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])
Logic-and-Law 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**>

Issue # 20 — March 2008

SORITES (Σωρίτης) ®

ISSN 1135-1349

Issue # 20 — March 2008

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

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

<**sorites@sorites.org**>

SORITES (Σωρίτης)

ISSN 1135-1349

ROLL OF REFEREES

SORITES (Σωρίτης)

ISSN 1135-1349

Issue # 20 — march 2008

TABLE OF CONTENTS

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 03-07
Abstracts of the Papers

ABSTRACTS OF THE PAPERS

LOCKE ON 'SUBSTANCE IN GENERAL'

by Matthew Carey Jordan

My goal in this paper is to answer two questions: what, if anything, did Locke have in mind when he spoke of substance in general? and did Locke affirm the existence of substance in general? Concerning the first of these, I argue that what Locke had in mind were bare particulars (or something very closely akin thereto). In the first part of this paper, I show why this interpretation of Locke is preferable to its two main rivals. Concerning the second question, Locke was agnostic about the existence of substance in general. He may not have wished to deny its existence outright, but he certainly did not affirm it. This claim runs counter to most readings of Locke, and I defend it in the second half of the paper. This defense rests on an examination of Locke's view of the relationship between conceivability and possibility, an aspect of the debate that most commentators have overlooked.

QUINE, THE NATURAL STANDPOINT, AND INDETERMINANCY

by M. G. Yoes

Quine's philosophy, early and late, proceeds from the natural standpoint, that is the explicit acceptance of science. This paper attempts to explain what this means and how it fits with his early criticism of reductive empiricism. A kind of horizontal reductionism remains, it is argued, which aims to explain the import of his thesis of the indeterminacy of translation. In the second part of this paper an argument is developed to cast doubt on the significance of this thesis. Because of the possibility of languages containing grue-like predicates, the very idea of determinacy loses all significance. The natural standpoint cannot rescue determinacy and thus cannot provide support for indeterminacy.

TEMPORAL AND COUNTERFACTUAL POSSIBILITY

by Muhammad Ali Khalidi

Counterfactual possibility is often thought to imply and be implied by temporal possibility. In other words, many philosophers subscribe to the following thesis and its converse: If x could have been F in another possible world, then x can become F in the course of time in the actual world. E.J. Lowe (2002) has recently challenged these theses, using two counterexamples. This paper argues that Lowe's counterexamples do not block the inference from temporal to counterfactual possibility, and although they do block it from counterfactual to temporal possibility, they do so only when F is a non-standard «irreversible property». This

raises a question as to whether talk of counterfactual possibility might be replaced by less controversial talk of temporal possibility.

Actualism and the Distinction of Truth over Truth in a World

by Edward R. Moad

Robert Adams characterizes actualism regarding possible worlds as «the view that if there are any true statements in which there are said to be nonactual possible worlds, they must be reducible to statements in which the only things there are said to be are things which there are in the actual world, and which are not identical with nonactual possibles.» In this paper, I will briefly explain actualism about possible worlds, showing that an essential pillar of the theory is the claim that truth is distinct from, and ontologically prior to, truth in a world. The rest of the paper is premised on the idea that an interesting philosophical defense of this claim calls for an analysis of truth itself, and is not intended as an objection to actualism, but rather to underscore the interest actualists should have in the question of what truth is. First, I will consider the idea, drawn from Adams and Alan McMichael, that truth differs from truth in a world in its being a matter of correspondence between a proposition and an independent object; that object being, in McMichael's words, the 'concrete universe'. Then, I will show that, given such an analysis of truth, the truth conditions for propositions about non-actual possibilities violate the central tenet of actualism, as articulated by Adams.

The Constitution Argument Against Conceptualism

by André Abath

According to philosophers such as McDowell and Brewer, the contents of perceptual experience are conceptual. This view came to be known as *Conceptualism*. However, a number of critics have argued that they are wrong in thinking this, for they claim that there is an argument, the so-called Fineness of Grain Argument, which is valid and sound, and has as its consequence the falsity of Conceptualism. Although McDowell and Brewer seem to acknowledge that the Fineness of Grain Argument, if valid and sound, has as its consequence the falsity of Conceptualism, they have ways of answering to the argument. In this paper, I will grant the proponents of Conceptualism that one of their ways of answering to the Fineness of Grain Argument is successful, and that the argument can be blocked. But I will argue that, even if this is the case, we have good reasons to think that Conceptualism is false. For there is another argument, that I will call the Constitution Argument, that has as its conclusion the falsity of Conceptualism. I show that the Constitution Argument is valid and sound, in which case we have good reasons to think that Conceptualism is false.

Free-Will and Determinism: A Debate in Sociology

by Jorge Gibert-Galassi

The matter of this paper is the problem of determinism in social sciences, from the general scientific pretense of achieving constant and univocal connections among events, states of things, as well as ideal objects. The historical obstacle that had placed social sciences in front of that pretense, it has been the fact of the individual freedom and the social contingency,

through the paradoxical question: How can we determine relations between the phenomena of social life if they are contingent in fact?. At an epistemological and theoretical level there is not consensus on a solution to this problem, which has questioned permanently the scientific status of the social disciplines. Some ideas are explored in this respect.

FREE AGENCY AND SELF-ESTEEM

by Robert F. Allen

In this paper I define the role of self-esteem in promoting free agency, in order to meet some objections to the content-neutrality espoused by the reflective acceptance approach to free agency, according to which an agent has acted freely if and only if she would reflectively accept the process by which her motive was formed — in other words, any volition the agent forms is an impetus to a free action just in case she would positively appraise its genesis. For primary self-esteem to exist it is enough to be capable of evaluating oneself, assessing, according to one's personal standards, the reasoning behind one's choices. Freedom lacks a social component; an alienated person may yet be free.

THE INTERPRETIVE MIND

by Peter Francis Colbourne

There are significant physical and intellectual perceptual barriers between our inquiring minds and the phenomena of the extant universe that make our relationship with the external world both complex and problematic. Focusing on scientific processes of inquiry, this paper explores those barriers through (a) a reanalysis of dualism, and (b) an analysis of a form of monism that has arisen out of recent neurobiological research. It is argued that objectivism as the primary principle of scientific inquiry is discredited and should be integrated with the subjectivism of the interpretive mind.

ON THE EMPIRICAL REALITY OF PURE MATHEMATICS

by Jonathan C. Sampson

This study explores the role of probability in an attenuated model of hyperbolic doubt. If formal certainty is withheld from statements about real numbers, then there exists a consistent predicate which varies extraneously by time and subjectivity. In this respect, the empirical factuality of the real number system is unambiguously negated.

NOZICK, PARFIT, AND PLATONIC GLASSES

by Wesley Cooper

The Closest-Continuer *schema* of identity is distinguished here from the Closest-Continuer *theory* of personal identity, the latter applying the former to personal identity by reference to the self's self-defining activity. Nozick's «Platonic glasses» mode of conceptualizing personal identity is defended against Parfit's objections and extended beyond hypothetical branching to the actual branching hypothesized by the «no-collapse» theories of quantum mechanics. The

reader may wish to consult Lev Vaidman's *Stanford Encyclopedia of Philosophy essay*, «Many-Worlds Interpretation of Quantum Mechanics», for an accessible and sympathetic treatment of this interpretation. (Vaidman 2002) See also David Deutsch's philosophical essay on what he calls the «multiverse» in *The Fabric of Reality*. (Deutsch 1997)

KANT AND THE EXPRESSION OF IMPERATIVES

by Ronald Cordero

According to a popular English translation of the *Foundations of the Metaphysics of Morals*, Kant says that all imperatives are expressed by an 'ought.' This, however, seems so clearly to be wrong that it is hard to suppose Kant said it. In this paper I discuss different senses in which imperatives can be said to be expressed and examine in particular the use of Kant's term 'sollen' for such purposes. I argue that what Kant says does not in fact commit him to the position that 'ought' judgments can express imperatives.

KNOWING THAT *P* RATHER THAN *Q*

by Bjørn Jespersen

I offer a two-tiered critique of epistemological contrastivism as developed by Jonathan Schaffer. First, I investigate the cornerstone of contrastivism, the notion of knowing the selected proposition *p* rather than the eliminated, or contrast, proposition *q*. Contrastivism imposes the *ternicity constraint* that the knowledge relation should span a knower and two propositions. However, contrastivism has yet to explain how to square this constraint with the required *contrast* between the selected and the eliminated propositions, and it is not immediately obvious how to accomplish this. I offer up for consideration the binary proposal that to know that *p* rather than *q* is to know that the conjunction of *p* and the negation of *q* is true. Second, I argue that contrastivist objects of knowledge ought to be hyperpropositions rather than functions from possible worlds to truth-values, as assumed by Schaffer.

THE TWO ENVELOPE PARADOX AND USING VARIABLES WITHIN THE EXPECTATION FORMULA

by Eric Schwitzgebel & Josh Dever

The present paper presents a diagnosis of what goes wrong in the reasoning in the «closed envelope» version of what is sometimes called «Two Envelope Paradox» or «Two Envelope Problem» or «Exchange Paradox». Plainly, some constraint on the use of variables within the expectation formula is necessary to escape the paradox. We argue that previous proposed constraints are too restrictive: One can avoid the paradoxical reasoning as long as the conditional expectations of the relevant variables are the same in each event in the partition — provided that all the relevant equations are linear.

HYPOTHESIS TESTING ANALYSIS

by Mikael Eriksson

Logic, as the theory of reasoning, traditionally focuses upon the validity of natural language arguments. During the millennia several logical systems have evolved, each using a specific set of logical constants validating some part of the natural language arguing. Therefore, at the time when reasoning of empirical knowledge entered the scene, it was not surprising to find logical systems having their set of logical constants validating that natural phenomenon. The aim of this paper is to question the strength of such systems and also to sketch a complementary logical system aiming at validating comprehensive empirical knowledge reasoning. In outline, this is done as follows. By listening to scientists discussing empirical issues, there has been historically accepted that they make arguing valid by using the notions of confirmation and falsification in a logical constant manner. Philosopher analyses of this phenomenon have evolved interesting logical systems, however including both insights and paradoxes. The aim here is to use a standard extensional logical system and add what I call a test predicate. The evolving system will derive theorems like empirical knowledge-gaining program having special non-extensional, non-inductive and correlative features suitable for comprehensive empirical knowledge gaining.

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 8-26
Locke on 'Substance in General'

LOCKE ON 'SUBSTANCE IN GENERAL'

by Matthew Carey Jordan

---

1. Introduction

The goal of this paper is to answer two questions: *what, if anything, did Locke have in mind when he spoke of 'substance in general'?* and *did Locke affirm the existence of substance in general?* Concerning the first of these, I will argue that what Locke had in mind were bare particulars (or something very closely akin thereto). In the first part of this paper, I will show why this interpretation of Locke is preferable to its two main rivals. Concerning the second question, Locke was agnostic about the existence of substance in general. He may not have wished to deny its existence outright, but he certainly did not affirm it. This claim runs counter to most readings of Locke, and I will defend it in the second half of the paper. My defense will appeal to a number of texts whose significance for this debate seems to have been overlooked by many commentators. By examining Locke's view of the relationship between conceivability and possibility, in particular, we can do much to elucidate his view of substance in general.

An appropriate way to begin is by presenting two key passages from Locke's *Essay Concerning Human Understanding*. The first is found in a chapter on space, and provides much of the impetus for skeptical readings of Locke on substance:

> They who first ran into the Notion of *Accidents*, as a sort of real Beings, that needed something to inhere in, were forced to find out the word *Substance*, to support them. Had the poor *Indian* Philosopher (who imagined that the Earth also wanted something to bear it up) but thought of this word *Substance*, he needed not to have been at the trouble to find an elephant to support it, and a Tortoise to support his Elephant: The word *Substance* would have done it effectually. And he that enquired, might have taken it for as good an Answer from an *Indian* Philosopher, That *Substance*, without knowing what it is, is that which supports the Earth, as we take it for a sufficient Answer, and good Doctrine, from our *European* Philosophers, That *Substance* without knowing what it is, is that which supports *Accidents*. So that of *Substance*, we have no *Idea* of what it is, but only a confused obscure one of what it does.[1]

The other is from a chapter titled «Of Our Complex *Ideas* of Substances.» Locke writes,

> if any one will examine himself concerning his *Notion of pure Substance in general*, he will find he has no other *Idea* of it at all, but only a Supposition of he knows not what support of such Qualities, which are capable of producing simple *Ideas* in us; which Qualities are commonly called Accidents.[2]

This passage, of course, provides us with Locke's famous definition of substance as

---

1. John Locke, *An Essay Concerning Human Understanding*, ed. Peter H. Nidditch (New York: Oxford University Press, 1975), II.xiii.19.

