ordinary language pronouncements was the very core of *linguistic phenomenology* as practiced by J.L. Austin. We are put in an imaginary context and asked what *we* or *the plain man* would say if confronted with these circumstances. The answer the

---

<sup>1</sup>. The *saying/showing* distinction is first laid down in *Tractatus Logico Philosophicus* but goes on through the later work of L. Wittgenstein (see Wittgenstein 1958, § 128) and onto many analytic philosopher's practice.

<sup>2</sup>. In fact, I believe that a similar methodology lies behind most philosophical appeals to *common sense* (see Gasparatou 2009b & 2010c). Even today, the same rationale justifies most of experimental philosophy's surveys of what laypeople say (see Gasparatou 2008, 2009a, 2010a&b).

philosopher gives or implies is taken to be *the only possible* answer, the answer *everybody* would give. The argument supporting this practice is that language is a rule-governed activity. Any competent speaker follows linguistic rules when using language. A philosopher is as competent a speaker as any.<sup>3</sup> So her or his answer will be the correct (and obvious) one.

Thus the Austinian question of *what we would say if...* is supposed to invoke our commonsensical, ordinary linguistic reaction, and, thus, to demonstrate an inter-subjective, theory free, *obvious* solution to some philosophical problem. For example, Austin writes:

...It is clearly implied [...] that the ordinary man believes that he perceives material things. [...] if this is taken to mean that he would *say* that he perceives material things, is surely wrong straight off; for 'material thing' is *not an expression which the ordinary man would use*— nor, probably is 'perceive'. ...*what the plain man would say instead is that he hears peoples voices or that he saw a flame or a river -and he would never describe those as «material things.* (Austin 1964, p. 7. Italics are mine)

The appeal to ordinary language is not always that straightforward. Consider the following examples:

Suppose that someone said, pointing to Nixon, «That's the guy who might have lost». Someone else says «Oh, no, if you describe him as «Nixon», then he might have lost; but, of course, describing him as the winner, then it is not true that he might have lost». Now, *which one is being the philosopher here, the unintuitive man? It seems to me obviously the second. The second man has a philosophical theory.* The first man *would say*, and with great conviction, «Well, of course the winner of the election might have been someone else...»... (Kripke 1996, p. 4. Italics are mine )

Cats might turn out to be automata, or strange demons...planted by a magician. Suppose they turned out to be a species of demons. Then, ... *the inclination is to say* not that there turned out to be no cats but that cats turned out not to be animals... (Kripke 1996, p. 122. Italics are mine)

I chose to quote Saul Kripke above, since Kripke strongly opposes equating ordinary usage and meaning. I am not going to get into his theory of meaning here. I will just emphasise that his appeal to ordinary language is not supposed to provide us with an analysis of meaning as such. Rather, the question about *what we would say* is supposed to put us in touch with our *intuitions*. Those intuitions will serve as the self-evident bond of a term's reference with necessity and help Kripke argue for his theory on meaning.<sup>4</sup>

But what is apparent, then, is that, even for him, intuition is expected to reveal itself in ordinary language. Ordinary utterances are where he looks in order to find intuitions. That's why Kripke, in the examples just mentioned, speaks as *the everyday person*, reporting to us what *any careful speaker's* reaction to a situation would be. The spontaneous responses of an average speaker are supposed to provide us with the *obvious* reply; hence the *obviously correct* answer. The difference is that in Austin's case, there is an implicit theory according to which most philosophical problems arise from the misuse of ordinary terms. Paying

---

<sup>3</sup>. For a discussion of this argument, its versions, advocates and critics, see also Gasparatou 2009a.

<sup>4</sup>. O. Hanfling (2000, pp. 222-244) also points out the relation between appealing to intuition and appealing to *what we would say*. In fact he thinks that for Kripke «the question is essentially about *what we would say* and the talk of intuitive content is confusing and unnecessary [...] It is in this way the question must be addressed and not by appeals to intuition» (pp.242-243).

attention to *what we would say* can provide us with a safe starting point where no misuse is possible. For Austin the *description* of ordinary language is supposed to be the explicit task of the philosopher. In Kripke's work, on the contrary, the appeal to ordinary language seems to be just the medium in order to evoke intuitions that reveal our tacit grasp of the notion of necessity.

What I want to suggest, though, is that, leaving aside such differences, the methodology is very similar. In both cases, the philosophers appeals to *what we say* and, in doing so, they use ordinary language as a self-evident stand from which they proceed to demonstrate *the obvious answer* to the problem that worries him. Our shared, commonsensical, spontaneous linguistic responses guarantee the truth of their conclusions.

This line of reasoning can be found in many different philosophers making different kinds of claims. *Would we call XYZ water* if we found ourselves in the Putnamian Twin Earth? (Putnam 1975) *Would we say we knew* if we had to face Gettier or Nozick's counter examples of *knowledge as justified true belief*? (Gettier 2000; Nozick 1981) The limits of language are expected to reveal the limits of our imagination and conceivability.<sup>5</sup> In Austin's words: «the powers of our imagination...[are] in curious ways enslaved by words» (Austin 1979, p. 67). Philosophers who differ in their attitude towards meaning still share this methodology in order to *describe* what would be conceivable or not to a community of speakers.

The first presupposition underlying such practice is that *there is* some common ground in language (*i.e. ordinary language*) that determines what is *conceivable* and what not. This view overlooks the difference between alternative natural languages, special languages or dialects. Moreover it neglects the fact that many factors interfere with our linguistic performance, a lot of which have nothing to do with linguistic competence. Whenever someone is asked *what would they say*, social rules about what would be polite or appropriate to say, background knowledge or interests partly suggest the answer.<sup>6</sup> So, a biologist's response on whether we would call a bean *alive* will probably vary from a chef's response, for example, although both would qualify as average or competent speakers. But I will not go into this in detail here. I will just take for granted that, despite such concerns, language can provide us with a common ground that allows (at least some minimum) communication between speakers, and translation between different languages. The question, though, is whether ordinary language pronouncements can be used in order to *solve philosophical problems*.

In fact, it seems that most stories described in thought experiments or imaginary examples are constructed in such a way, that the reader is somehow *forced* to give the particular answer the philosopher wants to give. For we are already guided as to how to take the situation, what aspects of the story to concentrate on, what to pay attention to and what

---

<sup>5</sup>. The use of thought experiments, imaginative examples or counterexamples seems to evoke *imagination* and *conceivability* as if those two notions were synonymous (see Sorensen 1998). Although the equation of these two notions might be part of the problem of this kind of argumentation, I am not going to discuss these two notions here, but rather use them interchangeably, just as those who appeal to them do.

<sup>6</sup>. Recent studies of *experimental philosophy* also suggest that social, cultural or education factors influence our responses on philosophers' thought experiments and examples. See Alexander & Weinberg 2007; Knobe 2007; Gasparatou 2008, 2010a&b.

to neglect; and, thus, how to answer when asked *what we would say then*. When we are told about Twin Earth, for example, we are put into a situation that would completely change our conception of the world, as we know it. Yet, we are driven to concentrate only on certain aspects of the story - water, or Putnam's Doppelgänger who can't tell an elm tree from a beech tree (Putnam 1975 & 2000) - and forget about what difference the discovery of Twin Earth would have made in every aspect of our understanding, our lives *and our ordinary language*.

However, one may reply to such an objection that selectivity is *always* the case. Even if we are in a real life situation, we concentrate on certain aspects of it. And in any kind of reasoning- especially in scientific studies- one has to neglect some aspects of the phenomena studied in order to draw interesting and valid conclusions.

What seems to create the problem though, is that these stories describe very peculiar, extraordinary circumstances; contexts which ordinary language is not qualified to deal with. We are told to imagine that we are brains in a vat, travelers in a planet where everyone has a Doppelgänger and where people «even speak English» (Putnam 1975). And we are told what we would or would not say and then conclusions are drawn about *what is to count as knowledge or meaning*, or *whether physicalism is valid*, for example. Let me remind you Jackson's thought experiment, which, I think, is typical:

Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there is about what goes on when we see ripe tomatoes or the sky, and use terms like red, blue and so on (...). What will happen when Mary is released from her black and white room or is given a color television monitor? *Will she learn anything or not? It seems just obvious that she will learn something* about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and Physicalism is false. (Jackson 1982, p.128. Italics are mine)

This is a case when there is no explicit appeal to «what we would say». It seems to me though, that again the argument is based on an appeal to the ordinary linguistic game of *learning*; or, to put it differently, on the ordinary usage of the word «learn», on what we *would count as learning*. We could easily paraphrase Jackson's question, then, and ask *wouldn't we say that she learns something?* without altering the argumentative potential of Jackson's reasoning a bit. And, again, the answer is considered as *the obvious* answer, the answer *everybody* would give.

Yet, in this thought experiment we are asked to imagine a very eccentric story. And we do; we can picture Mary in her black and white room. But, as Daniel Dennett (Dennett 1991 & 2014) points out, it is when we start focusing on the details that problems arise: How come Mary has never seen herself? Well, she obviously has no mirror, but how come she's never seen her hands? Or whatever part of her body? So, she is forced to have black and white clothes on, even gloves, as it seems; forced by some other scientists probably, or by some alien-scientists. But doesn't she ever take a bath? Well, we might even imagine that they drug her before they give her bath. None of it sounds very realistic but, yes, we can imagine all that.

What we have no clue about, is what it means to say that Mary «acquires (...) all the physical information there is about what goes on when we see ripe tomatoes or the sky, and use terms like red, blue and so on». «All the physical information» doesn't seem to have an

obvious role in ordinary language. So, what kind of inclination about *what we would say*, are we supposed to have on that case? Why wouldn't we finish the story as Dennett does? Saying that when Mary goes out of her room and those bad people who held her captive lie to her about the color of some object, Mary having «acquired all physical information», immediately corrects them and gives the right name for the color (Dennett 1991, 399-400).

Is this a *bad* thought experiment? Doesn't it manage to make the point it wishes to make? I think it does manage that; and this is why Dennett calls such thought experiments *intuition pumps* (Dennett 1991 & 2014). But what Jackson's case shows, is that the invocation of our inclinations (or intuitions) about *what we would say or think* is very vague and can be easily molded, especially when confronted with very peculiar circumstances. As Wittgenstein says:

It is only in normal cases that the use of a word is clearly prescribed; we know, are in no doubt, what to say in this or that case. The more abnormal the case the more doubtful it becomes what we are to say. And if things were quite different from what they really are- ...- this would make our normal language games lose their point... (Wittgenstein 1958, §142)

Thought experiments usually describe science-fiction circumstances. Yet extraordinary cases are extremely difficult to handle when appealing to ordinary language. We cannot appeal to our intuitions about what we would say when our most secure beliefs are being questioned (Fodor 1964). Normally, *the situation itself is supposed to provide us with the answer*. But, in such peculiar cases, the circumstances cannot prescribe what we would say. It is very difficult to tell what would count as *obvious* in cases we cannot recognize as being part of our ordinary *linguistic games*; that is, in cases where we have no rules about what would count as the correct thing to say.

That's what Austin also points out when he writes:

Suppose you live in harmony and friendship for four years with a cat: and then it delivers a philippic. We ask ourselves, perhaps, 'is it a real cat? Or is it not a real cat?' 'Either it is or it is not but we can't be sure which' Now, actually, that is not so: neither 'it is a real cat' nor 'it is not a real cat' fits the facts semantically.... With sound instinct, the plain man turns in such cases to Watson and says 'Well, now, *what would you say?*» How would you *describe* it? The difficulty is just that: *there is no sort of description which is not misleading... Ordinary language breaks down in extraordinary cases.* (Austin 1979, p. 67. Italics are mine)

And this is why the Austinian stories, although hypothetical, are never unrealistic. In such abnormal cases, we have no idea what to say or how to describe the situation. Kripke, on the other hand, does not seem to share the same reservations. Consider the example:

We might find animals in some part of the world which, though they look just like a tiger, on examination they were discovered not even to be mammals. Let's say they were in fact very peculiar looking reptiles. Do we then conclude on the basis of this description that some tigers are reptiles? We don't ... (Kripke, 1996, p.120)

In such a case, Austin would probably not know how to answer this question; Kripke does. And although they both seem to appeal to the tacit descriptions of ordinary language, they adopt different attitudes as to the scope of their demonstrative power. What then becomes apparent from such a disagreement is that the philosophical treatment of ordinary language is an issue that cannot be resolved by ordinary language itself.