2. *Essay*, II.xxiii.2.

«something, I know not what.» His use of the word 'something' here, along with some key passages from Locke's letters to Edward Stillingfleet, the bishop of Worcester, have led many commentators to believe that Locke does affirm the existence of substance in general.

Before moving on to the pertinent philosophical issues, a brief comment on Locke's terminology is in order. As with 'idea', Locke's use of 'substance' in the *Essay* is not univocal. My concern here is with what he refers to as 'pure substance in general' in the passage from II.xxiii cited above, *viz.* whatever it is (if anything) that underlies and supports the qualities of a thing.[3] For the sake of convenience, I will often follow Locke in referring to this just as 'substance' or 'substratum'. There is a potential confusion here, however, as Locke also uses 'substance' to mean «a particular thing,» in the sense of an Aristotelian primary substance. Context is usually sufficient to make clear which usage Locke intends, but (as we shall see later) there is at least one important passage in which his meaning is less than obvious.

2. What Locke means by 'substance in general'

Commentators have suggested at least three distinct ways of understanding what Locke meant when he spoke of substance. Some (e.g. Jonathan Bennett) have argued that Locke's conception of substance is merely relational; substance is that which supports qualities, and nothing more can be said. On this interpretation, Locke's substance in general consists of «bare particulars» which do not themselves have properties, hence no positive content is (nor could it be) included in our idea of it. Peter Alexander rejects this view. He maintains that Locke's ontology includes two ultimate, irreducible kinds of substance: material and immaterial. When Locke speaks of substance in general, what he has in mind is one of these two kinds of stuff. Alexander has little to say about the nature of the latter (beyond suggesting that Locke may have held perceptivity to be its defining characteristic), but argues that material substance, for Locke, is essentially solid stuff of which all material things are composed. This is the substance as *general essences* interpretation. It stands in contrast to the third main interpretation of Locke's view: substance as *real essences*. This has been defended by Nicholas Jolley, and hinted at by others (e.g. Michael Ayers, R. S. Woolhouse). The real essences interpretation claims that when Locke speaks of the real essence of a thing, he has in mind its substratum. On this view, 'real essence' and 'substratum' differ in intension but not extension. The two ideas are not equivalent, but in Locke's mind they pick out the same thing.

2.1 Substance as bare particulars

The bare particulars interpretation is the traditional way of reading Locke on substance. It is not difficult to see why. We have already noted two key passages (*Essay*, II.xiii.19 and II.xxiii.2) where Locke explicitly — and exclusively — characterizes substance in terms of that which supports qualities. There are many others, including the following:

> We have no such *clear Idea* [of substance] at all, and therefore signify nothing by the word *Substance*, but only an uncertain supposition of we know not what; (*i.e.* of something whereof we have no particular distinct positive) *Idea*, which we take to be the *substratum*, or support, of those *Ideas* we do know.[4]

---

[3]. To speak of substance as «underlying» and «supporting» qualities is, of course, to be rather vague. These notions will be clarified below; at this stage, however, nothing more precise needs to (or should) be said.

[4]. *Essay*, I.iv.18.

> our *Idea* of Substance, is equally obscure … in both [cases of material as well as immaterial substance]; it is but a supposed, I know not what, to support those *Ideas*, we call Accidents.[5]
>
> [substratum is] *we know not what Support of such Qualities as are capable of producing simple Ideas in us*[6]
>
> Your Lordship [Bishop Stillingfleet] … concludes that there is substance, «because it is a repugnancy to our conceptions of things … that modes or accidents should subsist by themselves;» and I conclude the same thing, because we cannot conceive how sensible qualities should subsist by themselves.[7]

On the basis of texts like these, adherents of the bare particulars interpretation maintain that for Locke, substance in general is understood purely in relational terms. Through sensation and reflection, we come to have ideas of various qualities, both primary and secondary. And because we cannot conceive of such qualities existing «on their own,» we suppose that there must be some *thing* in which they inhere — a thing which itself neither is a quality, nor has any essential qualities. Its nature is exhausted by its function; substance is merely that in which qualities inhere. One contemporary philosopher describes bare particulars as things which do not *have* qualities in the usual sense, rather, they are things to which qualities are *tied*.[8] Others seem more comfortable speaking of bare particulars as having properties, but emphasize that there is no property *F* such that it is essential to a bare particular that it have *F* in order to exist. As Jonathan Bennett puts it, «Lockean substratum-substance *cannot have a 'nature' at all*.»[9]

Besides being *prima facie* plausible,[10] the bare particulars interpretation makes sense in light of Locke's explicit concern to give an «Account of the Ways, whereby our Understandings come to attain those Notions of Things we have.»[11] We do seem to have an idea of substance, albeit an obscure and confused one, and thus Locke must explain where that idea comes from. He need not provide us with an elaborate metaphysical scheme, and except insofar as his epistemology places certain strictures on what can be known to exist, he does not do so. If Locke has bare particulars in mind when he speaks of substance in general, then (since no positive claim at all is made about their essence) he has arguably given an account of the sort necessary for his project to be complete, without engaging in the sort of «Speculations, which, however curious and entertaining,»[12] he generally resists.

---

5. *Essay*, II.xxiii.15.

6. John Locke, *Letter to the Bishop of Worcester* (London: A. and J. Churchill, 1697), text-fiche, pp. 10-11.

7. John Locke, *The Works of John Locke*, vol. 4 (London: Thomas Davison, 1823), 445.

8. J. P. Moreland, *Universals* (Chesham, UK: Acumen, 2001), 153. Moreland takes the «tied to» relation to be unanalyzable. Whatever it is, it differs from the exemplification relation; for Moreland, this is what accounts for bare particulars being genuinely *bare*.

9. Jonathan Bennett, *Locke, Berkeley, Hume: Central Themes* (Oxford: Oxford University Press, 1971), 62. Note the similarity between this statement and Gustav Bergmann's assertion that «Bare particulars neither are nor have natures» (cited in Moreland, 148).

10. I am speaking here of textual/interpretative plausibility, not philosophical plausibility.

11. *Essay*, I.i.2.

12. Ibid.

The account itself, while differing in important respects from the Aristotelian/Scholastic account of knowledge of substances, is rooted in that very tradition. Like many other early modern scholars, Locke's own philosophical training was Scholastic in character, and there can be no doubt that his mature thinking was shaped by it (though there is significant debate about the degree to which it was so shaped). Locke would have been quite familiar with the idea of prime matter, «pure potentiality» which is actualized by substantial forms. It is not at all difficult to see similarities between this doctrine and that of bare particulars which exist when tied to some property or other, but have no properties of their own. Locke's own words on the subject are reminiscent of Aristotle himself. Compare the following two passages (from the *Essay Concerning Human Understanding* and Aristotle's *Metaphysics*, respectively):

> The Mind … takes notice also, that a certain number of these simple *Ideas* go constantly together; which being presumed to belong to one thing … are called so united in one subject … Because, as I have said, not imagining how these simple *Ideas* can subsist by themselves, we accustom our selves, to suppose some *Substratum*, wherein they do subsist, and from which they do result, which therefore we call *Substance* … The *Idea* then we have, to which we give the general name Substance, being nothing, but the supposed, but unknown support of those Qualities, we find existing, which we imagine cannot subsist, *sine re substante*, without something to support them.[13]

> That is why someone might actually be puzzled about whether walking, flourishing, or sitting signifies a being; for none of these either is in its own right or is capable of being separated from substance, but it is more true that the walking or sitting or flourishing *thing* is a being — if indeed it is a being. This latter type of thing is apparently more of a being because it has some definite subject — the substance and the particular — which is discerned in such a predication; for this subject is implied in speaking of the good or sitting thing. Clearly, then, it is because of substance that each of these things is also a being, so that what is in the primary way, what is not something, but is without qualification a being, is substance.[14]

Obviously, these accounts are not identical — for one thing, Locke takes knowledge of qualities to be prior to knowledge (*if* we have such knowledge at all) of substance; Aristotle seems to reverse this order — but they are not terribly dissimilar. Given Locke's own exposure to Aristotle's philosophy and the Scholastic doctrine of prime matter to which it led, as well as the deep similarities between that doctrine and the bare particulars interpretation of Locke, it would not be surprising to find that Locke indeed had something of this sort in mind when he spoke of «substance in general.» Add to this the fact that it is natural to read Locke in this way, and the bare particulars interpretation is on solid footing. Indeed, it seems clear to me that this interpretation is correct.

This is not to say, however, that there are no problems with it. First and foremost, many philosophers have found the idea of bare particulars to be so absurd that it is simply inconceivable that a thinker of Locke's stature would countenance it. The principal philosophical challenge to bare particulars, of course, is that it seems nonsensical to speak of a thing which exists, but which exemplifies no properties. This is compounded by the fact that, on any construal of the doctrine, bare particulars seemingly must exemplify the property of being such that they exemplify no properties. Thus the very notion seems to be incoherent.

This objection need not trouble us, for the following reasons. First, the mere fact that a position appears to be philosophically indefensible does not entail that it is not Locke's view. The principle of charity dictates that we ascribe to Locke the most plausible view consistent

---

13. *Essay*, II.xxiii.1-2.

14. Aristotle, *Metaphyics* 1028a20, my italics.

with his writings, but insofar as those writings strongly lend themselves to a bare particulars interpretation, we have good reason to understand Locke in precisely that way — metaphysical convictions of our own notwithstanding. Second, it has been argued that the doctrine of bare particulars can, in fact, be defended against this sort of objection and is not as untenable as many philosophers seem to think it is.[15] Third, we can modify the interpretation in a way that is faithful to the basic contours of the bare particulars doctrine, but which avoids the charge of incoherence. In all the passages we have seen, Locke discusses our *idea* of substance in general. The content of that idea, as has been noted, is merely relational: substance is that which supports qualities. This does not rule out the possibility, however, that there are other things true of substance *itself*, and that we just don't know what those truths are. It may be that Locke «intends an implicit contrast between our idea of substance and that of an omniscient being … God's idea [of substance] would be whatever enables him to see what it is for properties to be coinstantiated, that is, to be properties of a single thing.»[16] It does not seem to me that such a re-interpretation is necessary, because I do not think that Locke himself took the notion of bare particulars to be incoherent — nor do I think he had mere ideas in mind. Nonetheless, if the incoherence objection is taken to be a serious difficulty for the bare particulars interpretation, then it should be noted that this reading avoids the problem without revising the interpretation in any significant way.

A second problem for the bare particulars interpretation, as Edwin McCann has noted, is that such a position seems incompatible with Locke's corpuscularian view of matter.[17] If material objects are ultimately composed of infinitesimal solid corpuscles, then what metaphysical work is left for bare particulars to do? It is generally agreed that Locke subscribed to Boyle's philosophy of matter, and granted this, it is not at all clear that his ontology has room for bare particulars.

In section 2.3, I will argue that belief in corpuscles is not incompatible with belief in bare particulars. More importantly, however, we should note here that any alleged conflict between Locke's corpuscularianism and his bare particulars view of substance is only problematic if Locke in fact affirmed both. I will argue later (section 3) that he did not, thereby dissolving this problem. For now, it remains to be shown that the two main rivals to the bare particulars interpretation — substance as general essences and substance as real essences — face far more serious difficulties.

2.2 Substance as general essences

Peter Alexander sees Locke as asserting the existence of two ultimate kinds of substance. There is material stuff, which is essentially solid, and there is immaterial stuff, whose essence is less clearly identified.[18] Everything that exists is composed of one of these two kinds of

---

[15]. Offering such a defense lies outside the scope of this paper; one good example is chapter 7 of Moreland's *Universals*.

[16]. Nicholas Jolley, *Leibniz and Locke* (Oxford: Clarendon Press, 1984), 77.

[17]. Edwin McCann, «Locke's Philosophy of Body,» in *The Cambridge Companion to Locke*, ed. Vere Chappell (Cambridge: Cambridge University Press, 1994), 80-1.

[18]. As noted earlier, Alexander sees perceptivity as the most likely candidate for the essence of Lockean immaterial substance; see his *Ideas, Qualities and Corpuscles: Locke and Boyle on the External World* (Cambridge: Cambridge University Press, 1985), 233-4.

substance. This interpretation is not obviously supported by the text, but Alexander contends that when all is said and done, it is the most reasonable way to understand Locke. Indeed, if Locke is to be understood as claiming that there *is* such a thing as substance in general, this view is more attractive than the bare particulars interpretation, if for no other reason than that it avoids the philosophical difficulties associated with that view.

In support of his interpretation, Alexander cites two important passages from Locke's correspondence with Stillingfleet. The first is a text which raises difficulties for the real essences interpretation, which Alexander sees as the only serious alternative to his position:

> my notion of these [real] essences differs a little from your lordship's; for I do not take them to flow from the substance in any created being, but to be in every thing that internal constitution or frame, or modification of the substance, which God in his wisdom and good pleasure thinks fit to give to every particular creature, when he gives a being: and such essences I grant there are in all things that exist.[19]

Here Locke can be understood as saying that there is some stuff which underlies the real essence of a thing; God modifies that stuff — material or immaterial, depending upon the sort of being — and substance thus modified is the real essence of the thing. For obvious reasons, this does seem problematic for anyone who wishes to maintain that Locke identifies substance with real essence. (*Prima facie*, it also appears to be problematic for the bare particulars interpretation — though only if Locke should be read as affirming the actual existence of substance in general.) However, while this passage is certainly *compatible* with Alexander's claim, no mention is made here of there being exactly two ultimate kinds of substance. It is a bit of a stretch to say that the passage supports the substance as general essences interpretation.

The second passage occurs in the context of the thinking matter controversy. Stillingfleet was one of a number of Locke's contemporaries who were troubled by his claim in the *Essay* that «GOD can, if he pleases, superadd to Matter a Faculty of Thinking.»[20] Whether or not this is so is one of the major points of contention in their correspondence, and at one point Locke writes,

> You say, my lord, «you do not set bounds to God's omnipotency: for he may, if he pleases, change a body into an immaterial substance;» *i.e.* take away from a substance the solidity which it had before, and which made it matter, and then give it a faculty of thinking, which it had not before, and which makes it a spirit, the same substance remaining. For if the same substance remains not, body is not changed into an immaterial substance, but the solid substance, and all belonging to it, is annihilated, and an immaterial substance created; which is not a change of one thing into another, but the destroying of one, and making another «de novo.»[21]

According to Alexander, Locke here denies that it would be possible to remove solidity from pure substance in general and then add thought to it; this would be «the *substitution* of one substance for another rather than the *changing* of one substance *into* another.»[22] But a careful reading does not bear this out. Locke's concern in this section of the correspondence is to show that thinking, understood as a power had by (Aristotelian primary) substances, is not

---

19. Locke 1823, 82.

20. *Essay*, IV.iii.6.

21. Locke 1823, 470.

22. Alexander 1985, 231.

incompatible with the quality of solidity. Stillingfleet explicitly affirms the existence of substance in general, and (as Locke points out) he believes that God could transform a material substance into a spiritual substance. Perhaps Stillingfleet is an advocate of bare particulars; the possible scenario for which he allows would then be something like the following. There exists some bare particular *B* which has the property of solidity. An omnipotent God could remove the solidity from *B* without destroying it. God could then give *B* the power of thinking, and the transformation from material to immaterial — according to Stillingfleet — would be complete. Locke, however, takes it one step further. Since Stillingfleet is committed to this state of affairs being possible, Locke uses it to demonstrate the compatibility of thinking with solidity:

> Further, you will not deny, but God can give it [i.e. *B*] solidity, and make it material again. For I conclude it will not be denied, that God can make it again what it was before. Now I crave leave to ask your lordship, why God, having given to this substance the faculty of thinking after solidity was taken from it, cannot restore to it solidity again, without taking away the faculty of thinking? When you have resolved this, my lord, you will have proved it impossible for God's omnipotence to give to a solid substance the faculty of thinking; but till then, not having proved it impossible, and yet denying that God can do it, is to deny that he can do what is in itself possible: which, as I humbly conceive, is visibly to set bounds to God's omnipotency; though you say here, «you do not set bounds to God's omnipotency.»[23]

The first thing to notice is that Locke's point in these passages is not to articulate his own positive doctrine of substance in general, but to show that Stillingfleet's own metaphysical commitments are inconsistent with his (Stillingfleet's) claims about the (im)possibility of thinking matter. Nonetheless, even if Locke were understood to be advocating a view of the nature of substance, the passage cited by Alexander simply does not mean what he claims it does. The key statement is «*if* the same substance remains not, body is not changed into an immaterial substance, but the solid substance, and all belonging to it, is annihilated.» Clearly, Locke is presenting Stillingfleet with a dilemma: either (A) God must destroy the material substance and replace it with an immaterial one, or (B) God may add both the quality of solidity and the power of thinking to one and the same substance. Stillingfleet rejects (A), so he must accept (B). Alexander's interpretation ignores the dialectic, and suffers accordingly.

While no passage in Locke can be cited in direct support of the substance as general essences interpretation, there are several texts which are problematic for it.[24]

> By *general Substance* here, I suppose, your Lordship means the general Idea of Substance: And that which induces me to take the liberty to suppose so, is, that I think your Lordship is here discoursing of the Idea of Substance, and how we come by it. And if your Lordship should mean otherwise, I must take the liberty to deny there is any such thing in *rerum Natura*, as a *general Substance* that exists it self, or makes any thing.[25]

---

[23]. Locke 1823, 471.

[24]. Edwin McCann has argued that the substance as general essences interpretation is ruled out by Locke's commitment to the possibility of thinking matter, since God's superaddition of thought to a material object would necessarily entail that the «object would have two distinct natures, and would belong to each of the two general kinds of substance» (McCann, 80). This objection is mistaken, however, since Locke is quite clear that thought is not an essential attribute of immaterial substance; for Locke, the key contrast is not between *material* and *spiritual* substances, but between *material* and *immaterial* substances. For Locke, thinking is essential to spirits (*qua* spirits), but immateriality is not. On Alexander's interpretation, the two ultimate kinds of substance are material and immaterial, so there is no inconsistency here.

[25]. Locke 1697, 52.

Locke's statements here are not strictly incompatible with Alexander's interpretation, but it seems very odd for him to say something like this if he understands substance in general in terms of general essences. Were that his view, one would expect an amendment to this passage, e.g. «I deny that there is any such thing as *one* general substance which underlies everything, but let us take note that there are exactly *two* such kinds of substance, and everything that exists is composed of one of them.» The fact that Locke refrained from making such a comment — and in such a natural place to do so — strongly suggests that he believed no such thing. Even more difficult for Alexander is Locke's statement that «the general idea of substance [is] the same every where.»[26] Here there is no hint of a distinction between material and immaterial substance, and «given Locke's general carelessness about observing the distinction between ideas and the things they are ideas of, he may be read as saying that substance or substratum is the same in [both].»[27] In the absence of any compelling textual evidence for Alexander's interpretation, this statement by Locke seems to count decisively against it.

2.3 Substance as real essences

Commentators who emphasize Locke's corpuscularianism as a key to understanding his philosophy — and there are many such commentators — often incline toward the substance as real essences interpretation of Locke. On this reading, Locke's distinction between substance in general and real essence is merely conceptual. In Michael Ayers' words, there is a «merely logical distinction … between substance and real essence … There are not two underlying levels, *first* the real essence, *then*, beneath it, the substance.»[28] Since the observable macroproperties of physical objects are, for Locke, determined by their microphysical structures,[29] the appeal of this interpretation is readily apparent. Furthermore, since substance in general is supposed to be whatever it is that supports qualities, and since Locke explicitly speaks of «*real Essences* … on which all the properties of the *Species* depend, and from which alone they all flow,»[30] the case for this interpretation is quite strong. And as with the general essences interpretation, this way of understanding Locke is considerably more attractive, philosophically speaking, than the bare particulars interpretation. The «corpuscular hypothesis … supplied [Locke] with a more significant content for the concept of body than philosophical tradition offered. There is a conceptual necessity about our concept of body which is more

---

[26]. Locke 1823, 33.

[27]. McCann, 80-1.

[28]. Michael R. Ayers, «The Ideas of Power and Substance in Locke's Philosophy,» *Philosophical Quarterly* 25 (1975): 16-7. It should be noted that Ayers has distanced himself from this interpretation in the years since the article was published. In *Locke: Epistemology and Ontology* (New York: Routledge, 1993), he writes, «Now the unknown cause of the union of the observable properties of iron, according to corpuscularian theory, is a particular or determinate constitution or modification of matter: precisely what Locke called the 'real essence' of the species. Yet in general Locke does not seem to have thought of the unknown substance and the unknown real essence of anything as identical» (Vol. II, p. 40). Nonetheless, his earlier work is often cited in support of the real essences interpretation, and with good reason.

[29]. See, e.g. *Essay*, II.xxxii.24: «all that the most expert Man knows, are but few, in comparison of what are really in that Body, and depend upon its internal or essential Constitution…»

[30]. *Essay*, III.v.14.

meaningfully filled by 'insensible particles' than by 'substratum'.»[31] The real essences interpretation also has the advantage of simplicity; it would be explanatorily superfluous to posit the existence of bare particulars *and* real essences, when one thing could do the work of both.

One difficulty for this interpretation, however, is that when Locke speaks of qualities depending on substrata and on real essences, he seems to have two different kinds of dependence in mind. Nicholas Jolley cites II.xxiii.3 of the *Essay*, where Locke writes,

> … therefore when we speak of any sort of Substance, we say it is a *thing* having such or such Qualities, as Body is a *thing* that is extended, figured, and capable of Motion; a Spirit a *thing* capable of thinking; and so Hardness, Friability, and Power to draw Iron, we say, are Qualities to be found in a Loadstone. These, and the like fashions of speaking intimate, that the Substance is supposed always *something* besides the Extension, Figure, Solidity, Motion, Thinking, or other observable *Ideas*, though we know not what it is.

and notes that it is «surprising,» on the real essences interpretation, «that he should have continued to describe the relationship between substance and the observable properties in terms of 'inherence', when what he was actually talking about was causal dependence.»[32] When Locke speaks of our inclination to posit the existence of a substratum as a support for qualities, he consistently uses language of this sort ('inherence', 'subsisting', etc.). We have already seen a number of examples of this in section 2.1. When Locke speaks of real essences, on the other hand, he consistently speaks of qualities which *flow from* those essences:

> How uncertain, and imperfect, would our *Ideas* be of an *Ellipsis*, if we had no other *Idea* of it, but some few of its Properties? Whereas having in our plain *Idea*, the whole Essence of the Figure, we from thence discover those Properties, and demonstratively see how they flow, and are inseparable from it.[33]

> … the Properties that flow from this Essence [of a triangle], are more than can be easily known, or enumerated. So I imagine it is in Substances, their real Essences lie in a little compass; though the Properties flowing from that internal Constitution, are endless.[34]

> … a Figure including a Space between three Lines, is the real, as well as nominal *Essence* of a Triangle; it being not only the abstract *Idea* to which the general Name is annexed, but the very *Essentia*, or Being, of the thing it self, that Foundation from which all its Properties flow, and to which they are all inseparably annexed.[35]

This does not entail that the two dependence relations could not possibly be the same, but if Locke had the substance as real essences interpretation in mind, he certainly could have expressed himself much more clearly. When Locke's words are taken at face value, it seems obvious that substance in general is that in which qualities inhere; qualities depend upon it insofar as they are not things which can subsist of themselves. The dependence here is ontological in character. Real essences, on the other hand, are what determine which observable qualities a thing has; qualities depend upon real essences insofar as a real essences

---

31. John W. Yolton, *Locke and the Compass of Human Understanding* (Cambridge: Cambridge University Press, 1970), 44.

32. Jolley, 86.

33. *Essay*, II.xxxi.11.