The extraordinariness of philosophical *thought experiments* however, is only part of the problem. Indeed, *the whole philosophical discourse is extraordinary*. The questions

philosophers ask are beyond the resources of ordinary language.<sup>7</sup> Describing a very peculiar situation, as happens in Jackson's thought experiment with Mary, for example, definitely makes it hard to appeal to our ordinary intuitions. The peculiarity of the situation is not the only difficulty. The root of the difficulty lies in the philosophical question posed by this thought experiment; *the question about physicalism* and whether it is true or false.<sup>8</sup> Questions such as this are not up to ordinary language to answer.

Philosophy starts when putting ordinary language and its resources under question. For it has the ambition to illuminate parts of the world that ordinary language does not pay enough attention to. Philosophers construct all kinds of hypothetical examples since, in order to explore possibilities, they explore all conceivable situations. Yet, ordinary language is inextricably tied to our existing language games and cannot be applied to cases or quests that lie far from the latter. And, for this reason, the question about *what one would say* is hardly relevant. It can be used as a device that could assure us that what we are talking about has not completely lost its meaning. But it is not *by itself* a self-evident premise for the demonstration of philosophical certainties.

In fact, if there is any need for philosophy (or science), it derives from the very need to evaluate, correct and revise our ordinary language descriptions. Ordinary language pronouncements can hardly offer the detached understanding or the objective explanations philosophers' seek. They cannot, by themselves, help when we ask general questions about *truth, meaning, necessity or knowledge, about other minds, self-identity or moral action*. They are oriented towards the practical purposes of everyday life, not the theoretical quests philosophers are occupied with, even if those quests had once arisen from everyday occurrences. That is, even if the sceptical doubt sprung after one's drinking or dreaming experience in everyday life, the moment he will start worrying about how we can establish that what we know about the world stands, is the moment when ordinary language can no longer be of any help.

J.L. Austin had realized that and that's why he tried to prevent that very step when one passes from the practical into the theoretical. Most of his work is dedicated in an attempt to «dismantle» philosophical doctrines «before (they) get off the ground» (Austin 1964, p.142). That's why, in Austin's view, *if something is «as good as a telephone [...], just is a telephone, no doubt about it»* (Austin 1964, p.119). For «everyday or practical or ordinary purposes» he adds and this is as far as he wants to go. But for Putnam or Kripke it would be outrageous to say: *«well, if it is as good as water, then it is water no doubt about it»!* For they too are trying to illuminate theoretical concepts and to offer a general, *detached, objective, true* conception of reference and meaning as they understand it. Those topics, as well as any of the philosophers' dilemmas, quests or theories, are alien to the ordinary language pronouncements. Therefore, it is both deficient and improper to treat such pronouncements as conclusive argument in support of philosophical theses.

---

<sup>7</sup>. I have argued elsewhere that experimental philosophers also make the same leap: when surveying laypersons beliefs about philosophical problems, they sometimes too use ordinary pronouncements to answer extraordinary questions (Gasparatou 2010a&b). Other than that, I believe that Austin would rather welcome experimental techniques in philosophy (see Gasparatou 2013). For experimental philosophy and its several appeals to intuitions, see also Knobe & Nichols 2008 & 2014.

<sup>8</sup>. In fact, supposedly appealing to ordinary language pronouncements usually amounts to ascribing to *soft naturalism*. See Gasparatou 2008.

As a concluding remark, I acknowledge that it is hard to provide a strict definition either of *ordinariness* or of *extraordinariness*. Here have used the term *extraordinary* to refer to whatever does not belong to our everyday practices or whatever may describe situations we do not normally find in everyday life. But there is also a more specific level of *extraordinariness* that has to do with any effort to go beyond the phenomena to an underlying system of explanation. Philosophy (and science) is extraordinary as far as they seek for such explanations. When suggesting a theory or a thesis that aims in a genuine solution to some underlying problem, then intuitions traceable by ordinary language are completely beyond the point.

#### REFERENCES

- Alexander J. & Weinberg J.M. 2007. Analytic Epistemology and Experimental Philosophy. *Philosophy Compass* 2.1: 56–80.
- Austin J.L. 1964. *Sense and Sensibilia*. Oxford University Press.
- Austin J.L. 1979. *Philosophical Papers*. Oxford University Press.
- Dennett, D.C. 1991. *Consciousness Explained*. Little, Brown and Company.
- Dennett, D.C. 2014. *Intuition Pumps and Other Tools for Thinking*. W. W. Norton & Company
- Fodor, J. 1964. On knowing what we would say. *Philosophical Review* 73: 198- 212.
- Gasparatou R. 2008. Armchair versus Questionnaire Polled Intuitions: Intuitions nevertheless. *The Reasoner* 2.11: 7-9.
- Gasparatou R. 2008. Species of Philosophical Naturalism, Science and Scienticism. *The International Journal of Humanities* 6.4: 27-35.
- Gasparatou R. 2009a. High-standard Epistemology and the Appeal to Intuition. *Filozofia* 64.7: 680-692.
- Gasparatou R. 2009b. Moore and Wittgenstein on Common Sense. *Philosophical Inquiry* 31.3: 65-75.
- Gasparatou R. 2010a. Experimental Appeals to Intuition. *Critica* 42.124: 31-50.
- Gasparatou R. 2010b. Folk Intuitions, Science-fiction and Philosophy. *Journal of Cognition and Culture* 10. 3-4: 377-382.
- Gasparatou R. 2010c. Philosophies of Common Sense from Reid to Experimentalists. *Journal of Scottish Thought* 3: 61-75.
- Gasparatou R. 2013. Naturalising Austin. *Acta Analytica* 28.3: 329-343.
- Gettier E.L. 2000. Is Justified True Belief Knowledge?. In S. Bernecker & F. Dretske (eds.) 2000. *Knowledge*. Oxford University Press: 13-15.
- Hanfling O. 2000. *Philosophy and Ordinary Language*. Routledge.
- Jackson F. 1982. Epiphenomenal Qualia. *Philosophical Quarterly* 32: 127-139.
- Knobe J. 2007. Experimental Philosophy. *Philosophy Compass* 2.1: 81–92.

- 
- Knobe J. & Nichols S. (eds) 2008. *Experimental Philosophy, Volume I*. Oxford University Press.
- Knobe J. & Nichols S. (eds) 2014. *Experimental Philosophy, Volume II*. Oxford University Press.
- Kripke S. 1996. *Naming and Necessity*. Harvard University Press.
- Nozick R. 1981. *Philosophical Explanations*. Oxford University Press.
- Putnam H. 1975. The Meaning of Meaning. K. Gunderson, ed. 1975. *Language, Mind and Knowledge*. University of Minnesota Press.
- Putnam H. 2000. Brains in a Vat. In S. Bernecker, F. Dretske (eds.) 2000. *Knowledge*, Oxford University Press.
- Sorensen R.A. 1998. *Thought Experiments*. Oxford University Press.
- Wittgenstein L. 1922. *Tractatus Logico-philosophicus*. C.K. Odgen, F.P. Ramsey, trans. Trench, Trubner & Co, London.
- Wittgenstein L. 1958. *Philosophical Investigations*. G.E.M. Anscombe, trans. Blackwell Publishers.

**Renia Gasparatou**

**University of Patras**

**gasparat@upatras.gr**



SORITES (Σωρίτης), ISSN 1135-1349

<http://www.sorites.org>

Issue # 21 — June 2008. Pp. 71-91

An Evolutionary Argument Against Epiphenomenalism  
and Type-Identity Theory

Copyright © by Brett Anderson and SORITES

## AN EVOLUTIONARY ARGUMENT AGAINST EPIPHENOMENALISM AND TYPE-IDENTITY THEORY

Brett Anderson

---

William James uses evolution to argue that our mental lives only make sense if epiphenomenalism is false (James, 1879). If epiphenomenalism is true, then given that humans developed over time due to natural selection, we would expect the a random relationship between our mental lives and our physical actions. To the contrary, our phenomenal experiences mirror our physical interactions, strongly indicating epiphenomenalism is false. Also, type-identity theories where mental events are identified with kinds of brain events, also fail because the identities, which are antecedent to any evolutionary context, fail to predict the situation we find, namely that our mental events are of the type we would expect if mental events have causal powers. First, I will present a short section on aspects of evolutionary theory that are pertinent to the argument. Then I will survey the historical roots of James' argument and similar contemporary arguments. Next, I will present James' argument against epiphenomenalism. I will then reconstruct the argument and respond to criticisms. Lastly, I will apply James' argument to certain type-identity theories.

### **Evolution**

In any evolutionary system there are replicators, whether organisms, traits of organisms, or behavioral strategies used by organisms. In addition, there must be one or more mechanisms that produce variety in the replicators. Thirdly, there is the environment where the various replicators differ in their reproductive success in any evolutionary system.

As far as biological evolution goes, the picture we find in the world is a constant flux between organisms and their environments. Each generation of organisms are shaped by their environment, while the environment is also changed by the organism and other forces. Contingency and complex interactions often make behavior of evolutionary systems difficult to predict. In general however, traits and strategies that tend to catalyze reproduction more than others will increase the proportion of the species population with those traits or strategies, and tend to increase the total species population.

This story is complicated by the fact that, as environments change, new experiences or organism/world interactions will trigger maladaptive behaviors, from an inclusive fitness point of view, in organisms. Behaviors that were once adaptive in the past can become maladaptive

due to environmental change. New objects or situations in a species' environment may trigger new selection pressures on the species' traits or behaviors. Whether traits or behaviors are fitness enhancing is purely contingent upon an organism's total composition and relationship to its environment. In addition, the variations involved in differential reproduction are limited to what the differentiation mechanisms, such as genetic mutation, make available; variations will be finite and fluctuating in number and kind. Of course, traits must be heritable in order for evolution to accumulate and maintain advantageous traits and behaviors at the generational level.

We can also apply evolutionary dynamics, not only to generations of organisms, but also to the lifetime of individual organisms that are capable of learning and reasoning. That is, a human individual can also be thought of as a sort of evolutionary system, in that behavioral changes can be based on past feedback and model testing. Humans are not just limited to random mutations in their behavior, but can model various actions in response to new situations and predict their outcome (Dennett, 1995, p. 53). Humans can also learn from the past behaviors of both themselves and others to modify their behavior. Behavioral strategies can be selected within the life of the individual, instead of relying exclusively on the slower and less efficient trial and error modification at the generational level.

There is a continuum of how evolution accounts for some human traits. On one end, there are robust genetic/developmental interactions that reliably produce certain traits. At the other end, behaviors are selected via learning, without change at the genetic level. Other than the fact that evolution has developed most of our basic capacities and systems, nothing in my argument turns on the details. In any case, it will be helpful to keep in mind this general picture of human evolution for what follows.

### **Spandrels and Complexity**

Here is a faulty argument for the causal efficacy of consciousness: traits of organisms are adaptations and must have causal powers if they are to be selected by evolution. Consciousness is a trait of some organisms, therefore it must have causal powers if it was developed by natural selection (Robinson, 2003). The problem with this sort of evolutionary argument is that consciousness could be a spandrel (Gould and Lewontin, 1979). Not all traits of organisms are the result of selection pressure. The sound of a heart beating is a trait many creatures share. However, it seems clear that this trait is not an adaptation, the adaptation is the pump. Evolution selected for the pumping action of the heart. A heartbeat is what you get with functioning organic hearts. Similarly, consciousness could be a free rider on the adaptation of the brain, and so evolution cannot be used to show that epiphenomenalism is false.