34. *Essay*, II.xxxii.24.

35. *Essay*, III.iii.18.

cause things to have such-and-such particular qualities. This sort of dependence is determinate in character. Locke is quite plain about this:

> *Essence* may be taken for the very being of any thing, whereby it is, what it is. And thus the real internal, but generally in Substances, unknown Constitution of Things, whereon their discoverable Qualities depend, may be called their *Essence*.[36]

Thus we are faced with the same textual difficulty that plagues the general essences interpretation: in all of Locke's work, there is no passage in which he actually affirms it, in spite of abundant opportunities to do so. In the words of one commentator, «on the whole, [the substance as real essences interpretation] leaves entirely unexplained large and central stretches of both the *Essay* and the Stillingfleet correspondence in which the notion of substance is discussed.»[37] One might also note that the notion of real essence, while not as hotly contested an issue as substance, is also discussed at length in the *Essay* and in the Stillingfleet correspondence — yet time and time again, Locke fails to make explicit his alleged identification of real essence with substance in general.

Even worse for the real essences interpretation is that Locke does make explicit statements which seem to contradict it. We have already seen that Locke believes there is one idea of substratum, common to all things, but «if substance were identical with real essence, substance in gold would not be the same as substance in silver; what makes them particular substances is their *different* real essences or inner constitutions.»[38] If the real essences interpretation is correct, then Locke's words cannot be taken at face value when he claims that «the general idea of substance [is] the same every where.»[39]

On this point, we should also note a criticism Locke makes of the Aristotelian doctrine of substantial forms:

> Concerning the real Essences of corporeal Substances, (to mention those only,) there are, if I mistake not, two Opinions. The one is of those, who using the Word *Essence*, for they know not what, suppose a certain number of those Essences, according to which, all natural things are made, and wherein they do exactly every one of them partake, and so become of this or that *Species*. The other, and more rational Opinion, is of those, who look on all natural Things to have a real, but unknown Constitution of their insensible Parts, from which flow those sensible Qualities, which serve us to distinguish them one from another, according as we have Occasion to rank them into sorts, under common Denominations. The former of these Opinions, which supposes these *Essences*, as a certain number of Forms or Molds, wherein all natural Things, that exist, are cast, and do

---

36. *Essay*, III.iii.15.

37. McCann, 83.

38. Alexander 1985, 216.

39. Locke 1823, 33. The real essences interpretation cannot be salvaged here by claiming that the «idea of substance» is just the abstract idea of real essences. We have already seen that Lockean real essences are «inner constitutions» or «modifications,» and the passage cited here continues, «the general idea of substance being the same every where, the modification of thinking, or the power of thinking *joined to it*, makes it a spirit, without considering what other modifications it has … on the other side, substance, that *has the modification* or [sic] solidity, is matter» (my emphases). Substance in general is what gets modified; real essences are specific modifications thereof. Also, as will be discussed later, whatever «idea of substance is the same every where» needs to apply to immaterial as well as material things. It is not at all clear that the substance as real essences interpretation can do so.

equally partake, has, I imagine, very much perplexed the Knowledge of natural Things.[40]

There are two things to notice in this passage. First, Locke again makes it clear that real essences determine the sensible qualities of a thing, and therefore determine the species to which a thing belongs. So the substance as real essences interpretation would again contradict the earlier claim that substance is the same in all things. Second, it is interesting to see that Locke refers to something «they know not what.» This phrase is so closely tied to Locke's view of substance in general that it is hard to explain its occurrence here as mere coincidence — especially since the fourth edition of the *Essay* (the one being cited here) was published *after* the controversy with Stillingfleet. Locke certainly knew what this phrase would remind his readers of. If 'they know not what' is meant to conjure thoughts of a ubiquitous something in which qualities inhere, then Locke cannot have been identifying real essence with substance in general.

Textual issues aside, there are other problems for the real essences interpretation. The great appeal of this interpretation is that it makes sense of how Locke's corpuscularianism fits into his doctrine of substance, and with it, how his doctrine of substance fits into his overall project of offering an alternative to the Scholastic and Cartesian views of the world. However, the emphasis on corpuscularianism that drives this interpretation leads, itself, to two serious difficulties. First, it simply pushes the problem of substance in general back a step. We have already seen (in section 2.1) that the problem arises «because we cannot conceive how sensible qualities should subsist by themselves.»[41] Of course, this inconceivability is not unique to *sensible* qualities. The problem cannot be solved by appealing to the physical constitution of material substances, because that would leave us wondering what it is in which *micro*properties inhere. If the fact «that a certain number of … simple *Ideas* go constantly together; which being presumed to belong to one thing … are called so united in one subject, by one name»[42] is a philosophical puzzle when those ideas are the furriness, quadrupedality, and meowing of a cat, then it is also a puzzle when those ideas are the solidity, extension, and shape of a corpuscle. If there needs to be an explanation for the unity of the former set, there needs — it would seem — to be an explanation for the unity of the latter set.[43]

Second, this interpretation seems to ignore that Locke is a dualist. Even if he affirms materialism about human persons, there can be no serious doubt that he acknowledges the existence of at least one immaterial substance: namely, God.[44] And while the substance as real essences interpretations is in line with Locke's philosophy of body, it is very difficult to see how it could be used to make sense out of the relationship between substance in general and particular immaterial substances. The real essence of a body is its microphysical structure.

---

40. *Essay*, III.iii.17.

41. Locke 1823, 445.

42. *Essay*, II.xxiii.1.

43. It should be pointed out that this line of thinking does not commit Locke to the view that there *is* a substratum; as I have mentioned, I take him to be agnostic on that issue — perhaps the problem described here is simply insoluble (at least for us). It does, however, seem to rule out corpuscularianism as a *solution* to the problem as Locke himself describes it.

44. See *Essay*, IV.10.

What, then, is the real essence of an immaterial thing? It is not at all clear how one would begin to answer this question, and it is perhaps telling that defenders of the real essences interpretation tend to focus exclusively on Locke's view of material substances. Perhaps there is an answer available; if so, it escapes me. For now, we may note that Locke was a metaphysical dualist, and any interpretation of his (somewhat opaque) views on *substance* in general must account for his (quite clear) views on substance *in general*. The substance as real essences interpretation fails to do this, and should be rejected.

Thus far, our goal has been to answer the question, *when Locke speaks of 'substance in general', what does he have in mind?* I have argued that what he has in mind is precisely what interpreters have traditionally thought: bare particulars (or something closely akin to them). The other leading interpretations — general and real essences — are implausible, mainly for textual reasons. Neither position is explicitly affirmed by Locke, even though he has abundant opportunities to make his position on the matter clear. Furthermore, both of these positions seem to contradict what Locke actually *does* say on the topic of substance. Turnabout is fair play, however, and in the next part of this paper, I will argue that Locke was agnostic about the existence of substance in general — in spite of the fact that this flatly contradicts a number of his own statements on the topic. My contention is that careful exegesis, along with an understanding of how Locke viewed the relationship between inconceivability and possibility, will be sufficient to show that he did not affirm the existence of substance in general.

3. Locke's agnosticism about the existence of substance in general

In Locke's correspondence with Stillingfleet, the topic of substance (in particular, its knowability) takes center stage. As mentioned earlier, Stillingfleet was a professional clergyman. He had published a book called *A Discourse in Vindication of the Doctrine of the Trinity* in which he accused Locke (though he did not name him outright) of compromising orthodox Christian faith, Trinitarianism in particular. Since Stillingfleet understood divine triunity according to the traditional conception of three persons existing in one *substance*, he saw Locke's sarcastic comments about substance in the *Essay* as an implicit attack on the rationality of the Christian understanding of God. And there can be no doubt that Locke was quite sarcastic in places. We have already noted (in the introduction to this paper) his famous analogy of the Indian who claimed that the Earth was supported by an elephant, the elephant by a tortoise, and the tortoise by «something, I know not what,» just as philosophers are wont to claim that substance is what undergirds the sensible qualities of things. Locke went on to say of «those who lay so much stress on the sound of these two Syllables, *Substance*» that «It helps not our Ignorance, to feign a Knowledge, where we have none, by making a noise with Sounds, without clear and distinct Significations,»[45] and that

> were the Latin words *Inhaerentia* and *Substantia*, put into the plain English ones that answer them, and were called *Sticking on*, and *Under-propping*, they would better discover to us the very great clearness there is in the doctrine of *Substance and Accidents*, and shew of what use they are in deciding of Questions in Philosophy.[46]

Later, in the chapter on substance, Locke writes,

---

[45]. *Essay*, II.xiii.18.

[46]. *Essay*, II.xiii.20.

> here, as in all other cases, where we use Words without having clear and distinct *Ideas*, we talk like Children; who, being questioned, what such a thing is, which they know not, readily give this satisfactory answer, That it is *something*; which in turn signifies no more, when so used, either by Children or Men, but that they know not what; and that the thing they pretend to know, and talk of, is what they have no distinct *Idea* of at all, and so are perfectly ignorant of it, and in the dark.[47]

On the basis of statements like these, Stillingfleet accused Locke of «almost discard[ing] Substance out of the reasonable part of the World.»[48] In his reply to Stillingfleet, Locke is quite anxious to rebut this charge:

> I do not understand what is *almost to discard Substance out of the reasonable part of the World.* If your Lordship means by it, That I deny or doubt that there is in the World any such Thing as Substance, that your Lordship will acquit me of, when your Lordship looks again into that Chapter, which you have cited more than once, where your Lordship will find these Words.[49]

Locke acknowledges that he does not think we can have a clear idea of substance in general, but insists that this is not the same as denying its existence. Indeed, he sees a striking similarity between his own description of substance as something we posit because we can't conceive of qualities subsisting alone, and a statement of Stillingfleet's about such subsistence being «repugnant» to us:

> What now can be more consonant to it self, than what your Lordship and I have said in these two Passages is consonant one to another? Whereupon, my Lord, give me leave, I beseech you, to boast to the World, That what I have said concerning our general Idea of Substance, and the way we come by it, has the Honour to be confirmed by your Lordships Authority.[50]

In light of such statements, it is easy to see why most commentators see Locke as affirming the existence of substance in general, and it is fair to say that there is a burden of proof on anyone who suggests otherwise. Michael Ayers is one who is dismissive of such suggestions:

> Among the many other passages that should help to settle the question, it is worth mentioning Locke's extended, indignant disclaimer in reply to Stillingfleet, who had complained that he seemed to «deny or doubt that there is in the world any such thing as substance». Yet every theorist can resort to epicycles, and if we are prepared to postulate enough insincerity, secret doctrine, ambivalence and confusion on Locke's part, it is just possible to maintain that 2.13.16-20, understood as a scornful rejection of the whole notion of a substrate, represents Locke's true views about substance.[51]

Of course, in the correspondence with Stillingfleet, it cannot be denied that Locke *is* occasionally insincere and sometimes secretive. As Lex Newman has pointed out, «Given the

---

47. *Essay*, II.xxiii.2.

48. Edward Stillingfleet, «The Objections against the Trinity in Point of Reason Answer'd» in *Three Criticisms of Locke* (New York: Georg Olms Verlag, 1987), 234.

49. Locke 1697, 6.

50. Locke 1697, 22. The «two passages» mentioned here are Stillingfleet's statement that «… it is a Repugnancy to our first Conceptions of things, that *Modes* or *Accidents* should subsist by themselves, and therefore the *Rational Idea* of *Substance* is one of the first, and most natural *Ideas* in our minds» (Stillingfleet, 236), and Locke's claim that «'because we cannot conceive how simple ideas of sensible qualities should subsist alone, or one in another, we suppose them existing in, and supported by, some common subject.' Which I, with your lordship, call also substratum» (Locke 1823, 13).