In response to this difficulty, Shaun Nichols and Todd Grantham argue that the complexity of a trait is *prima facie* evidence that a given trait is an adaptation and not a spandrel (Nichols and Grantham, 2000). The problem with this argument is that complex traits may very well have complex epiphenomena. For example, Owen Flanagan argues that dreams are spandrels and that while dreaming might have effects on our behavior, it is not an adaptation since evolution did not select for its development. The physical structure of the brain is the adaptation that is functional in the sense that the brain does a set of things that secure fitness, not dreaming (Flanagan, 2000). The same could be said for phenomenal consciousness, complex physical structures can surely have complex epiphenomena, hence the complexity of consciousness is not much evidence that epiphenomenalism is false. In addition,

we need to know what sort of complexity is indicative of adaptation. A heartbeat is complex if you look at the variation in time lapses between beats over time, but (so we assume) the heartbeat is not an adaptation, it is a free-rider on the blood pumping function of the heart.<sup>1</sup>

Given these considerations the fact that a trait is complex is perhaps some evidence that it is an adaptation, but it is not clear how much epistemic weight we should give complexity. In the case of consciousness, the complex adaptation of the brain could very well produce the complex spandrel of phenomenal consciousness and complexity by itself gives no clear reason to think consciousness has causal powers. Williams James' argument that I discuss below is similar to Nichols and Grantham's in that they both use a probabilistic reasoning, rather than one than relying exclusively on thought experiments or the concepts of necessity and possibility. In what follows, I argue there is a strong probabilistic evolutionary argument that epiphenomenal dualism (either substance or property) and certain type-identity theories are false. What we must look for is a pattern in the sort of phenomenal consciousness we experience, not just its complexity, to complete the argument that phenomenal consciousness has causal powers.

### **Past and Present Evolutionary Considerations About the Mind**

Alvin Plantinga's argument against naturalism goes roughly as follows. Given various theories regarding how semantics links up with syntax, it is unlikely that the majority of our beliefs are true, most importantly beliefs in evolution and metaphysical naturalism. For contemporary theories describing the linkage between semantics and syntax, semantics are causally irrelevant; causal powers reside in syntax. Since evolution does not select for things that cannot affect the behavior of the organism, evolution does not select for semantics. Given metaphysical naturalism, there is no preordained matching of syntactic properties with semantic content. Hence, the syntactic properties of our beliefs, which are available for selection, will most likely be out of synch with the content of our beliefs. Our mental lives, given evolution and metaphysical naturalism, would most likely not be veridical. So, we have a defeater for the bulk of our beliefs, including beliefs about the truth of metaphysical naturalism and evolution (Plantinga, 2002, pp. 1-12).

C.S. Lewis had similar thoughts:

If, on the other hand, I swallow the scientific cosmology as a whole, then not only can I not fit in Christianity, but I cannot even fit in science. If minds are wholly dependent on brains, and brains on bio-chemistry, and bio-chemistry (in the long run) on the meaningless flux of atoms, I cannot understand how the thought of those minds should have any more significance than the sound of the wind in the trees. (Lewis, 1962, pp. 164-5)

Epistemological worries about the nature of the mind, given it was designed by evolution, have been with us since Darwin:

...[T]he horrid doubt always arises whether the convictions of man's mind, which has been developed from the mind of lower animals, are of any value or at all trustworthy. Would any one trust in the convictions of monkey's mind, if there are any convictions in such a mind? (Darwin, 1959, p. 285)

---

<sup>1</sup>. Although it should be noted that even if we agree that both dreaming and heart beats are not adaptations, because they have effects in the world they are available to future selection pressure.

Those who assume a naturalistic outlook have also used evolution to draw conclusions about the mind. Herbert Spencer noticed that pleasurable experiences correspond to beneficial states of the organism in relation to evolutionary fitness. The same is true for painful experiences, which tended to correspond to acts that are harmful from an evolutionary point of view (Spencer, 1888, pp. 272-284). As Grant Allen puts the point:

Finally, it may be added, without entering into the ultimate question of the connexion between physical and psychological states, that there seems a certain concinnity and fitness in the correspondence between these feelings and their objective counterparts: in other words, the consciousness of Pain or Discomfort bears somewhat the same relation to other conscious states as the physical fact which underlies it bears to other conditions of the system. (Grant Allen, 1877, p. 20)

### **James' Argument Against Epiphenomenalism**

William James uses the observed alignment between the quality of our experiences and the state of our physical bodies, from the point of view of inclusive fitness, and forms it into an argument against epiphenomenalism in «Are We Automata?» (James, 1879, pp. 17-8).

But if pleasures and pains have no efficacy one does not see (without some such a priori rational harmony as would be scoffed at by the «scientific» champions of the Automaton-theory) why the most noxious acts, such as burning, might not give a thrill of delight, and the most necessary ones, such as breathing, cause agony. (James, 1879, p. 17)

James argues that the epiphenomenalist must accept some sort of parallelism in order to explain the congruence between an organism's actions and the phenomenal consciousness caused by physical events. Since this result would be unsatisfactory to the naturalistically minded, James concludes that we should abandon epiphenomenalism.

...[T]he study *a posteriori* of the *distribution* of consciousness shows it to be exactly such as we might expect in an organ added for the sake of steering a nervous system grown too complex to regulate itself... But, if it is useful, it must be so through its efficaciousness, and the Conscious-Automaton-theory must succumb to the theory of Common Sense. (James, 1879, p. 18)

So, if we accept epiphenomenalism and evolution, the fact that the qualities of our experiences generally align with our interactions with the world becomes extremely surprising. James is basically adding a premise to the standard evolutionary argument for the efficacy of the mental discussed above. William S. Robinson calls this surprising alignment of our behavior with our conscious qualia the «hedonic/utility match» (Robinson, 2006, p. 8). If pleasure and pain were efficacious in producing behavior, we would expect evolution to employ them in keeping organisms away from danger (situations that would decrease fitness) and employ them in pulling organisms towards certain goals and interactions that increase fitness. If mental events have causal powers, it would follow creatures that feel pleasure towards things that are harmful would be selected out of the population, since the motivational direction of pleasure will cause them to seek that which lowers their reproductive fitness. Conscious organisms have had certain phenomenal qualia selected for, and others selected out, such that organisms generally have the «right» phenomenal desires towards events, resulting in fitness increasing behavior.

If consciousness did not have causal powers, if it was epiphenomenal, then evolution could not select for it, and most likely our conscious experiences would not align with our physical interactions. If evolution could not select for qualities of experiences, eating might be painful, and concussions might produce pleasure, etc. So, given that our species evolved

contingently with whatever was available to natural selection at the time on our planet, the probability that brain states are so well aligned, while also being epiphenomenal, is low at best. Since naturalism rejects any sort of parallelism or preordained congruence between minds and bodies to keep minds and bodies in alignment, the truth of evolution and the general alignment of our desires yields the conclusion that epiphenomenalism is probably false.

William James uses much the same sort of evolutionary reasoning as Plantinga's argument against naturalism. Plantinga's argument is regarding beliefs while James' is about consciousness. Plantinga's argument applies to beliefs generally, while James' examples include only pleasure and pains as the data to be explained (I'll discuss this more below). Both use the fact that evolution only selects for things with causal powers as the main lever in their respective arguments. James assumes that evolution and naturalism are true and then evaluates epiphenomenalism, while Plantinga surveys of modern theories of how semantics relates to syntax and uses it as an epistemic lever against evolution and naturalism.<sup>2</sup>

### **James' Argument Generalized**

Here is a response to James' argument that is offered by William Robinson (2006):

Epiphenomenalists can meet James' argument, however, by supposing that both the pleasantness of pleasant feelings and the feelings themselves depend on neural causes (and analogously for painfulness and disliked qualities). So long as both types of neural events are efficacious in the production of behavior, their combination can be selected for, and thus the felicitous alignment of feelings with evaluation can be explained. (Robinson, para. 16)

Robinson follows the examples of James' original argument, delight and agony. I take it that Robinson is arguing that given the equa-probable chance that natural selection would select types of neural events that cause certain behavior that also happen to have the appropriate desire as epiphenomena, the situation we observe is unsurprising. If all we have to account for is two qualia, pleasure and pain, James argument does not give much reason to think epiphenomenalism is false.

James argument, as he presents it in «Are We Automa?» uses pleasure and pain as the only examples. Robinson gives a plausible response to this formulation of the argument. However, I think we can interpret James argument in a much broader way. Pleasure and pain are only examples of a general picture of how our mental lives are related to our physical interactions with the world. In the realm of the desires, our range of feelings is much greater and finer grained than just pleasure and pain. We experience jealousy, burning, itching, tickling, fatigue, humor, rapture, satiation, anguish, catharsis, pride, longing, all of these feelings, the phenomenal aspects are different.

Once we appreciate the richness and variety of our experiences, we can see that the entirety of our phenomenal consciousness generally exhibit the alignment James, Huxley and Allen noticed. Our phenomenal experiences are what we would expect to find if our mental lives had causal powers and could be the subject of selection pressure. If phenomenal

---

<sup>2</sup>. Recently Sharon Street makes a similar argument that she levels against realist theories of values. She argues that it is implausible that evolution would select for creatures that would track objective values. Her reasoning is that objective values would not have casual powers and could not influence evolutionary development. Hence if there are objective values, we are probably massively wrong about their contents, since it is unlikely the values evolution selected for matched objective values (Street, 2006).

consciousness was efficacious, evolution would want us to feel pain when getting burned, so that we avoid being burned. Evolution would want us to be nervous if we detect danger, so that we are ready to fight or flee. Evolution would want sex to feel good, if conscious experience had causal powers, so that we would pursue it. On the belief end, evolution wants (would select for us) to correctly see and hear an approaching predator so we can respond appropriately to it. If phenomenal consciousness has causal powers, not just any phenomenal consciousness experiences will do, only very specific experiences will have the complex content to drive the correct behavior (again, the correct behavior from an evolutionary point of view). In short, our experience of rationality, coupled with Darwinism, only make sense if our conscious experiences can affect the world. Epiphenomenalism has no resources to explain why our mental lives seem to play a causal role in our behavior.

Along with William Robinson, I've used the term 'alignment' to describe how our experiences are related to our bodies' interactions with the world. It is time to cash out this chip:

*Alignment:* A phenomenal experience is aligned if it is of the sort we would expect to find if conscious states had causal powers.

When a phenomenally conscious event is aligned with a physical interaction, the first person experience mirrors the third person description of the interaction. Emotional experiences, such as pleasure and pain, align with the physical when the normative evaluation or normative direction of the experience mirrors the actual or future result of the interaction for the inclusive fitness of the organism. Pain due to bodily injury is aligned because bodily injury tends to decrease fitness. Conversely, if a conscious organism's body is damaged and there is an attending conscious experience with no pain, there is no alignment between the mental and the physical. In the presence of objects in the world that could physically harm an organism, natural selection would select for avoidance behavior. The phenomenal experience of pain is in alignment when it mirrors, all things being equal, the tendency of the interaction to hinder evolutionary fitness. A bear biting your arm is a threat to your fitness. A bear biting your arm is painful, it has the normative direction of badness that mirrors the fitness decreasing act of the bear biting your arm. This makes sense if painfulness, and consciousness in general, is able to cause behavior. The experience of pain tends to interact with various beliefs to form behavior patterns aimed at ending pain. This behavioral direction of pain, the tendency to act with the relief of pain in mind (again all things being equal) is what we would expect of pain could affect our bodies. If phenomenal consciousness were not efficacious, we would not expect this to be the case. Even if consciousness experiences covary one to one with physical interactions, we would expect a random distribution of the mental over the physical, if epiphenomenal dualism were the case.

Sensory experiences are aligned when behavioral abilities with objects match or mirror the perceptions of the object. Imagine testing a subject and determining their abilities in relation to a brick. Their abilities include, say, the ability to see edges, discriminate between colors, correctly infer the distance from the brick, etc. If these behavioral abilities are mirrored in the conscious experiences of the subject, those conscious experiences align with the physical. Functional abilities towards objects are reflected phenomenologically in aligned conscious events. Optical illusions, and visual hallucinations would be examples of conscious experiences that are misaligned, since the experiences inaccurately describe the properties of the objects in the organism's environment and so hinder successful interaction. When I successfully pick up a glass of wine in front of me, if consciousness has causal powers, then

the specific type and distribution of colors and tactile sensations I experience make it seem that they have causal powers. If consciousness has causal efficacy, no other type and distribution would allow for success: a field of blue and the chirping of a bird, or some other set of phenomenal experiences will not allow me to pick up the glass, barring some sort of lucky coincidence. If the type and distribution of phenomenal consciousness does not have a role to play, we would not expect the content of phenomenal consciousness to mirror the biological descriptions of our behavioral abilities. Our sensory experiences and the beliefs that result from them are generally what we would expect if our phenomenal experiences can affect our bodies, given that we are products of natural selection. In our world at least, beliefs flow from phenomenal experiences, the relationship is not random. My experience of a red light in front of me usually results in the belief that there is a red light in front of me. The experience of a red light does not usually form the belief that there is a chicken in front of me. My functional ability to correctly stop at red lights and go at green ones mirrors my phenomenal experiences of the two traffic light states. When aligned, phenomenal experiences tend to map over functional abilities that in turn mirror salient differences in the world that affect inclusive fitness. Again, in the case of hallucinogens and optical illusions, the alignment of the sensory consciousness comes apart. Phenomenal experiences of different colors align if the phenomenal experience matches discrimination abilities with the colored objects. Aligned phenomenal experiences are isometric with the third person description of those abilities. That is, if a conscious organism has the ability to distinguish between salt and pepper and that difference is mirrored in phenomenal experience, that that experience is aligned with the physical. We would expect this to be the case if consciousness had causal powers. Evolution would tend to design creatures that had phenomenal experiences that accurately modeled the salient properties of the outside world. If there is the head of a bear sticking up out of the bushes, we would expect creatures whose contents of phenomenal conscious could affect behavior to have the phenomenal experience of the head of the bear sticking up out of the bushes. Organisms who only experienced a blank field of gray in the same situation and whose phenomenal experiences could cause behavior, would be more likely to be eaten.

Our basic desires, such as for food, sex, and rest are obviously linked to basic fitness enhancing goals. We also have robust abilities to form correct beliefs about medium sized objects with which we directly interact. These belief forming mechanisms are also linked to the fulfillment of these basic biologic goals. These basic desire and belief forming systems are well aligned. Alignment of the mental with the physical grows stronger the more basic the biological fitness goals and belief producing mechanisms. That is, our phenomenal experiences in these areas are what we would expect if the content of our phenomenal consciousness could affect our behavior. To put it another way, compared to the possibility, or conceivability of experiencing a random set of qualia, our normal conscious experiences exhibit a great degree of alignment. Of course, the alignment between the mind and body is not total. We make errors, we brake down, and some interactions are beyond our power to affect the outcome. In organisms built by evolution, we would expect glitches and massive failures under certain conditions to be the consequence of a design full of trade-offs and jury-rigged mechanisms.

The fact that our experiences do not always align with our bodies' physical interactions, as in the case of LSD or optical illusions, tells us that alignment is a relationship to be explained. Alignment is the normal case, but it is not ubiquitous. The question of why some mental/physical events are aligned while others are not calls for an answer. I discuss this occasional misalignment more in a later section.

To sum up this section, the normative directions of our desires often mirror the impact the accompanying physical interaction has or will have for inclusive fitness. Desires that exhibit this property are aligned. For example, damage to the organism, threatening survival and reproduction, produces a negative normative reaction in our consciousness. Similarly, our conscious sensory experiences that are aligned mirror our physical interactions. For the normal driver, the experience of red and green stop lights mirrors the functional fault lines with regards to pressing the brake or gas. If sense experience misses a difference or inserts a difference that is not actually there, the abilities to interact are impoverished in the same way and the phenomenal experiences are misaligned to the same degree. Some readers might be thinking that I'm pointing out the obvious. I hope the idea that the mental and physical generally align will be easy to accept. It represents our normal everyday experience. However, once we articulate our experiences as generally aligned with our bodies' interactions, we can add the diachronic element of evolution and ask whether various models of the relationship between mind and body account for the alignment we experience.

Now, back to the Robinson's objection. Certainly, Robinson has answered the binary interpretation of James' argument. It is certainly possible that the brain processes that cause successful action also cause the phenomenal consciousness of that brain process (Robinson, 2006, p. 9). However, once we look at the problem at a more fine grained level, at the variety and type of our phenomenal consciousness, we see that it is extremely unlikely, unless parallelism is assumed, that the belief/desire experiences, align at all with the what is going on in the world. Our experience is too rich and varied for the scenario Robinson outlines to be plausible.

### **The Story Our Phenomenal Consciousness Tells**

So, what relationship does phenomenal consciousness have to our desires, beliefs and resulting behavior? If phenomenal consciousness is pure sense data that is unavailable for uptake in language, then there is no intentional or motivational content to link up our consciousness experiences with our physical interactions in the world. Ned Block illustrates the position:

...[S]uppose I have an auditory experience as of something overhead, and a simultaneous visual experience as of something overhead... The phenomenal contents of both experiences represent something as being overhead, but there is no common phenomenal quality of the experiences in virtue of which they have this representational commonality... (Block, 1995, pp. 234-5)

We can formulate this as an objection to James' argument. If qualia have no content that links up to our functional and behavioral interactions, then there is no surprising fact of alignment to be explained. There is no surprising alignment because the two things in question, our physical interactions with the world, and our phenomenal consciousness, have no commonality to be compared. As Kim writes:

Second, ordinary sensory concepts, like «pain,» «itch,» and the rest, have motivational/behavioral aspects in addition to qualitative/sensory aspects, and it is clear that the motivational/behavioral component of, say, pain can be given a functional account. (Kim, 2005, p. 170)

In order for James argument to have any force, the 'motivational/behavioral' and the 'qualitative/sensory' aspects must have something to do with one another. Clearly identical qualia can have different intentional content forming directions. Hearing the roar of a lion can produce different intentional contents. If I hear the roar in a zoo, the intentional direction of the roar usually becomes a belief about a real lion. If I hear the same roar in a movie theater,



I will most likely not infer that there is a real lion present and I will instead believe that I'm hearing a recording. The intentional dimension of qualia is contextual to our other beliefs. Also, differing qualia can have the same intentional contents as Block indicates above.

Despite these considerations, surely qualia carry intentional direction, though this relationship is not one of entailment or necessity. Clearly if you vary qualia, while holding everything else constant, content varies and it varies in definite ways. The totality of our mental lives determines our beliefs and desires, but clearly qualia play a role. The formation of our beliefs and desires vary in definite ways with the sorts and combinations of qualia we experience and the formation of beliefs and desires will result in changes to our behavioral dispositions. As David Chalmers writes:

There are deep and fundamental ties between consciousness and cognition. On one side, the contents of our conscious experiences are closely related to the contents of our cognitive states. Whenever one has a green sensation, individuated phenomenally, one has a corresponding green perception, individuated psychologically. On the other side, much cognitive activity can be centered on conscious experience. We know about our experiences, and make judgments about them... These relations between consciousness and cognition are not arbitrary and capricious, but systematic. (Chalmers, 1996, p. 172).

And later:

The primary nexus of the relationship between consciousness and cognition lies in phenomenal judgments. Our conscious experience does not reside in an isolated phenomenal void. We are aware of our experience and its contents, we form judgments about it, and we are led to make claims about it. (Chalmers, 1996, p. 173)

Chalmers uses the term 'phenomenal judgments,' and distinguishes these from beliefs (Chalmers, 2006, pp. 172-3). However, I think it is clear that though Chalmers inserts this middle step between our qualia and our beliefs and desires, he thinks there is a tight relationship between qualia and the resulting beliefs and desires. At this point, I can and should be neutral regarding the nature of this relationship, all I need to show is that our phenomenal experiences, our beliefs and desires, and our behavioral patterns vary in definite ways. That is, whatever the relationship between qualia and intentional states, there certainly is a correlation between what bodies do and what minds experience. When I have the phenomenal experience of a spoon in front of me, all things being equal, I will tend to form beliefs that there is a spoon in front of me and act accordingly. When I feel a sharp pain in my leg, all things being equal, I will form a desire to stop the pain. However, if the quale of pain doesn't have a role in motivation, then varying the quale would not vary the behavior. It is clear, at least in our world, that varying the phenomenal experience, is at least correlated with varying belief and desire formation in definite ways. So, although there is not an entailment relationship between qualia and the content of beliefs and desires, a given set of qualia, tends to form a given set of beliefs and desires.

### **The Alignment Between the Mental and the Physical is Not Total**

The congruence, or alignment, that James, Huxley and Allen noticed is not total. There are times where our phenomenal experiences get things wrong on either the belief or desire end. We often have desires that are contrary to desires that evolution would select for, and often we get things wrong about the world, appearances do not match reality, which usually impedes evolutionary goals. For example, auditory hallucinations due to schizophrenia cause the normal alignment to come apart. The phenomenal experience of a person who is

hallucinating will be out of synch with the world and therefore out of synch with the goals evolution selects for organisms. We can see that someone whose phenomenal experiences are not accurately reflecting properties in the world are usually worse off in terms of inclusive fitness than other members of the population. What this reminder helps elicit is that there is a great degree of congruence between how we experience the world and how we act in it, but this congruence can be upset. However, this picture is what we would expect of a system designed by evolution. Such systems have bugs and glitches, since they are designed with whatever variation produced, to be good enough to survive and reproduce. The tradeoffs and jury-rigged nature of systems designed by evolution result in conscious organisms that have various degrees of alignment with the behaviors and physical interactions that will serve the ends that evolution selects to increase fitness.

We can at least conceive the situation in which our phenomenal consciousness is totally unrelated to our environmental situation (the belief side) and the desires that are at least consistent with furthering evolutionary goals (the desire side). We can now push our imagination a little farther, and imagine a being whose phenomenal life is totally random, and totally unrelated to what is going on in the world. This existence would be a strange one. Such a human in our world would probably not do to well in an evolutionary sense, or any sense at all. But *our* phenomenal lives make sense most of the time, and we can see this by taking note of cases such as LSD and schizophrenia where the normal mental and physical alignment come apart. In our world false beliefs and persons whose burns produce no motivational behaviors are not aligned with the beliefs and desires evolution would select for us. In our world, misaligned minds produce maligned bodies.

### **Are Qualia Functional?**

In *Sweet Dreams* (2005), Daniel Dennett reports on an informal experiment he presents to clarify the nature of qualia. Two similar pictures of a kitchen are flipped back and forth on a screen. It takes most people quite some time to notice the difference in the color of a cabinet panel in the two images. Once the difference is pointed out however, it is readily apparent. Dennett then asks, «...*before you noticed* the panel changing color, were your color qualia for that region changing?» (Dennett, 2005, p. 83). If the answer is no, then your qualia is constituted by how you are disposed to report and behave. If the answer is yes, then you probably think there is logical space, or residue, left after the functional analysis is done (Dennett, 2005, p. 82-91). I think Dennett presents a good test for one's views about qualia. Either qualia are exhausted by some sort of functional/behavioral account, or qualia cannot be entirely, or at all, reduced in a functional/behavioral analysis.