51. Ayers 1975, 3-4.

charges of heresy that lurk, it is understandable that Locke would resort to being cagey.»[52] But we need not attribute to Locke either ambivalence or confusion. In spite of the «disclaimer» referred to by Ayers (and quoted above), Locke is remarkably consistent in his remarks about the existence of substance in general. When pressed by Stillingfleet, he repeatedly agrees that we are unable to conceive of qualities subsisting of themselves. He says (again, as quoted above) that he has the very same reasons for believing in substance as Stillingfleet does.[53] But never does he come out and say, «Yes, Bishop, there *is* a substratum.» Given the amount of space dedicated to this topic, and Stillingfleet's repeated badgering of him on this issue, it would be astonishing if Locke really believed in the existence of substance in general and yet completely failed to say so explicitly. Since he did not say so, it is reasonable to conclude that he did not, in fact, believe in it.[54] There are three main (kinds of) texts in which Locke appears to deny this. The first are passages like the one noted above, where Locke grounds the existence of substance in general on the inconceivability of qualities existing without a support, and draws parallels between this inconceivability and the repugnancy of which Stillingfleet speaks. Many commentators have construed this as an argument to the effect that the existence of a substratum is logically necessary. The second are passages where he distinguishes between agnosticism about the *being* of substance from agnosticism about our *idea* of substance. The third is one isolated passage where Locke comes very close to an explicit affirmation of the existence of a substratum. I will deal with these texts in reverse order.

3.1 Equivocation: 'substance' versus 'substances'

Early in the correspondence with Stillingfleet, there occurs a passage which many have seen as conclusive proof that Locke does believe in substance. He writes, «having every where affirmed and built upon it, That a *Man* is a Substance, I cannot be supposed to question or doubt of the *being of Substance*, till I can question or doubt of my own *being*.»[55] It is surprising, however, that this statement has been taken to be so decisive, since it contains a rather blatant equivocation. The charge against Locke is that he has undercut the rationality of belief in substance in general, not substance*s*, understood as particular beings. Clearly, the latter usage is being employed here. It is obvious from the *Essay* that Locke believes in substances in this sense, and he is correct to say that rejecting the existence of such beings would be tantamount to denying or questioning his own existence. That, however, is not what is at issue; this passage simply changes the subject. We must look elsewhere for evidence that Locke's sarcastic comments about substance in the *Essay* do not capture his real view. Before moving on, however, it is worth noting that Locke's equivocation here is further evidence of his unwillingness to affirm that there is such a thing as substance in general. If he wished to acknowledge its existence, why not say so plainly here, of all places?

---

[52]. Lex Newman, «Locke on the Idea of Substratum,» *Pacific Philosophical Quarterly* 81 (Sept. 2000): 317.

[53]. We shall see later (section 3.3) that this fact is meant to undermine Stillingfleet's position.

[54]. Note, however, that we may conclude merely that Locke *did not affirm* the existence of substance in general. His silence here does not give us reason to think he *denied* its existence; a non-affirmation is not the same thing as an outright rejection. Again, Locke should be read as an agnostic about whether there is such a thing as substance in general.

[55]. Locke 1697, 32.

3.2 'Being' and 'idea'

Immediately prior to the statement just considered, Locke makes a claim which he repeats several times in letters to Stillingfleet. He says,

> The other thing laid to my Charge, is, as if I took the *being of Substance* to be doubtful, or render'd it of by the imperfect and ill-grounded Idea I have given of it. To which I beg leave to say, That I ground not the *being* but the *Idea* of Substance, on our accustoming our selves to suppose some *Substratum*; for 'tis of the *Idea* alone I speak there, and not of the *being of Substance*.[56]

Now, the first thing to notice here is that Locke again refrains from explicitly affirming that there is such a thing as substance. As before, it is fair to wonder why he does so, since it is clear that a single, unambiguous affirmation of belief in substance in general would be sufficient to end the controversy. Second, he acknowledges that our idea of substance is obscure in virtue of its source; neither sensation nor reflection can provide us with a simple idea of substance, and the idea is not «suggested» to us by other ideas, like the ideas of existence and unity are.[57] Locke's point here is that he has not *denied* that there is such a thing as a general substratum, which is certainly correct. A non-denial, however, hardly amounts to an affirmation. And if our *idea* of substance is on such admittedly shaky ground, it is no wonder that Stillingfleet continued to press Locke on this subject. Finally, Locke's consistent use of words like 'accustom' and 'suppose' in this context suggest that he is less than totally confident in the rational justifiability of belief in substance in general.

Stillingfleet had picked up on this last point, asking Locke whether this «*Custom* [is] grounded upon true Reason or not?»[58] In reply, Locke repeats what he had already said about our inability to conceive of qualities subsisting alone, «Which I think is a *true Reason*, because it is the same your Lordship grounds the Supposition of a *Substratum* on, in this very Page; even on the *repugnancy to our Conceptions, that Modes and Accidents should subsist by themselves*.»[59] Clearly, what Stillingfleet wanted to know was whether we are justified in believing in a substratum. Locke avoids this question, instead acknowledging that we cannot help but form such a belief when we reflect on qualities themselves. His point is psychological, not epistemological. Whether this «true reason» is a *good* reason is not addressed. Locke's continued strategy of pussyfooting around the real issue does give *us* good reason, however, to suspect that he is not being entirely forthright about his own views. Stillingfleet is seeking one of two things: either an admission from Locke that the arguments of the *Essay* undercut the rationality of belief in substance in general, or an outright, unequivocal statement to the effect that the existence of substance in general can be established. Locke understands this, but never budges from his position that our idea of substance is grounded in custom and supposition:

> Your lordship indeed tells me, that I say, «that in these and the like fashions of speaking, that the substance is always supposed something;» and grant that I say over and over, that substance is supposed: but that, your lordship says, is not what you looked for, but something in the way of certainty by reason.

---

56. Ibid.

57. *Essay*, II.vii.7.

58. Stillingfleet, 236.

59. Locke 1697, 34.

> What your lordship looks for is not, I find, always easy for me to guess. But what I brought that, and some other passages to the same purposes for, out of my Essay is, that I think they prove, viz. that «I did not discard, nor almost discard substance out of the reasonable world.» For he that supposes in every species of material beings, substance to be always something, doth not discard it out of the world, or deny any such thing to be.[60]

Again, we see the same features of Locke's account emerge: (1) substance is (merely) *supposed*, (2) no admission of certain knowledge concerning substance, and (3) a noncommittal attitude toward the actual existence of substance in general. Newman gets it exactly right when he says that the consequence of these passages «is not to upgrade the certainty of Locke's account of substratum from supposition to demonstrative reason; the effect is to downgrade the certainty which Stillingfleet purports, to that of mere presumption.»[61]

3.3 The logical argument: repugnancy, inconceivability, and impossibility

In sections 2.1 and 3.2 of this paper, we have seen that Locke explains our custom of supposing the existence of a substratum by noting that we cannot conceive of qualities existing «unsupported.» Because Locke speaks of inconceivability, many commentators have understood him to be making a substantive modal claim. As Peter Alexander puts it, «the statement that we cannot imagine qualities existing without support [is] making a logical, rather than a psychological point, that is, [it is] the assertion that it is *inconceivable* because logically impossible that qualities should exist without support.»[62] *Prima facie*, this claim is strengthened by the fact that, as we have seen, Locke claims parity between his notion of inconceivability and Stillingfleet's notion of repugnancy, and the latter clearly *does* ground belief in substance in reason What many commentators have overlooked, however, is that while Locke does not explicitly develop an account of modal epistemology, he does seem to have clear thoughts about what we may (and may not) infer about the world on the basis of inconceivability. Loosely, his view is as follows. There are at least two kinds of inconceivability; we can label them 'positive' and 'negative'. A state of affairs is *positively* inconceivable (PI) when that state of affairs is incompatible with knowledge we have of necessary truths (e.g. the essential properties of geometric shapes). States of affairs which are PI can be known to be impossible. On the other hand, a state of affairs is *negatively* inconceivable (NI) when we lack a clear and distinct idea of how that state of affairs could come to be. Thus positive inconceivability stems from our knowledge, but negative inconceivability stems from our ignorance. If a state of affairs is NI, it may, for all we know, be possible. We certainly cannot know that it is *im*possible, and thus we must be agnostic about its modal status.

Let *S* be the state of affairs described by the proposition 'qualities exist of themselves, without any substratum in which to inhere'. For Locke, *S* is NI. We cannot know *S* to be impossible, and we should be agnostic about whether or not *S* is actual. Therefore Locke is guilty of precisely the charge of which Stillingfleet accuses him: there can be no certainty about the existence of substance in general, and hence no knowledge of it, either. The reason

---

60. Locke 1823, 444-5.

61. Newman, 29.

62. Peter Alexander, «Locke on Substance-In-General,» *Ratio* 22 (1980): 97.

Stillingfleet is unable to make the charge stick, however, is that when he (Stillingfleet) speaks of *S* being repugnant or inconceivable, he means that *S* is PI. Locke avoids the charge of heresy by using the same words as Stillingfleet, but changing their meaning; what *he* means is that *S* is NI.[63] An examination of some key texts will bear this out.

There are two principal contexts in which Locke makes clear his views on positive and negative inconceivability. One is the thinking matter controversy, where he argues that, for all we know, God could create a purely material being with the attribute of thought. The other is his argument for the existence of God and his defense of *ex nihilo* creation. Here is what Locke says:

> We have the *Ideas* of a *Square*, a *Circle*, and *Equality*; and yet, perhaps, shall never be able to find a Circle equal to a Square, and certainly know that it is so. We have the *Ideas* of *Matter* and *Thinking*, but possibly shall never be able to know, whether any mere material Being thinks, or no.[64]

> If it be said, there was a time when no Being had any Knowledge, when that eternal Being was void of all Understanding. I reply, that then it was impossible there should ever have been any Knowledge. It being impossible, that Things wholly void of Knowledge, and operating blindly, and without any perception, should produce a knowing Being, as it is impossible, that a Triangle should make it self three Angles bigger than two right ones. For it is as repugnant to the *Idea* of senseless Matter, that it should put into it self Sense, Perception, and Knowledge, as it is repugnant to the *Idea* of a Triangle, that it should put into it self greater Angles than two right ones.[65]

> But you will say, «Is it not impossible to admit of the *making any thing out of nothing*, since we cannot possibly conceive it? I answer, No … 'tis an overvaluing ourselves, to reduce all to the narrow measure of our Capacities; and to conclude, all things impossible to be done, whose manner of doing exceeds our Comprehension.[66]

Notice that in each of the first two passages, Locke juxtaposes the modal claim in question with a necessary truth from geometry, our paradigm for knowledge of necessity, possibility, and impossibility. In the first one, he counsels agnosticism concerning the existence of thinking, merely material beings. In the second, he states that what we know about thought entitles us to the strong claim that thinking *cannot* come from non-thinking; the denial of this is PI. And in the third passage, he fends off a potential criticism by noting that creation *ex nihilo* is merely NI. It may well be inconceivable, but only because of our own cognitive limitations. In sum, Locke holds that inconceivability is sometimes an indicator of necessary truth, but only in the case of «certain Relations, Habitudes, and Connexions, so visibly included in the Nature of the *Ideas* themselves, that we cannot conceive them separable from them, by any power whatsoever.»[67] On the basis of everything Locke says about qualities, inherence, and substrata, it should be clear that the so-called «logical argument» for the existence of substance in general does not meet Locke's own criteria for logical necessity. We cannot be certain that there is such a thing, and Locke is entitled to nothing stronger than

---

63. This interpretation is similar, though not identical, to that defended by Lex Newman in the paper cited earlier.

64. *Essay*, IV.iii.6.

65. *Essay*, IV.x.5.

66. *Essay*, IV.x.19.

67. *Essay*, IV.iii.29.

agnosticism on the subject. He should be read as affirming precisely that.

3.4 «It cannot be doubted…»

If any doubt remains about Locke's real attitude toward the existence of substance in general, there is one more passage worth noting. It has not received any significant attention in the secondary literature, which is surprising, since it speaks directly to the issue at hand. We have seen that interpreters of Locke who wish to argue that he affirmed the existence of substance in general typically cite his repeated statements of the following kind: «Your lordship … concludes that there is substance, «because it is a repugnancy to our conceptions of things … that modes or accidents should subsist by themselves;» and I conclude the same thing, because we cannot conceive how sensible qualities should subsist by themselves.»[68] Following Lex Newman, I have argued that the comparisons Locke makes between his own views and Stillingfleet's are intended to weaken confidence in the latter, not to strengthen it in the former. There is another comparison made by Locke, however, which seems to settle the matter once and for all. After insisting to Stillingfleet that «the *being of Substance* is not shaken by what I have said,»[69] he goes on to say, «It cannot be doubted but there are distinct Species of separate Spirits, of which yet we have no distinct Ideas at all: It cannot be questioned but Spirits have ways of Communicating their Thoughts, and yet we have no Idea of it at all.»[70] *Prima facie*, Locke appears to be agreeing with Stillingfleet here; there are, he suggests, indubitable truths about which «we have no distinct Ideas at all.» But I do not think it is anachronistic to point out that it *can* be doubted that «there are distinct Species of separate Spirits,» and it *can* be questioned whether «Spirits have ways of communicating their thoughts.» Stillingfleet himself may have been unwilling to subject these claims to criticism, so in that respect Locke's examples are well chosen. But insofar as Locke's beliefs about substance are what is at issue, these examples are quite odd — *if* what he wishes to communicate is a sincere, confident affirmation that there is such a thing as substance in general. Indeed, Locke had already addressed the issue of our knowledge of spirits in the *Essay*:

> we have no certain information, so much as of the Existence of other Spirits, but by revelation. Angels of all sorts are naturally beyond our discovery: And all those intelligences, whereof 'tis likely there are more Orders than of corporeal Substances, are Things, whereof *our natural Faculties give us no certain account at all.*[71]

If the very existence of angels and other spirits is a matter about which we cannot acquire certainty (apart from divine revelation), then surely the same is true of the distinctions between such species and their capacity for interspiritual communication.

This is a very important point, and even at the risk of repetitiveness, the dialectic needs to be made clear. Stillingfleet has accused Locke of undercutting the rationality of belief in substance. This is tantamount to heresy, since if substance is called into question, so — in Stillingfleet's eyes — is the doctrine of the Trinity. Understandably, Locke wants to avoid being labeled a heretic, but he also wants to promulgate his philosophical views, including his

---

68. Locke 1823, 445.

69. Locke 1697, 33.

70. Ibid.

71. *Essay*, IV.iii.27, my emphasis.

agnosticism about the existence of substance in general. So in the correspondence with Stillingfleet, he softens the critical claims he made in the *Essay* and dances around Stillingfleet's accusation, while never actually affirming (or denying) that it is rational to believe in substance. Throughout, he hints at his real view, including in the passage being considered here. The existence of substance in general, he states, is *just as certain* as the existence of «distinct Species of separate Spirits.» That is to say, it is not certain at all.

Matthew Carey Jordan
The Ohio State University
jordan.384@osu.edu

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 27-36
Quine, the Natural Standpoint, and Indeterminancy

QUINE, THE NATURAL STANDPOINT, AND INDETERMINANCY

by M. G. Yoes

Part I: Quine and the Natural Standpoint

1. Quine, following Russell and inverting Husserl, turns his back on the skeptic. He declares that science and especially physics is to be taken a face value and that philosophical problems will be those left by and those created by this assumption. The skeptic's fatal question whether knowledge is possible and the consequent need for the metaphysical underpinnings of a first philosophy, are swept away by this declaration. Presumably science is basically correct, and the new question is how have we, or how could we have, come to know it? This is the natural standpoint, a standpoint from which philosophical problems are chosen, formulated, analyzed and in due course solved.

This is not a naturalism of method. Quine adopts the content of science by taking at face value what it tells us about the world. The philosopher is no outside critic, no skeptic working at arm's length, but a member of the firm, and as such is entitled to use all available resources. The philosopher is expected to deal with the most general and far-ranging problems in terms of these resources: whether numbers or sets exist, for example, or whether dispositionals can be properly analyzed. Here there is neither room nor need for an additional first philosophy; for on this anticartesian view, science is its own first philosophy.

In «Two Dogmas of Empiricism» vertical reductionism is laid waste.[1] The empiricist dogma of confirmation of isolated hypotheses by isolated data together with its corollary, the unrevisability of universally confirmed (that is, analytic) sentences, are rejected. This dogma is replaced by the doctrine of the interconnectedness of all sentences — and not merely all sentences of a given theory, but all sentences of the language as a whole. (This is how the language becomes theory laded and loses its neutrality.) The truth value of any sentence, however empirical or however analytic, is in principle revisable. There is no saying how any evidence will be or should be accommodated; indeed, there is no definable logical relation of confirmation between evidence and hypothesis. Under the new doctrine of interconnectedness of all sentences, the empiricist ideal of a vertical reduction of all statements to a base of sense data statements becomes an impossible dream.

These familiar Quinean doctrines have Quinean consequences. If the overall systematic virtues of conservative connections with the past and inherent simplicity are present and only a few predicted sensory collisions are missed, then a mountain of theoretical terms, or rather their surrogate quantificational apparatus, and their corresponding theoretical entities have a

---

[1]. W. V. Quine, *From a Logical Point of View* (Cambridge: Harvard University Press, l981), Essay 4.

reason for being. Such terms are no longer embarrassing, forever unpaid promissory notes. They have their own respectable posited reality, and reduction to something more solid is neither possible nor desirable. Posits are in, reductions are out.

2. Yet in *Word and Object* and earlier in «The Scope and Language of Science», there appears a kind of *horizontal* reductionism.[2] Thus the philosopher, in the role of full scientific partner, seeks an implementation of scientific results in philosophical research. The philosopher looks for solutions in terms certified by accepted scientific theory, in particular physical theory. The goal is implementation of scientific theory in the task of solving philosophical problems. Indeed, all scientific results are available to the full partner, and, it appears, *only* such results are available.

Thus adopting the natural standpoint entails more than taking science as given and turning away from skeptical questions. It also means taking science as the ultimate source of a necessary certification of notions to be used in philosophizing from the natural standpoint. Again theoretical terms are the life's blood of science, and there can be no vertical reduction of them to sense data or stimulus meaning. But there is an intended or ideal horizontal reduction, or something approaching it, of all other notions to those of high science.

Again this is not a question of method. It is not enough merely to emulate scientific method, just as it is not enough merely to have access to scientific results. One must be *bound* by the results of existing science. There is metaphysics beyond the method. As in logical empiricism, the language of science dominates; but with a difference. In Quine's book there is no independent from-the-ground-up empiricical analysis of the language of science, no empirical explanation of why science makes sense and theology does not. Science just is, and it just is our only touchstone.

This feature of the natural standpoint helps explain a long-standing puzzle for many readers of *Word and Object* who have not forgotten the lessons of «Two Dogmas». Given that theoretical terms are an essential part of scientific procedure, why not treat «analytic», «synonymous», «meaning», and the other terms of traditional philosophical semantics as theoretical terms in a theory of translation or of a theory of language generally? Why is the author of «Two Dogmas of Empiricism», of all people, following a narrow behaviorism, a classical reductivist philosophy? Wasn't it Quine who delivered the fatal blow to reductionist empiricism? Traditional philosophical semantics may turn out to be inferior as a theory, but it seems that the merits of the case are obscured by an arbitrary behaviorism and that the case for philosophical semantics can not get a fair hearing. Anyway, indeterminacy of translation is hardly a startling result, one hears, given behavioristic assumptions.

The natural standpoint explains the unabashed behaviorism. For behaviorism is an obvious application of high science to the problem of translation, a piece of philosophical engineering. As such it is well motivated. We cannot merely introduce theoretical terms as part of a philosophical theory of some subject matter and call them legitimate, however closely scientific methods are followed in building up that theory. Those terms must first somehow be measured against the already legitimate notions of science, whatever those may be.

Chapter II of *Word and Object* is a careful and determined effort to take the measure of

---

2. W. V. Quine, *Ways of Paradox* (New York: Knopf, l966), p.230f; and W. V. Quine, *Word and Object* (Cambridge: MIT Press,l960), p. 22f.

the terms of philosophical semantics in just this way. The outcome is negative but not forgone. The meaning of indeterminacy is that traditional philosophical semantics applied to the problem of radical translation cannot be cashed in scientific-philosophic currency. The evidence for this is that implementation of physics for translation in the behavioristic theory spelled out in Chapter II fails radically to model the traditional theory. The resulting indeterminacy of translation marks the limit of scientific questions about translation and meaning.

While behaviorism is an obvious implementation of the natural standpoint, it is not, as Quine should be the first to admit, uniquely determined by it. Legitimate philosophical questions are no more uniquely answered by the natural standpoint than any other scientific questions. Thus while the behavioristic theory of Chapter II may be an obviously acceptable implementation of physics for language theory, there is no proof that it is the only possible one. The indeterminacy of translation and other results, then, are relative both to this implementation *and* to the natural standpoint generally. Just as there is always room for new discoveries in science, there is always room for new implementations, new engineering solutions.

The overall project in *Word and Object* is not pursued from a perspective outside all theories, or even outside all common sense or scientific theories, as is most of «Two Dogmas» and even some important parts of Chapter I. It is written entirely from the natural standpoint. Since «Two Dogmas» there has been in Quine's work both a growing emphasis on the importance of standpoint and a shift to the natural standpoint. Failing to follow this shift of standpoints may account for the chronic misunderstanding of Quine's project by philosophers who, by tradition and commitment, minimize assumptions and maximize critical questions. *Word and Object*, being written mostly from the natural standpoint, is heavy with presumptions.

3. This flatfooted physicalism may grate on friends of «Two Dogmas « who admire the revolutionary doctrine that all intellectual undertakings from physics and common sense to theology are myth eaten posits and the view of language as an articulated whole.»[3] Some sentences are closer to the intersection of experience and language and are more quickly given up in the face of contrary experience and less easily defeated by theoretical considerations. And some are further away and less influenced by the vagaries of sense but more easily defeated by theoretical considerations. Science apparently has the pragmatic advantage of better predictions; indeed, one might say that this pragmatic advantage is a defining characteristic of science.[4] But the point is that the high line on high science is not prominent in «Two Dogmas». Likewise in much of Chapter I the whole of human theories viewed from outside are mere posits. From that lofty view there is not even a unique best physics, many possible conflicting ones being tied for first place. So, much of Chapter I is just a natural development of «Two Dogmas».

---

[3]. An example might be Putnam in his review of Quine's *Quiddities* in *London Review of Books*, 21 April 1988, p.12f.

[4]. W. V. Quine, *The Pursuit of Truth* (Cambridge: Harvard Univerisity Press, 1990), p. 20: «…when I cite predictions as the checkpoints of science, I do not see it as normative. I see it as defining a particular language game, in Wittgenstein's phrase: the game of science, in contrast to other good language games such as fiction and poetry.»

But not all. Quine recognizes at the end of that chapter that we have no choice but to judge from some standpoint, from within some theory, and it might as well be the best available. What saves us from a relativism in which criticism is pointless is that «… we continue to take seriously our own particular aggregate science, our own particular world-theory or loose fabric of quasi-theories, whatever it may be.»[5] Now it is this shift to the natural standpoint which fixes the frame of reference for the translation theory and the ontology of *Word and Object* and the epistemology and value theory of later work. It is a question of language theory within science, philosophy of language from the natural standpoint. Ontology, epistemology, philosophy of mind, value theory: all are naturalized in turn.