If qualia are thought of as at least partially constituted by this extra-functional residue, then James' argument gains a foothold. For whatever qualia is, whether partially or completely separate from functional/behavioral analyses, the surprising motivational and intentional direction of our experience clearly demands an explanation. Evolution cannot select for this residue, since it is of course not constituted by the casual-functional. As Kim reminds us regarding the 'residue' Dennett speaks of above:

...[T]he mental residue, insofar as it resists physical reduction, remains epiphenomenal. It has no place in the causal structure of the world and no role in its evolution and development. (Kim, 2005, p. 171)

However, if our phenomenal consciousness does not have causal powers that link up with the sorts of experiences we have, cases of misalignment will not be seen by evolution. Evolution

only selects for inclusive fitness, if a physical state increases inclusive fitness, it will be naturally selected. This situation is akin to Plantinga's scenario, where the bulk of our beliefs are false (Plantinga, 2002, pp. 1-12). The difference in the imagined case is that it is our qualia's relation to interactions in our environment under consideration, rather than the truth of our beliefs, although certainly the situations have deep affinities. Under epiphenomenal dualism and metaphysical naturalism, we probably would have random experiences since evolution designed brains without appeal to the contents of phenomenal consciousness.

If qualia cannot be functionally reduced, then whatever dualism you choose, excepting parallelism, the relationship between the mind and body would most likely not result in qualia that followed our functional interactions. Given that epiphenomenal dualism is true, we would not expect the alignment we find, since evolution would only select for physical causal structures, and the causal mechanism between mind and body would not mirror these functional relationships of the physical structures. The mental and the physical would wander completely separate as the shifting interactions of evolution selected for the physical relations and not the mental contents. The probability that naturalism coupled with epiphenomenalism can explain this alignment by chance is not equa-probable, but extremely remote.

### **William Robinson's Second Objection**

William Robinson presents another objection, to James' argument. He gives two possible analyses of the term «pleasure» and argues that on both accounts James' argument has no force. The first analysis is a functional one, «...[F]or any activity of experience, x, to be pleasant is for x to be something we generally pursue without coercion.» (Robinson, 2006, p. 8). Robinson goes on to correctly argue that under this account of pleasant activities or experiences, it is not surprising that:

...[E]volution should favor organisms for which the beneficial and the usually pursued without coercion should go together and, given our dispositional hypothesis about pleasure, it follows directly that the beneficial and the pleasant should go together. (Robinson, 2006, p. 8)

He goes on to argue that «physicalism, epiphenomenalism or interactionism...» has access to this response to James argument and so, if we take the definition of pleasantness above, James argument loses its force against epiphenomenalism (Robinson, 2006, p. 9). This cannot be correct. Epiphenomenal dualism does not identify physical events or states with its mental equivalents; it is a causal relationship between two different things. In this case, there is the mental state, the phenomenal experience of pleasure and the dispositional behavior we use to pick out that mental state. For epiphenomenal dualism, the two are not the same thing, the physical causes the mental. Hence, there is space for the phenomenal quality of the mental state and the physical state to come apart. That is, in order for an epiphenomenalist to account for the alignment between the mental and the physical, there must also be a parallelism so the disposition to usually pursue something actually feels pleasant rather than painful.

We can state this response as a dilemma. Either we define the term «pleasant» behaviorally, in which case there is no need for epiphenomenalism. On the other horn, if mental terms such as pleasantness do not allow for translation into functional terms, then epiphenomenalism, with its one-way causal arrow, gives no reason to think mental events and physical interactions will have any relevance to one another.

Robinson argues «These reflections make it plausible that evolution would have led to a system in which the brain that organizes speech related behavior concurrently with its organization of speech...» (Robinson, 2006, p. 4) and later:

...[W]e should think of the occurrents and the associated behavior as common effects of an organized brain under current stimulation. The view is compatible with the claim that a naturalistic theory of content can be given. It is compatible with holding that the contents of occurrent beliefs and desires depend on facts about how a brain is organized. It is compatible with there being a regular relation between changes in sensory inputs and changes in brain organization that cause both changes in occurrent beliefs and changes in associated behavior. That is to say, it is compatible with holding that changes in beliefs can track changes in sensory inputs. Thus, whatever contributions to content assignment are made by tracking of worldly states by neural event types will be available to epiphenomenalists. (Robinson, 2006, p. 13)

But how is this so? The story would have to go something like this: the content of phenomenal consciousness is assigned by the causal-functional relationships of the organism, including the goals and beliefs of the organism and the world. The story offered *must* be something like this, since evolutionary systems are contingent and responsive to local varying conditions. The same brain state may play different roles in different systems and be assigned different content. Content cannot be antecedently assigned to brain states if you expect content/behavioral alignment. For the pre-assignment of mind/brain events will most likely not match the contingent problems and pressures for which the brain event was a selective fit. This psycho/physical law would be a strange one. It must map phenomenal experience over causal/functional relationships in order to account for the alignment between the mental and the physical. Epiphenomenalism requires constraints on the distribution of mental states over physical states. In normal causal relationships, not all the properties of things are transmitted across causal relationships. In the case of epiphenomenal dualism, the causal relationship holds across very different things. What reason at all is there to think the mental would track the causal-functional relationships of physical states? What reason is there to think that the causal laws involved are ones where the physical states track only certain causal relationships (over, say, the Brownian motion relationships of molecules in the brain) to create mental states?

The epiphenomenalist could respond that all physical interactions cause mental states, but that what we call mental states are only the complex causal interactions of the brain at a certain level. This is a possible response, but then the epiphenomenalist has won the battle but lost the war. For if the epiphenomenalist takes this tack, then they are admitting that at the end of the day physicality causes mentality, and the mental states track the causal functional relationships of the brain. Then what is the motivation for epiphenomenal dualism? It seems like an unnecessary middleman if in the end the mental state is only regurgitating the causal-functional interactions of the brain. For the motivation behind forms of dualism is that our mentality cannot be reduced or explicated in physical or functional terms. The experience of blue is the sort of thing that does not admit of a reductive explanation. No causal explanation of how our brain distinguishes blue things will tell us what it is like to experience blue. However, if our physicality causes mentality, and the content of that mentality is mapped totality over the casual-functional relationships of the brain, there will be no content over and above the causal-functional to satisfy the motivation of dualism. For any content over and above the functional would not be selected by evolution and would probably not align with the physical functional. If phenomenal experience is thought to go over and above the causal functional, then James' argument kicks in, and the alignment of our qualia cannot be explained, since evolution will not have access to feedback regarding misalignment. The

burden of the epiphenomenalist, whether property or substance, is to articulate how or why the physical world causes a mental life that maps over the causal functional.

The relationship that Robinson presents, where mental states track the causal functional relationships of brain states, would be fortunate for us. We would be very lucky for there to be such a causal law, constantly causing mental lives that make sense and presenting phenomenal lives that exhibit alignment. It seems unlikely that existence would include such a law. In any case, such a law certainly isn't needed to get creatures where they are going, the casual structures selected by evolution do just fine getting organisms to track the world and pursue goals. While it is conceptually coherent for there to be such a causal relationship between minds and bodies, either under property or substance dualism, it does seem parallelistic for there to be a causal relationship between the mental and the physical that mirrors the causal-functional interactions that evolution accesses to design and build organisms. It is not plausible to think that that mind/body causal relationship would trace over the causal-functional relationships of the physical world.

David Chalmers articulates the parameters for a theory of consciousness:

For a final theory, we need a set of psycho physical laws analogous to fundamental laws in physics. These fundamental (or basic) laws will be cast at a level connecting basic properties of experience with simple features of the physical world. The laws should be precise, and should together leave no room for underdetermination. When combined with the physical facts about a system, they should enable us to perfectly predict the phenomenal facts about the system. (Chalmers, 2006, p. 277)

The conundrum that Chalmers, and all non-interactive dualists place themselves in is that if qualia do not have completely functional bases, then we could not predict the content that any particular quale would have. That is, there seems to be no reason why any physical structure causes, or has the property of, etc. a certain phenomenal content. As Jaegwon Kim states, if qualia is nonfunctional, we cannot see how to link up a certain physical event with a certain quale (Kim, 2005, pp. 168-69). However, non-interactive dualists cannot explain, while being naturalists and Darwinists, why our qualia are distributed in such a way that makes it seem like our phenomenal experience is efficacious, when evolution cannot select for it.

Chalmers rightly claims that qualia do not shift in our world with identical functional relationships (identical at the level of behavior). He thinks this principle of organizational invariance is a constraint on a final theory of consciousness (Chalmers, 1996, p. 276). We can think of the general pattern of alignment as another constraint. No assignment without alignment. The problem is that under a dualism where the mental cannot affect the physical, the relationship between the mind and body must overlap the causal functional relationships.

The ontology that this leads us to might truly be called a double-aspect ontology. Physics requires information states but cares only about their relations, not their intrinsic nature; phenomenology requires information states, but cares only about the intrinsic nature... We might say that internal aspects of these states are phenomenal, and the external aspects are physical. Or as a slogan: Experience is information from the inside; physics is information from the outside. (Chalmers, 1996, p. 305)

The problem occurs if we couple this with Chalmers claims elsewhere that zombies are possible (Chalmers, 1996, pp. 94-99). For then this causal or lawful relationship between phenomenology and information states that expresses alignment becomes anthropocentric, that

is, a lucky break for us that we should not expect in a naturalistic universe. In a natural universe, we would expect a random distribution of phenomenology over information states.

To put the point in another context, Ned Block and Jerry Fordor have argued that functionalism cannot account for qualia (Block and Fordor, 1972). Sydney Shoemaker responds with the argument that we would not have epistemic access to qualia that was not based on functional relationships (Shoemaker, 1975). However, assuming a non-functional account of qualia, evolution will not select for the type of qualia, since it is in virtue of casual relations that evolution selects. Since as Shoemaker pointed out, physical relations cannot filter out the content of a phenomenal consciousness not based on physical relations (Shoemaker, 1975, p. 255). Chalmers dismisses Shoemaker's objection because he claims that it rests on a causal theory of knowledge. Chalmers goes on to sketch an account of how we access our phenomenal lives that does not rest on a causal account of knowledge (Chalmers, 1996, pp. 196-200). The problem for this sort of account is again that evolution works only with causal structures. However we view Chalmers' version of phenomenal access from the synchronic case, as we evaluate mind/body relationships from the diachronic frame of evolutionary dynamics, Chalmers' account fails. There is no alternative picture of qualia access for evolution, for natural selection has no access. Phenomenal information above and beyond the causal/functional cannot be a candidate for uptake by natural selection.

One objection at this stage is to point out the possibility that qualia might be partially, but not entirely, reducible to functional properties. The relationships between qualia are functionally reducible, but the content is not. The relationships between the colors of a chessboard are functionally reducible, but the colors themselves that fill in between the functional fault lines are not reducible. First of all, I am dubious that one could make sense of the idea that only portions of qualia are functional. However, even if this idea is coherent, the assumption that the functional boundaries would be reflected in the non-reductive aspect of qualia is *ad hoc*. There seems to be no reason why the proportion or aspects of qualia that are functional and non-functional would fit over so nicely over the functional/behavioral cracks. Why is it not the case that the non-functional would invade the functional, leaving us with mis-aligned minds? Again, given a naturalistic universe, this sort of anthropocentric assumption is not one we can allow ourselves.