The appearance of the observation sentence does not mark a departure from the holism of «Two Dogmas» after all, a retreat to Protokalsaetze and positivistic ideas about language. The radical holism of that earlier work flows from the standpoint there taken, a minimalist standpoint, as far as possible, outside common sense and scientific theories of the world. From the minimalist standpoint, radical holism remains. On the other hand the role of observation sentences in *Word and Object* is surely nearly implied by the natural standpoint which figures so prominently. The implementation of physics at least for language theory is a precise sort of behaviorism whose aim is to reconstruct, in terms acceptable to physics or high science, as many of our notions about translation as possible. The observation sentence is the payoff.[6] Indeterminacy sets in exactly where physics thus implemented leaves off.

What troubles the friends of «Two Dogmas» and no doubt contributes to widespread misunderstanding is that Quine seems in places to treat the natural standpoint not as method or standpoint but fact. Granted he emphasizes that what appears a hopeless relativism from outside of theory building becomes, from within a given theory, a firm ground for judging truth and avoiding relativism. Internal realism, indeed. But in answering critics who may not share the standpoint and in criticizing others, he does not argue the utility of method but the obviousness of scientific result (no change without a change in subatomic states, etc.) The standpoint is transparent even at the level of philosophical exchange.[7]

This position can be usefully contrasted with that of Hume's Newtonian project for human nature. Hume meant to apply the same or similar techniques, look for confirmation in experience, organize things around one or two central ideas, etc. But he surely did not intend to limit himself to concepts which physical science used, accepted and certified or to the use of well-established scientific theories. Perhaps Hume thought there was one and only one reasonable method, a method based on impressions and ideas, and Newton used it for the physical world and he would use it for the human world. The logical positivists, perhaps following Hume, looked to empirical methods. But Quine begins and ends with scientific results.

Moreover, the much discussed additional indeterminacy of the indeterminacy of translation

---

5. *Word and Object*, p.24.

6. In *Word and Object* (p.44) Quine confesses that the observation sentence *does* seem to be the Protokalsatz, but that it differs in that observationality is a matter of degree. And of course the particular account of language he gives even under the natural standpoint is contextualist and holistic notwithstanding the observation sentence.

7. Perhaps this is evident in Quine's reply to Chomsky in *Words and Objections* (p. 303) and in his review of Nelson Goodman's *Ways of Worldmaking* in the *New York Review of Books*, November 23, 1978.

is additional because the natural standpoint, engineered with the scheme of stimulus, stimulus meaning, and the like, leaves untouched the question of which translation scheme is the right one. It is the burden of Chapter II is to show this, to show that the natural standpoint cannot be implemented in this way so as to yield a scientific account which settles most questions of translation. The natural standpoint encompasses a doctrine that is the major premise to the argument that this indeterminacy is additional: science suitably extended by philosophical work is the test of what can be, of what can be true and of what can be a matter of fact. Most questions of translation fail the test and therefore are not questions of fact at all.

4. It is hoped that much of this reconstruction will sound right to Quine's readers. But the key idea of horizontal reduction may not. For the careful reader will know that in places Quine seems explicitly to reject reduction to physics. In his review of Goodman's *Ways of Worldmaking* he objects to Goodman's acceptance of a large patchwork of worlds or world versions, from physics through common sense and on to worlds created by the arts.[8] He sees Goodman's acceptance of multiple mutually exclusive physical theories, which is consonant with his own view, as leading Goodman unnecessarily down a slippery slope to a multiplicity of strange worlds: Quine would stop relativization at physical theory.

But for the maker of many worlds there appears no reason to give physical theory a special place. There can be multiple equally right conflicting physical theories. There can also be multiple equally right conflicting or merely incommensurable versions of all kinds, from physical theories to common sense or artistic versions. Physical theory is but one among many world versions which can be right. Now the answer to this, Quine says, is «… *not* that everything worth saying can be translated into the technical vocabulary of physics; not even that all good science be translated into that vocabulary.»[9] The special standing of physical theory turns, rather, on its central task of universal coverage: nothing happens without «… some redistribution of microphysical states.»[110] Any putative counterexample to this merely calls for a revision in physical theory. Thus the job description of physics bestows a special standing, a standing which only those who reject the enterprise of physics altogether can deny.

From outside the natural standpoint one can see many competing, equally good, irreducible, conflicting physical theories; but the relativism of **outside** falls before the immanent notions of truth and matter of fact of **inside**, inside the physical theory bequeathed us by history and tradition. This standpoint is privileged since it alone answers the need for universal coverage. No version of any other kind can even compete in this arena since any real competitor is bound to be another physics. This move of Quine's invites the reply that job definitions are neither analytic nor necessary by Quine's own doctrines. No untenable dualism in language theory; but likewise none in theory of science as well.

Philosophers who think there is more in heaven and earth that is dreamt of in a complete physics face no challenge from Quine to produce physical definitions or to produce translations into physical theory. There might be much «worth saying» even in another science which fails in principle to be translatable — worth saying, to be sure, but factually empty. One is reminded of tamer logical positivists who proclaim ethics cognitively meaningless

---

8. See *Theories and Things* (Harvard University Press, 1981), pp. 97f.

9. W. V. Quine, *Theories and Things* (Harvard, 1981), p. 98.

10. *ibid.*, p. 98.

notwithstanding its importance (emotional or otherwise). Indeed one reason Quine may have for not wanting to call this reductionism is that he does not seem to believe that indeterminacy has consequences for the **practice** of translation or that arguments against the sensibility of common psychological language imply that such talk should be abandoned.

Still it seems not far wrong to say that the physicalism espoused in the claim that nothing happens without a redistribution of underlying microphysical states speaks of a kind of reductionism even if it is of neither the term-by-term nor the translational sort. For as Quine sees it this implies that from within the natural standpoint matters of fact are determined by distributions of microphysical states and only by such distributions. Thus if conflicting manuals of translation are physically equivalent in the sense of resting on no distinct distributions of microphysical states, then the translation is indeterminate, and there is no fact of the matter which separates them. It is certainly a kind of ontological reduction.

The philosopher's job, from the natural standpoint, is to implement physics or high science for the solution of philosophical problems; and where there is impossibility of implementation there is indeterminacy; and where there is indeterminacy, there is no fact. In seeking implementations one is doing what any scientist may do, which is to use the results science has so far established; and insofar as one wants a *factual* philosophical outcome, one must use *only* such results. Because of physic's special job, its results are the ultimate results. They alone determine the facts.

If this account is correct, then perhaps Quine's philosophy can be more directly understood. The natural standpoint is a unifying principle. Not that it all comes crashing down when the natural standpoint is rejected. Much would remain, of course. Perhaps sweeping away the metaphilosophy would leave the philosophy largely in place. In any case this naturalism is not easily disposed of. What are the alternatives?

Part II: Gruese and Indeterminancy

5. The role of the natural standpoint and the consequent method of horizontal reduction are central in Quine's philosophy. That is the burden of these observations so far. Are there any problems arising out of these doctrines? Indeed there are. What follows is an application of the doctrine, namely the famous thesis of the indeterminancy of translation, and a discussion of a serious problem it reveals.

Radical translation is translation from an independent language, the language of the Other. Thus for all the translator knows the language of the Other contains predicates like «grue». What bearing does this possibility have on the doctrine of indeterminacy of translation?

The Other speaks Gruese. We would say, characterizing the situtatuon in our language, that the Other's occasion sentence «Grue» prompts the Other to affirm when presented with a green stimulus before t and prompts the Other to affirm when presented with a blue stimulus at some time later than t.[11] The affirmative stimulus meaning of «Grue» for all speakers of Gruese is at each time the same: for all times before some time t, it includes all green stimulations but no blue ones; and for all times after t, it includes all blue stimulations but no green ones. Thus, the affirmative stimulus meaning of «Grue» contains grue stimuluations and

[11]. Of course Goodman's definition is formulated in terms of things examined before t and things not so examined. He is after confirmation. Here, since the subject is stimulus meaning, being prompted by a stimulus replaces examinining an object for color.

the negative stimulus meaning of «Grue» contains only stimulations other than grue. The observationality of «Grue» may be high, as high as «Green» in English.

The translator who translates «Grue» as «Green» makes an error if the time is not right. The Other who translates «Green» as «Grue» makes an error if the time is not right. To say this defends the **determinacy** of the translation of «Grue», a defense which is necessary for Quine since in his translation theory all observational sentences are subject to determinate translation. If there is a language like Gruese, like all languages, all of its observation sentences have a determinate translation.

By Quine's account, then, the Other's sentence «Grue» is subject to determinate translation. It is a solid fact that «Grue» translates into some observation sentence of English. This is not to say that translators can make no errors with these sentences. Determinacy is not incorrigibility. A translator may collect his evidence on matters of stimulus meaning with great care, generalize from the evidence with exquisite subtlety, and in the end be mistaken. The point, Quine reminds us, is that in the case of observation sentences there is something to be mistaken about. Since «Grue» is observational in Gruese it is a scientific fact that any observational sentence of English whose stimulus meaning is the same as «Grue» is a correct translation of «Grue».

6. The thesis of indeterminacy of translation is naturally of much philosophical interest. Perhaps it would not be far wrong to compare its importance with that of other famous negative theses such as the theorems of Goedel and Church and the indeterminacy principle of Heisenberg. And just as the fascination of Goedel's theorem depends on there being completeable systems, indeterminacy gains its interest by contrast with determinacy. If there were no significant class of computable functions, there would be less interest in the theorem that establishes the existence of noncomputable ones. The careful constructions of Chapter II of *Word and Object* — stimulus, stimulation, modulus of stimulation, stimulus meaning, observation sentence, etc. — are all in service of finding a precise, scientifically certified sense of determinant translation. How far does such a precisely defined determinacy take us? The thesis of indeterminancy of translation is that this determinancy cannot take us all the way, that in the end a large area of indeterminacy must remain. But observational «Grue» presents a difficulty for that determinacy, a difficulty which subsequent paragraphs explore. If this difficulty cannot be met, the importance of the indeterminacy of translation thesis may be compromised.

7. Though thoroughly observational, the determinacy of «Grue» falls far short of the determinacy of Quine's standard examples. Occasion sentences with little variation in stimulus meaning within the linguistic community are sentences whose stimulus meaning is little effected by extraneous clues, collateral information. «Gavagi»'s stimulus meaning may vary from expert hunter, to stay-at-home novice who knows nothing of gavagis. The Other's word for bachelor, if there is one, varies widely in stimulus meaning across the linghistic community. Technical and scientific terms may vary widely and have thin stimulus meanings. Color words, but not only color words, vary little and thus are counted observational sentences when used as such. The translator-engineer, applying physics to translation problems through stimulus response theory, may for various reasons adjust the boundaries of the observational; observationality is a matter of degree. Still *sameness* of stimulus meaning across the population is a safe guide to observationality. From there it is just a problem of finding a word or phrase in the home language which has the same, or roughly the same, stimulus meaning.

Nevertheless, evidence of observationality does not increase the likelihood that «Green» is the right translation of «Grue». For unlike other observational sentences, «Grue» is itself underdetermined by scientific results on stimulus meaning. Nothing in the stimulus meaning of any observational predicate, whether grue-like or not, differentiates between mistaken translation and change of meaning. On rechecking the stimulus meaning tables for «Grue» the translator after t finds that green stimuli no longer prompt native affirmations. Nothing in the stimulus meanings will decide whether the original translation was in error, there was a change in meaning, or a live grue predicate has been discovered. Whether Gruese is rife with the likes of «Grue» or has only a few such sentences does not matter. Each such sentence is prima facie a counterexample to the claim that all observational sentences are determinately translatable.

If only there were some universal mark of the grue-like pathological predicate or sentence, we could solve this problem by simply setting them aside. It seems doubtful that any such mark can be found. And without an independent characterization of gruelike sentences for the business of radical translation generally, we cannot be sure that **any** observational sentence is translatable. And this is a difficulty apart from the humdrum problem of always being somewhat uncertain about the conclusion of an induction. Given that «Grue» is observational and indeterminate, and given that there is no independent way for the radical translator to spot gruelike sentences in the speech of the Other, no amount of empirical evidence about the stimulus meaning of any given sentence can increase the probability that it is not one of the gruelike and that its stimulus meaning is a guide to its translation. Unless we have an independent way of spotting them, some new way must be found to show that «Grue» and its ilk are, despite appearance, determinately translatable.