The epiphenomenalist must argue that our experience of controlling our body is an illusion. However the data that the epiphenomenalist cannot deny is that our experiences are consistent with we would expect to have if it were true that our mental lives could affect our bodies. In short, the mental and the physical exhibit alignment. That is, our conscious experience exhibits the relationship one would expect if evolution created humans and consciousness had causal powers, and could therefore be selected by natural selection. An epiphenomenalism, whether substance or property, cannot account for this alignment, since the content of the consciousness is not available for selection, and hence misalignment would be the predicted outcome. Misalignment would be the predicted outcome since we have no reason to think the causal relationship of the physical to the mental would be one that mirrors causal functional relationships. The other option for the epiphenomenalist is to assume that the type of conscious experience is what it is because it mirrors the functional relationships of the body. This collapses into parallelism since in a naturalistic universe, we would not expect a contingent causal law to hold up a metaphysical mirror to the causal/functional relationships of our bodies. Secondly, the motivation for dualism is lost if the

epiphenomenalist takes this stance, since it is being admitted that the causal/functional accounts for the content of consciousness.

### **Type-Identity Theory**

In this section, I argue that type-identity theory suffers the same fate as epiphenomenal dualism when faced with a generalized version of James argument. My targets in this section are type-identity theories that identify mental events with kinds of physical events. I will use the term ‘type-identity theory’ with the understanding that the types involved are types of physical events. This includes theories such as David Lewis’ functionalism, where functional roles are used as reference fixers (Lewis, 1980).

Thomas Polger claims that identity theory solves the problem of mental causation because, «...[M]ental states ipso facto have the causal powers of physical states.» (Polger, 2004, p. 35). This second statement is true but I will argue that under type-identity theory the specific causal powers of a physical kind have nothing to do with the specific type of mental event it is. For physical kinds, functional powers will vary with the causal context in which it is placed, while the mental kind does not vary. The identities are set and cannot vary, but the uses evolution finds for physical kinds will vary with context and evolution will not take account of the mental side of the equation. Hence, while type-identity theory provides a coherent model of how the mental causes the physical it does not stand up to James’ argument, because it cannot explain the alignment we experience.

Many type-identity theorists hold that types of mental events are identical to kinds of physical events. The composition of physical stuff determines content of the mental event on the other side of the biconditional. One hybrid position is Lewis’ functional identity theory outlined in «Mad Pain, Martian Pain» (1980). Lewis argues that mental states will be picked out functionally, but the identity conditions of the mental event will be of types and will not vary under different functional contexts (Lewis, 1980, p. 218). The sort of phenomenal experience inheres in the physical state, rather than its functional relationships.

Evolution cannot select for the content of the mental state, it only selects for the causal properties of the physical state. Given the contingent and varying nature of evolutionary histories it is highly unlikely that the content of the mental state would align with the casual role of the physical state. The alignment is that the contents of our phenomenal experiences are of the sort that we would expect if phenomenal consciousness had causal powers. Evolution cannot select for matches between the two. Mental states can cause physical states under identity theory, but they cannot cause physical states because of, or in virtue of, the phenomenal content of the qualia. Mental states cause other physical states because of their physical properties and powers. Under type-identity theory the content of phenomenal consciousness is not news evolution can use.

If we did not know how the identities were distributed, that is, if we didn’t know which mental events were identical to which physical events, and we were given a population of rationally behaving creatures, what would we assume about their phenomenal experiences? Surely there are a vast number of phenomenal experiences, and also a vast number of environmental interactions and evolutionary goals. Given this fact, it seems unlikely that the creatures in question would have evolutionary histories which selected for phenomenal contents that aligned with the physical interactions and behaviors which evolution selected. Following the same structure of Plantinga’s thought experiment in his argument against naturalism, we would expect from an identity theory in a given population of rational creatures

is that their phenomenal experiences to be distributed randomly amongst physical events (Plantinga, 2002, pp. 1-12). Their phenomenal experiences would most likely be random since evolution selects for causal powers using whatever variation produces, there is no connection between casual effectiveness and phenomenal experience for the identity theorist. The phenomenal experiences that are identical to say, picking up and sipping a glass of wine, would not, given the local and contingent way evolution selects for brain events find themselves tokened in the situations where conscious creatures found themselves picking up and sipping wine glasses. Type-identity theory combined with naturalism and evolution predicts the wrong result, it predicts that most species will have phenomenal lives that do not have much to do with what their bodies are doing, or what is going on in their environment, or what is good for them from a fitness point of view.

This situation is akin to Fred Dretske's soprano situation. A soprano sings 'break' and the glass breaks, but it did not break because of the content of the word, the glass broke because of the causal relationships of the voice and the glass. (Dretske, 1988, pp. 79-80)

If beliefs and desires explain behavior in this way, then what we believe and desire (the content of our beliefs and desires), however useful it might be for predicting what we are going to do, will not be a part of the explanation of what we do. (Dretske, 1988, p. 81)

Any content of the note would have done the job as long as the changes of air pressure were above a certain frequency and volume. However, it certainly seems like our phenomenal consciousness (the analogue of the content of the soprano's voice) is what causes our behavior, (the glass breaking). But assuming naturalism we should not expect there to be mind/body identities that, luckily for us, align our phenomenal consciousness (the soprano singing «break») with the physical kinds that evolution has selected, (possibly only in our species) to get the job done (the air pressure changes that break the glass). Assuming naturalism and type-identity theory we should not expect the content and the action of mental/physical states to match or align. In the case of the soprano, the content of the note cannot matter. In the same way, the content of the mental state, again assuming type-identity theory, is not under any selection pressure by evolution, which can only select for causal structures that affect fitness. The same physical kinds could very well play different causal roles in different conscious organisms.

The identity theorist, given the truth of naturalism and evolution, is forced to say that mental states cause other mental/physical states because of their physical powers, but not because of the intentional direction of the qualia. The identity theorist might respond: but our phenomenal consciousness does in fact align, our environmental situations and evolutionary goals. Yes, but identity theory has no way to explain this except by an extremely implausible appeal to luck. In our own case, we are merely lucky that the mind/body identities conspire so nicely to give us the phenomenal appearance of having rational lives. We are merely fortunate, according to type-identity theory, that our phenomenal experience mainly tells the story of our interactions with things, and imparts to us the motivations that evolution selected for us to feel. Without this appeal to luck, type-identity theory needs some extra parallelistic bridge law or principle to explain the surprising alignment of the content of our conscious lives with our behavior. However, given the brute nature of identities, an appeal to parallelism does not seem to be an option, since a parallelism seems to imply that the identities could have been different, while the nature of the identity relations precludes this possibility. Causal-role functionalism does explain the general congruence between the mental and physical because phenomenal experiences are constituted by their causal role in the system.



David Lewis (1980) thinks it is possible that the experience of pain could come apart from the behaviors we normally associate with pain. However, while this situation is not the norm, adding the diachronic nature of evolution makes it unreasonable to assume pain would accompany the functional behavioral dispositions we associate with it. David Lewis discusses the possibility of the mad man, who feels pain the way we do, but whose behavior is not what we usually associate with pain. Our qualia of pain could separate from the functional and behavioral dispositions that usually accompany pain. Lewis thinks that our theory of mind cannot make mad pain the norm although he thinks mad pain is a possibility (Lewis, 1980, p. 219). The assumption is that it is clear that mad pain is not the norm and Lewis thinks that one's species provides the standard of pain by which mad pain is not the norm (Lewis, 1980, pp. 219-220). But why is it not the norm?

If pain is identical to a kind of brain event then it will not vary with the functional role in which that type of brain event is placed. However, Lewis also argues that the concept of pain is picked out functionally, Martians can feel pain too (Lewis, 1980, p. 216-7). Evolution designs organisms functionally. If a brain event increases inclusive fitness, it will tend to increase its proportion of the population. Any given mental event, under type-identity theory, while it has causal powers because it is also a physical event, the functional properties will vary with the causal context in which it is placed. The sort of mental event the brain event also is not the basis of selection by evolution. Hence, if pain is identical to a type of brain event, there is no reason to suppose that the contingent way evolution works would select pain to fit the functional role of avoidance and escape.

### **Polger and Dennett on Zombies**

We can place this discussion in the debate over the possibility of zombies. Thomas Polger interprets Daniel Dennett as giving those who think Zombies are possible a dilemma. The dilemma is that either the concept of a Zombie turns out to be incoherent (i.e. the level of sophistication needed for functionally identical zombies makes us realize that they are conscious after all), or the possibility of zombies entails epiphenomenalism of consciousness (Polger, 2004, p. 233).

In order to meet Dennett's challenge Polger goes on to distinguish between two types of epiphenomenalism. There is what is called Huxley's epiphenomenalism, where two mechanisms have different physical effects, but these differences do not alter the function of the system in which they are interchangeable. His example is replacing the carburetor in his car with a fuel injector alters some detectable physical effects but it does not alter that functional system of the car. Both regulate the flow of gas to the engine. They fulfill the same function, but they are different physically, and fulfill their functions in different ways (Polger, 2004, pp. 227-232). «It is a mistake to think that because some mechanism is not required, because it is inessential, that it is epiphenomenal.» (Polger, 2004, p. 227). Polger goes on to remind us that even though a mechanism may be replaceable it still has causal powers that fulfill its function in the system. Hence, mechanisms that can be replaceable are not epiphenomenal if they play the same functional role. The upshot of Polger's argument is that using this distinction we can create a concept of Zombies that answers Dennett's Dilemma. Zombies are possible because we could replace some mechanisms in a brain that preserves the functional identity, but consciousness could be changed or missing. Since the type of phenomenal consciousness, or a lack of, is linked to a type of physical kind, no matter where it is or the role in which it is placed (Polger, 2004, pp. 230-231).

Although Polger secures a place for the causality of the inessential mechanisms, this is not sufficient to avoid epiphenomenalism of phenomenal consciousness and the force of James' argument. If mental items do not get their phenomenal or intentional content from their functional relationships, then the content they have cannot be due to an evolutionary or developmental selection. Therefore, it is quite unlikely, given naturalism and evolution, the contents would align so well with our functional structures. The contents of our phenomenal consciousness do generally align with our functional relationships. Hence, it is unlikely that Polger's Zombies are possible.

A type-identity theorist has no way to connect up the causal powers of a brain state with the content of phenomenal experiences, the other side of the biconditional. For the identification of a phenomenal experience with a brain state does not vary with its causal powers. Evolution only sees (selects for) causal powers. Within a naturalistic framework it is extremely unlikely that the story the phenomenal content of our mental lives tells about our interactions with the world and the goals evolution pressures us to pursue, would match up so well (but not entirely) with the story of what our interactions are really like. For the contingent interactions between organism and world, the unique evolutionary path of each species and each individual, will constitute the selection level from which the brain processes will be selected. Evolution selects for brains that increase inclusive fitness. It selects causal functional answers to environmental problems. It is extremely odd that the identities between the mental and the physical match up so well with our species' evolutionary history in order to elicit the experience of alignment.

Polger's argument is that a type-identity theory can produce a coherent account of zombies that does not make consciousness epiphenomenal. Different mechanisms, with different physical properties, can fulfill the same functions in a rational creature, so zombies are coherent (assuming a type-identity theory) because differing phenomenal experiences may very well hold for each type of mechanism. However, Polger's identity theory combined with evolution has another consequence. It cannot explain why we are in fact not zombies! The mind/body identities are set and do not vary with functional content. While evolution does its work, it does not filter for the mental side of the identity. The identities will have no relation to the contextual, varying, local situations and prior histories that evolution trades on. For example, the type or types of physical events that are identical with pain would most likely not find themselves in the functional role of avoidance of fitness decreasing situations.

Polger's zombie is a coherent creature if take in isolation. However, once we add the diachronic aspect of natural selection, we can see how the *ad hoc*, scavenging, opportunistic selection process will not heed the phenomenal side of the mental/physical identities, only the physical side. The roles evolution finds for brain events, will most likely not make sense, match or align with the mental side. Since this chaotic distribution of phenomenal experience is not what we actually find in the world, type-identity theory is most likely false.

## Objections

One objection might be that in the case of minds, evolution does not really select for events, it selects more at the level of modules. That is, it selects for physical systems that can reliably give the appropriate outputs in lots of different circumstances, rather than an *ad hoc* selection of brain states that produce the right behavior for each case. Evolution relinquishes selection control at the level of individual actions, and instead controls at the level of systems.

Hence, the picture painted of evolution is not realistic enough to draw any conclusions about the metaphysics of mind.

In response, my argument does not depend on any details about the evolution of the human mind. For it does not matter what level evolution selects at, whether at a level of individual brain states, or at the souped-up level of modules that can reliably produce fitness enhancing behavior in novel situations that the species has not yet experienced. The mind/body type-identities are set prior to any evolutionary situations in which they are selected for or against. However the identities are set, whether according to some principle or relation they do not vary according to the wider causal interactions in which they are set, hence it is highly unlikely to expect that they would overlay so nicely over the wider environments in which we act.

Another objection might be that I have overestimated the role that evolution had in determining our behavior. That is, if natural selection in general is not responsible for the structure of the brain, then James' argument fails. In rejoinder, you may insert whatever mechanisms you would like in place of natural selection. Any causal mechanism will not take account of properties that are not causal. Hence, any hodge-podge of mechanisms that brought us to where we are cannot filter out non-causal properties. It is unlikely that alignment would be the norm. Secondly, I can allow that a large amount of behavior is not the result of selection pressure, our basic perceptual experiences and desires such as sexual pleasure, hunger and pain are surely the result of selection pressure and exhibit strong alignment.

### **Teleo-Functionalism**

Teleo-functionalism in the philosophy of mind includes evolutionary history and normativity in the nature of mental states. The content of the mental state is at least in part determined by the historical situations, or future fitness in which the organism evolves. Thomas Polger argues that as a realizer of mental states, an account of the evolutionary history of an organism lacks the causal powers to realize mental states (Polger, 2004, pp. 167-69). Teleo-functional states are not synchronic and so lack the causal powers in the present to produce mental states. Eric Sidel also argues that such teleofunctional accounts make the content of mental states epiphenomenal (Sidel, 2001, pp. 139-166). The causal efficacy resides in the physical structures of the organism when the mental states occur. If the changes in an organism's environment are slow enough, teleofunctional determinants may overlap to a great extent over what organisms experience. However, this is often not the case. While I do not have the space to argue the point, I think it should be clear that if Polger and Sidel's arguments that teleo-functionalism leads to some form of epiphenomenalism are correct, then James' argument provides good reason to think the past and the future do not describe the nature of the mental. Evolution provides the filter that selects various realizers of mental states. However, evolutionary history does not realize the mind. The casual interactions of physical systems perform that task.

### **Conclusion**

Jaegwon Kim clearly states the conundrum that qualia present:

What we miss, something that we need to know in order to design a pain experiencing machine, is a connection between the causal work of the pain box and the arising of pain when the box is activated. Why pain rather than itch or tickle? The machine would try to flee when its skin is punctured even if we had, wittingly or unwittingly, designed itch or tickle in the box. What this shows is that we cannot distinguish pain from itch or other sensations by their causal work; our

strong intuition is that even if pain is associated with scratching behavior (like itch) or squirming behavior (like tickle), as long as it is felt as pain-as long as it hurts- it is pain. Pain may be associated with certain causal tasks, but these tasks do not define or constitute pain. Pain as a sensory quale is not a functional property. In general, qualia are not functional properties. (Kim, 2005, pp. 168-9)

What I have hoped to show is that including evolution in our considerations leads us to the conclusion that the picture of qualia Kim articulates cannot be correct. Assuming a type-identity theory, evolution would not be able to track the contingent evolutionary history of humans. Our qualia would most likely be random and have no relation to our physical environment. Any sort of epiphenomenal dualism falls into the same trap. The ontology of the relationship between the minds and body is either parallelistic, or predicts random qualia experiences as the mental will not feed back into the physical interactions to which evolution has access. If it is not possible to make a machine that we know is feeling pain rather than an itch when it is functionally expressing avoidance behavior, then evolution cannot either. If we cannot functionally reduce qualia, then evolution cannot either and cannot select for it. The only reason the quale of pain is what it is, is because evolution has selected for it to get us to do the right thing by the light of inclusive fitness. The functional characteristics, all things being equal, of pain as an alarm for damage to the organism that will threaten basic evolutionary goals, follows what it feels like. A theory of the mind must specify how mental states get their content, and moreover, that content must match our actual experiences. There can be no content assignment without alignment. The only theory that predicts the alignment we find in the world, given the truth of metaphysical naturalism and Darwinism, is a functionalism where mental states are constituted by causal relationships.

### References

- Allen, G. (1877) *Physiological aesthetics*, (London: Henry S. King & Company)
- Block, N. (1995) On a confusion about a function of consciousness. *Behavioral and brain sciences*, 18, 227-287
- Block, N. & Fodor, J. (1972) What psychological states are not. *Philosophical Review* 81, no. 2, 159-181.
- Chalmers, D. (1996) *The conscious mind: In search of a fundamental theory*. (New York: Oxford UP)
- Darwin, C. (1959) Letter to W. Graham, July 3, 1881. (In F. Darwin (Ed.), *The life and letters of Charles Darwin* (p. 285). New York: Basic Books)
- Dretske, F. (1988). *Explaining behavior*. (Cambridge: MIT Press)
- Dennett D. (2005) *Sweet dreams*. (Cambridge: MIT Press)
- Dennett, D. (1995) *Darwin's dangerous idea*. (New York: Simon and Schuster)
- Flanagan, O. (2000) *Dreaming souls*. (New York: Oxford University Press)
- Gould, S. J. & Lewontin, R. (1979) The spandrels of San Marco and the panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the royal society*, B205, 581-98
- James, W. (1879) Are we automata? *Mind*, 13, 1-22
- Kim, J. (2005) *Physicalism, or something near enough*. (Princeton: Princeton University Press)

- Lewis, D. (1980). Mad pain and martian pain. (In N. Block (Ed.), *Readings in Philosophy of Psychology vol. 1* (pp. 216-222). Cambridge: Harvard University Press.)
- Lewis, C. (1962) Is theology poetry? (In, *They asked for a paper: Papers and addresses* (pp. 150-165). London: G. Bles)
- Nichols, S. and Grantham, T. (2000) Adaptive complexity and phenomenal consciousness. *Philosophy of Science*, 67, 648-670
- Plantinga, A. (2002) «The evolutionary argument against naturalism.» (In, J. Beilby (Ed.), *Naturalism defeated? Essays on Plantinga's Evolutionary argument against naturalism.* (pp. 1-12). Ithaca: Cornell University Press)
- Polger, T. (2004) *Natural minds* (Cambridge: MIT Press)
- Robinson, William (2006). *Evolution and epiphenomenalism*. Retrieved September 15, 2006, from Iowa State University, Department of Philosophy website: <http://www.public.iastate.edu/~wrob/EvoEpi.pdf>
- Robinson, William (2003) *Epiphenomenalism*, Retrieved November 4, 2006, from Stanford encyclopedia of philosophy: <http://plato.stanford.edu/entries/epiphenomenalism/>
- Saidel, E. (2001) Teleosemantics and the epiphenomenality of content. (In J. S. McIntoch (Ed.) *Naturalism evolution and intentionality* (pp. 139-66). (Calgary: University of Calgary Press.)
- Shoemaker, S. (1975) Functionalism and qualia. *Philosophical studies* 27, 291-315
- Spencer, H.(1888) *The principles of psychology, vol. 1* (New York: Appleton and Company)
- Street, Sharon. (2006) A darwinian dilemma for realist theories of value. *Philosophical Studies*, 127, 109-166

**Brett Anderson**  
**[djbraski@yahoo.com](mailto:djbraski@yahoo.com)**

# THE SORITES Charter

· Version 2.2 ·

Copyright © 1995-2006 by the **SORITES** team

---

## PREAMBLE

---

This Charter is an updated and amended version of the original «Copyright Notice and Legal Disclaimer» appeared in Issue # 01 of *SORITES* (April 1995). It is a covenant between the **SORITES** team, as a publisher, and its contributors and readers, prescribing their respective duties and rights, as well as the terms and conditions under which those rights have to be exercised.

Persons dealing with **SORITES** in either capacity, as contributors or as a readers (or users), bind themselves to abiding by these regulations, as does the **SORITES** team itself.

This Charter aims to safeguard all legitimate interests in a balanced way, so as to ensure maximum access to, and use of, the materials we produce, to serve the scholarly public at large, to enhance the prestige of our contributors and the influence of the journal and its publisher, while preventing dishonourable behaviour and academic fraud.

This Charter bears a family resemblance to various schemes: the GNU Free Documentation, Creative Commons, Creative Archive and Open Content licenses. Moreover, the current version borrows a few notions and phrases from them. Nevertheless this Charter stands on its own and must not be interpreted as adjusted to any of those models. Although all those approaches share a strong leaning towards open and free content-divulgence, a distinctive view prevails in each case.

Openness and freedom are broad and loose concepts admitting of sundry conceptions. 'Free content' has been commonly agreed upon to stand for content which readers can and may honestly use without hindrance or threat of reprisals — rather than as content they necessarily get for free — but there are several views as to the justification and implementation of those liberties. A thorough elucidation of those problems goes beyond the scope of this Preamble, though.

In order for the current scheme to stand out as a differentiated dispensation, it will be called '**the allodial approach to intellectual property**'. Four features make it unique:

- The first one is that, while it allots to everybody an unhampered access to intellectual goods **without discrimination**, it also recognizes the right to private intellectual ownership, construed — according to Civil-Law systems — as twofold: an entailed, perpetual, untransferable and non-renounceable entitlement to authorship-recognition and to the integrity of the work; and an alienable right to the exclusive use and profit from that work.
- The second one is that it regards a compilation (such as a journal Issue) as a collective work, and hence as an intellectual entity jointly authored and owned by the Editor and the contributors taken together as a group, while each article, by itself, is an intellectual entity belonging to its author alone.
- The third one is that it resorts to the legal notion of *servitude* or *easement*, which enables us to reconcile all those rights and liberties, in the same way as happens with land freeholds: certain positive or negative interests, franchises or privileges (*servitudes*) may be conferred on non-owners over an estate for certain legitimate reasons and purposes.
- The fourth one is that it rules out any registration requirement, as contrary to the rights to privacy, unobstructed access and uncontrolled dissemination.

\* \* \*

The **SORITES** team will publish revised versions of the **SORITES** Charter, similar in spirit to the present version, but different in detail, to address new problems or concerns. Authors and readers having obtained certain rights in virtue of the current Charter will keep those rights and will never be subject to restrictions brought in by later versions.

Everybody is free to copy this Charter, and to modify or attune it to their own needs, provided appropriate authorship credit is given to the **SORITES** team.

---

## DEFINITIONS

[D01] A **separate material** is an intellectual product which is released, broadcasted or published as an intangible entity standing on its own rather than as a section or a division of another intellectual product.

Parts of an aggregated work — such as essays gathered into a collection — are not separate materials, since they are not released as intellectual products on their own.

[D02] An **Issue of the *SORITES* journal as a whole** is either:

- (i) a single document or file (be it its official version or any unofficial rendering also released by the **SORITES** team as an undivided file, such as the PDF file or a print-out thereof); or
- (ii) a set of files produced by slicing one of the entire-file versions in order to facilitate handling, browsing or downloading (in which case the conveyor is bound to distribute the whole collection taken together at the same time).

[D03] A reader of **SORITES** is an individual who, abiding by all the terms laid down in this Charter, accesses the **SORITES** site, or a mirror thereof, or receives or uses any file or document directly or indirectly proceeding from that site.

[D04] As a rule, institutions and organizations are not readers. They are considered users only when this Charter explicitly confers a certain right on them.

[D05] Private use is personal use by an individual. Sharing downloaded or stored files within an organization is a divulgation, not private use.

[D06] An Issue of the journal *SORITES* is an aggregated work; the articles it contains are not separate materials.

[D07] A copy of a material is either a print-out on paper or a reproduced file or set of files.

[D08] A reproduced file, or a set of files, is either (i) a duplication of the original file, or set of files, as released by the **SORITES** team; or (ii) any result of converting the downloaded file or files to a different format by applying conversion tools, provided the content is not garbled or curtailed.

[D09] A different format is any modification which leaves the content intact and unaltered.

[D10] The phrase 'copy or partial copy' covers any print-out, extended quotation, electronic posting or display as well as any other reproduction.

---

## ARTICLES

Art.01. — **SORITES** is both a resource site, <<http://sorites.org>> (or set of sites, such as <[sorites.info](http://sorites.info)>, <[sorites.net](http://sorites.net)>, etc), and a digital journal, *SORITES*, consisting in a series of issues. Both of them belong to the publisher, the **SORITES** team.

Art.02. — Each Issue of the *SORITES* journal as a whole is a collection, an aggregated work, jointly owned by the contributors and the publisher.

Art.03. — Individual papers remain the private property of their respective authors (the contributors) encumbered with the following servitudes:

- (i) No part of any such paper may be printed or displayed elsewhere, or incorporated into a publication of any sort, unless *SORITES* is therein clearly and explicitly acknowledged and cited as the original source of publication.
- (ii) In particular, when a contributor posts a version of his or her paper on a home page, institutional web-site or subject repository, a link must be made to the version displayed at the **SORITES** web-site.
- (iii) When contributors merge a recast version of a substantial portion of an article published in *SORITES* into another paper, they must also give due credit to the *SORITES* source and clearly inform their audience how to access the **SORITES** site (including a reference to the official web-site, <http://sorites.org>).
- (iv) Contributors adhere to the other terms and provisions contained in this Charter, grant **SORITES** readers and users the rights provided by Art.06 and consent to abide by the publisher's claims and commitments (arts. 04, 13 and 14).
- (v) When contributors republish their papers, they are free to apply any other License, provided clauses (i), (ii) and (iii) of this article are honoured.
- (vi) Any individual who signs a co-authored article is a contributor who accedes to the above burdens being laid on the paper. Arrangements between co-authors cannot stand in the way of that undertaking.

Art.04. — By submitting a manuscript to *SORITES*, contributors declare: that it is an output of their personal research; that it has not been published yet (in the sense of appearing under an ISSN or an ISBN); that it is not currently submitted for publication elsewhere; and that it complies with established Copyright laws. Acceptance of a manuscript by the **SORITES** team is done in good faith under the assumption the declaration is truthful.

Art.05. — The authors of the included papers and the publisher — whether jointly or separately, as the case may be — hereby reserve all rights not conferred to other parties in this Charter.

Art.06. — Subject to honouring the conditions set forth in this Charter, readers are hereby accorded the following rights:

- [i] The right to freely access, view, read, search, store, and link to, any material offered at the **SORITES** site, especially any Issue of the *SORITES* journal, any paper therein included, and any other file or document displayed or supplied at the **SORITES** site.
- [ii] The right to duplicate single detached articles for a personal, private and fair use done in good faith. Fair use encompasses teaching, study, criticism and review.
- [iii] The right to reproduce and distribute on any medium *verbatim* copies of any **separate material** released at the **SORITES** site, provided the material is reproduced **in its entirety** and preserved as an Invariant Document, unaltered in its text, titles and headings.

Art.07. — When a copy of a separate material published by **SORITES** is incorporated into a volume or a package containing other materials, the content obtained from the **SORITES** site must be printed on separate sheets, clearly demarcated, by page breaks, from those accompaniments. No text or logo from other sources must be mingled with, or interpolated into, the **SORITES** material.

Art.08. — The freedom awarded to readers in Art.06 means that:

- (1) No fee will be charged or requested by the **SORITES** team.
- (2) No registration will be demanded.
- (3) No one will be locked out of accessing the **SORITES** site.



- (4) As regards issues of the *SORITES* journal and papers therein published, exercising those rights is not conditional upon reciprocation or remuneration of any sort, whether moral or material, or upon any consideration other than respect for intellectual property claims raised in this Charter.
- (5) As regards other materials offered at the **SORITES** site, moral consideration may be demanded, such as not using those materials for any unethical purpose or activity.

Art.09. —

- [1] Private business companies are not entitled to incorporate a copy of an Issue of the *SORITES* journal into any commercial material they sell for profit without the previous conformity of the **SORITES** team.

However they are hereby allowed to combine a whole Issue of the *SORITES* journal — or any other material offered as a separate unit at the **SORITES** site — into publications they circulate at no charge, provided they observe the other provisions of this Charter, especially Art.07.

- [2] Learned societies and academic establishments are hereby authorized to sell integral copies of issues of the *SORITES* journal or any other material offered as a separate unit at the **SORITES** site.
- [3] Search engines and indexing data-bases are entitled to store, reproduce and display integral or partial copies of any material published by the **SORITES** team, including issues of the *SORITES* journal, even of detached papers, provided due credit and links to the original **SORITES** site (or a mirror thereof) are ostensibly shown.

Art.10. —

- [i] Except as provided by Art.09[3] or by fair use legal provisions, no copy or partial copy of a detached paper may be distributed without its author's acquiescence.
- [ii] The **SORITES** team hereby authorizes any such transmission subject to the author's approval, provided the copyright box on top of the paper is literally incorporated into the copy or partial copy.

Art.11. —

- [i] No part of any Issue of the *SORITES* journal may be delivered to a plurality of individuals — whether in writing or by any other means — unless the source is clearly and explicitly recognized. Furthermore, such a recognition must comply with the following stipulations.
- [ii] No part of any Issue of the *SORITES* journal or of any paper therein included may be conveyed to others by means of reproduction, quotation, copy or paraphrase, without a clear and explicit acknowledgement of the Issue of the *SORITES* journal, the author's name and the paper's full title.
- [iii] Whenever the quotation occurs within a publication, it is also mandatory to mention the date of publication and the official pages (as shown within the Copyright box on top of the paper), as well as the ISSN (1135-1349) and the official web-site: <http://www.sorites.org>, or a mirror thereof such as <http://www.sorites.info>.
- [iv] When the quotation takes place in an electronic medium, a link to the **SORITES** site (or a mirror thereof) is also imperative.

Art.12. —

- (1) Any use of copies or partial copies of an Issue of the *SORITES* journal, or of papers therein included, not permitted in this Charter will be a violation of the contributors' and the publisher's intellectual property.

- (2) Distribution of copies of detached papers without the author's consent is hereby prohibited unless pursuant to Art.09[3].
- (3) Derivative works are unlawful, except in so much as they are provided by Art.09[3].
- (4) Shortening, abridgment or truncation is only licit as provided either by Art.09[3] or else by fair use terms in accordance with intellectual property legislation.
- (5) In any case, the two following activities are absolutely forbidden: 1. change of words or any other semantic or syntactic alteration; 2. false, faulty, obscured or disingenuous attribution.
- (6) The **SORITES** team reserves the right to sue malicious or reckless transgressors of this Charter as culprits of plagiarism, forgery or other misdemeanours, requesting both compensatory and punitive remedies. It also reserves the right to publicly denounce serious copyright violations.

Art.13. —

- [i] Any Material displayed at the **SORITES** site, including issues of the *SORITES* journal, is provided as *is*, without any guarantee of any kind.
- [ii] While the **SORITES** team tries to ensure the correctness of all the content it publishes, it makes no representation or warranty as to the accuracy, completeness or fitness for any purpose of such a content and — to the maximum extent allowed by law — disclaims any pledges or assurances, whether express or implied, to the effect that any published material is free of defects or non-infringing.
- [iii] By furnishing those documents, the **SORITES** team does not underwrite any commitment except in so much as explicitly espoused in the present Charter.
- [iv] Any views expressed in papers published in **SORITES** are the opinions of the contributors and not those of the **SORITES** team. The publisher disowns being answerable for the contributors' ideas or arguments.

Art.14. — Under no circumstances and under no legal theory shall the **SORITES** team, or any contributor or supplier of **SORITES**, be liable to any person for any indirect, special, incidental or consequential damages for loss of good will, mistaken belief, confusion, delusion or any and all other damages or losses suffered as a result of downloading, reading, or using any materials retrieved from the **SORITES** site or included in any Issue of the *SORITES* journal. The user assumes, at their sole risk, full responsibility for relying upon any opinion or information contained in any material published or displayed by **SORITES**.

Art.15. —

- [i] By opening, accessing, downloading, storing, reading or in any other way using any material offered at the **SORITES** site or any part thereof, readers enter into a synallagmatic contract with the **SORITES** team binding themselves to full acceptance of the above stipulations. Breach of any of those provisions entails automatic forfeiture of all rights derived from this Charter and termination of the contract.
- [ii] If, upon downloading or storing a file containing an Issue of the *SORITES* journal or a part thereof, or receiving an electronic or printed copy thereof, or any other material obtained from the **SORITES** site, a reader fails to assent to the conditions stipulated in this Charter, he or she is obliged to immediately discontinue reading, storing, sharing or using the material in any way.
- [iii] In the event of such a termination, though, third-party rights derived from usage prior to the termination will survive.
- [iv] Readers to whom previous versions of this Charter bestowed rights in excess of those granted here, or subject to less stringent conditions, may continue enjoying those rights.

- [v] The **SORITES** team will revise this Charter at its sole discretion. Acquired rights will endure, though.
- [vi] In the event that there exists any inconsistency between this Charter and legal commandments, the latter will take precedence.
- [vii] If a reader cannot satisfy simultaneously the duties under this Charter and commitments stemming from custom, contract or other sources, they must either extricate themselves from those commitments or else refrain from accessing **SORITES** and from using any material released by the **SORITES** site.
- [viii] This Charter is the complete and exclusive agreement between the **SORITES** team, its contributors, and its readers which supersedes any proposal or prior deal, oral or written, and any other communication relating to the subject matter of this Charter.
- Art.16. — Rights granted in this Charter are contingent upon third-party intellectual-property claims.
- Art.17. — In compliance with Spanish Law, all issues of the *SORITES* journal are legally registered, electronic copies being deposited with the competent authorities, namely the Legal-Deposit office of the Autonomous Community of Madrid (Registration: M 14867-1995.)
- Art.18. — Although a mirror of **SORITES** is displayed at the internet host <[www.ifs.csic.es/sorites/](http://www.ifs.csic.es/sorites/)> belonging to the Spanish Academy of Sciences, or C.S.I.C., thanks to a permission graciously given by the system's administrators, the journal is not sponsored or endorsed by that institution, the only owner and publisher being the **SORITES** team.

---

Madrid. 2006-10-05 [V.2.2]

(Previous versions: 1995-04-10; 1996-12-31; 1998-04-27; 2002-10-03; 2004-12-31;  
2005-01-18 [V.2.1])

Lorenzo Peña (The **SORITES** Team)

<<http://sorites.org/copyrigh.htm>>

## *Release Notice*

---

This issue of **SORITES** (Σωρίτης) was intended to come out in June 2008. Owing to a number of practical difficulties the publication has been interrupted for six years and a half.

At long last, the issue is now made available in several formats. Its official version is the PDF document released with filename:

**sorites21.pdf**

Although each version, whether official or not — as initially released on January 12, 2015, by the **SORITES** team — is an entire seamless file, it may be splitted down into chunks in order to facilitate downloading, browsing, transferring or e-mailing. In such cases, the unity of this issue of **SORITES** as a whole must be preserved by keeping the ensuing collection intact.