Again, the thesis of indeterminacy of translation takes its philosophical import from the contrast with that limited class of cases for which determinacy of translation is well-defined and well-established. Without a clear sense of determinacy of translation, there is no interesting sense of indeterminacy of translation. Indeterminacy would be everywhere and thus nowhere.

Such a situation would also be bad for the feasibility of a behavioristic account in general. For if such an account does not cover observational sentences, it is not likely to cover any other class of sentences. Without determinacy, the stimulus-meaning account, though intended to be a test of how far a scientific theory of meaning could take us, would lose its credibility and the famous thesis of indeterminacy of translation its point.

It might be pressed that these examples confuse indeterminacy of translation with indeterminacy of induction. We may never find out the actual stimulus meaning of «Grue» and thus be defeated in our inductions from cases. Still the translation of «Grue» into English is just Goodman's famous formula, whether we can ever discover that fact or not. The evidence that «Grue» is to be translated «Green» can accumulate, but evidence that «Grue» is to be translated by a version of Goodman's formula will not accumulate. Quine's scientific recipe for radical translation, at the first significant series of blue stimulations' prompting of «Grue» from the native, calls for the conclusion that a change of meaning has occurred or that a batch of linguistically deviant natives had been used or that some strange error was made in recording data. Thus Gruese, even in its observational base, completely escapes Quine's method.

8. This suggests the following argument. There can be two equally good translation

manuals according to one of which «Grue» is always translated as «Green» and according to the other as «Green» up to t and «Blue» otherwise. Both accord equally well with the hard factual evidence. Therefore «Grue» has no determinate translation. This seems to undermine the whole notion of determinacy as it is set out in *Word and Object.*

It might be said that this argument fails to appreciate that in Quine's book translation is naturalized. Radical translation takes place within the assumption that science, specifically physics, is true. It is this natural standpoint that gives point to the enterprise. The question the theory of translation tries to answer is that given this, how can we formulate a scientific account of translation between radically different languages? That is, how much of our naive theory of meaning can be scientifically reconstructed? Now given the the theory of evolution is fundamentally correct Gruese is impossible, for projection of «Grue» and the like would promote catastrophe. Speakers of Gruese, or those with a gruese-like prelinguistic quality space, would have fallen by the wayside. Thus within this overall assumption, the grue problem cannot arise.

Perhaps, one might argue, the grue problem does not arise within the natural standpoint. But Quine himself in «Indeterminacy Again»[12] shows that the native's physics may be quite different from ours and that our translation of his physics up to observationality is radically indeterminate even on the assumption that our physics is the right one. The grue argument here presses the point that the qualification «up to observationality» be dropped. And with it the whole idea of determinate translation even within the natural standpoint. Moreover, the evolutionary argument may fail to be convincing for a reason cited above, that there is no known means of giving an independent characterization of grue-like predicates.

9. This argument from «Grue», if it succeeds, prompts a new perspective on translation. Evaluation has a hidden link to habit, custom, tradition. A valid deductive argument remains defective if a premise is false or it begs the question. A inductive argument with all true premises remains useless if it is based on unprojectible (unentrenched) predicates. Samples are representative only if taken according to accepted practice. These are all instance or rightness, and if Goodman is right, rightness arises from habit or tradition. Analytical hypotheses, like first-level hypotheses about emeralds in a given language, must both fit the evidence and be *right*. Analytical hypothese must meet Quine's conditions (1)-(4) and pass some further test of rightness. Perhaps the principle of charity is best regarded as a principle of rightness. The argument here shows that the rightness requirement for translation extends all the way down, all the way down to all sentences. What such a test might be is an intriguing question. But however that question is answered, uniqueness and thus determinacy of translation are not to be expected *at any level however theoretical or observational.* Rightness in translation theory, as rightness in metaphysics, bestows objectivity but not uniqueness. What goes for translation goes for theory generally. We have no firm foundation; we are at sea with only our traditions

12. *Journal of Philosophy* (1973).

and habits to carry us along.

M. G. Yoes
Professor of Philosophy, Emeritus
Department of Philosophy
University of Houston
mgyoes@comcast.net

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 37-42
Temporal and Counterfactual Possibility

TEMPORAL AND COUNTERFACTUAL POSSIBILITY

by Muhammad Ali Khalidi

Among philosophers working on modality, there is a common assumption that there is a strong connection between temporal possibility and counterfactual possibility. For example, Sydney Shoemaker (1998, 69-70) writes:

> It seems to me a general feature of our thought about possibility that how we think that something could have differed from how it in fact is [is] closely related to how we think that the way something is at one time could differ from the way that same thing is at a different time. In possible worlds jargon, the ways one and the same thing of a given sort can differ across worlds correspond to the ways one and the same thing of that sort can differ at different times in the same world. Could I have been a plumber or an accountant instead of a philosopher? The answer seems to be yes — and this goes with the fact that we acknowledge the possibility of a scenario in which someone who was exactly as I was at some point in my life undergoes a series of changes resulting in his eventually being a plumber or an accountant.

In a footnote, he acknowledges that this connection between counterfactual and temporal possibility needs to be qualified in response to an objection by Randy Clarke: «The property of being the child of someone who has visited Paris is not one that one can have and then lose; but it is one that one can have in the actual world and not have in some other possible world.» (1998, 75n.9) Therefore, he restricts the inference from temporal to counterfactual possibility and vice versa to «non-historical properties.»

In the passage quoted above, Shoemaker stops short of endorsing strict implication relations between temporal and counterfactual possibility. Still, there is a presumption in the work of many philosophers that a *de re* modal claim concerning the possession of a property by an individual usually implies and is implied by a temporal claim concerning the possession of that same property by that individual. For example, it might be argued that Alfred could not possibly have been non-human on the grounds that Alfred cannot become non-human, or that Bertha could possibly have been a plumber on the grounds that she can become a plumber. Generally, we can frame the relevant principles as follows[1]:

(1) If x could have been F in another possible world, then x can become F in the course of time in the actual world.

(2) If x can become F in the course of time in the actual world, then x could have been F in another possible world.

These principles are rarely if ever stated in this stark fashion, but implicit appeal to them (especially the second) abounds in the work of essentialist philosophers. To cite just one

---

[1]. One can also formulate contrapositive equivalents of these claims, e.g. (1') If x cannot become F in the course of time in the actual world, then x could not have been F in another possible world. Corresponding claims can also be framed in terms of necessity rather than possibility.

recent example, in a book on scientific essentialism, Brian Ellis (2001, 32) characterizes accidental properties as «properties that a thing can acquire or lose without ceasing to be a thing of the kind it is.» The connection between what is accidental (i.e. non-essential, and therefore contingent) with what can be actually acquired or lost in the course of time seems quite widespread.

However, these principles have been questioned by at least one recent writer on modality. E. J. Lowe thinks that it is neither the case that counterfactual possibility implies temporal possibility ((1) above), nor that temporal possibility implies counterfactual possibility ((2) above). He argues against (1) as follows: «… a sculptor could perhaps have given a somewhat different shape to a statue that he has just made — but it doesn't follow that the statue can now be made to take on that different shape.» (2002, 80) Though Lowe does not elaborate further, it is not difficult to conjure up a case to illustrate the claim. Imagine that a sculptor has taken a block of marble and fashioned it into a bust of Socrates with a small nose. Let us grant that it may not be possible for that sculpture to be reshaped at some later time such that it has a large nose. This would be the case, for example, if the sculpture has been fashioned from a single piece of marble that cannot be repaired by reattaching parts that have been cut from it. However, the sculpture might originally have been capable of being fashioned into a head with a large nose (indeed, the sculptor might first have carved it in such a way as to have a large nose, then made the nose smaller). Thus, this sculpture could have had a different shape from the one that it actually has (counterfactual possibility), but it may not be possible for the sculpture to acquire that shape in the future (temporal possibility).

As for (2), Lowe imagines a certain plant that can acquire a different shape from the shape it now has, though it could not have had a different shape in another possible world. He writes: «it might be that a plant of a certain kind *must* possess a certain shape at a certain stage of its existence — for instance, when it is a seed — even though it is possible for it to change its shape thereafter.» (2002, 80; original emphasis) He holds that it could take on a different shape in the future, when it is transformed from a seed into a seedling. But «it is true to say of it, when it is a seed, that it could not have had a different shape from the shape that it actually has at that moment.» (2002, 80) Hence, even though it is temporally possible to take on a different shape, it is not counterfactually possible for it to have had that shape.

In the rest of the paper, the two principles and their respective counterexamples will be considered in reverse order. After describing a modified version of the second counterexample, in section 2, an attempt will be made to show that it does not, on closer inspection defeat (2). Then, in section 3, it will be argued that the first counterexample does defeat (1), but for a certain set of non-standard «irreversible properties». (The «historical properties» for which Shoemaker issues his restriction are a special case of such properties.) Section 4 will conclude by claiming that it has not been shown that temporal possibility does not imply counterfactual possibility, and that although it has been shown that counterfactual possibility does not generally imply temporal possibility, the exceptions to this principle are nonstandard properties.

2. Does Temporal Possibility Imply Counterfactual?

Let us begin with a modified version of the second counterexample, which purports to show that temporal possibility does not imply counterfactual possibility. When it comes to *de re* modality, it always helps to personalize matters. Consider a certain caterpillar Charlie in the larval stage at time t_1. At that time, it seems correct to say that Charlie must necessarily

be larva-shaped (she cannot, for example, be butterfly-shaped and remain a larva). However, it is clearly possible, and in many cases it is actually the case, that Charlie can become butterfly-shaped at some future time. Is it really impossible for Charlie to be butterfly-shaped in some other possible world? It might be argued that she could not, on the grounds that at that very stage in her life cycle, she must be larva-shaped. However, since we are dealing with *de re* modality, the question is not whether Charlie-as-larva could have been butterfly-shaped, but whether *Charlie* could have been butterfly-shaped. And it is clearly possible for Charlie to have been butterfly-shaped without ceasing to be Charlie. To see this more clearly, let us suppose that we allow some time to pass in the actual world, after which Charlie has been transformed first into a pupa, and then into an imago (butterfly). If we assume uncontroversially that the organism maintains its identity through these transformations, then Charlie the larva will have become Charlie the imago.[2] At that point in time, call it t_2, if we ask ourselves whether Charlie could have been a butterfly, the answer is clearly that she could have been, since she is actually a butterfly, and actuality implies (counterfactual) possibility. Therefore, since it is true of Charlie at t_2 that she could have been a butterfly, it is also true of Charlie at t_1, that she could have been a butterfly.

But what should we make of the claim that Charlie is necessarily larva-shaped at the larval stage of her existence, or that she cannot possibly have been butterfly-shaped while at the larval stage? It is true that it is impossible for a larva to be butterfly-shaped (according to biological laws as we know them). But that does not show that it is impossible *for Charlie* to be butterfly-shaped. The claim that it is impossible for a larva (while remaining a larva) to be butterfly-shaped is quite different from the claim that it is impossible for Charlie (while remaining Charlie) to have been butterfly-shaped. The first is an essentialist claim about properties, while the second is an essentialist claim about an individual's possession of a property. Accordingly, insofar as we are concerned with *de re* modalities pertaining to the possession of certain properties by individuals, the first claim does not lead us to the conclusion that an individual's possession of a property may be temporally possible but counterfactually impossible.

We could try to relativize the counterfactual claim to a time, but that would not help. It might be said that Charlie at time t_1 (while in the larval stage) could not have been butterfly-shaped. But if we assume necessity of origin for organisms, so that Charlie necessarily originates from some particular fertilized egg, then there is a possible world in which Charlie was born at some time prior to the time at which she was actually born. In particular, there is another possible world in which Charlie hatched earlier than t_1 and was already transformed into a butterfly by t_1. Therefore, it is not even true to say of Charlie at time t_1 that she could not have been butterfly-shaped at t_1.[3]

Is this case like Lowe's? It would appear so. In his example, a particular plant, which is a seed at a particular time, could not have had some other (non-seed) shape at that time. But if we frame this in terms of a *de re* modal claim about that particular plant, call it Danny, then it is clearly the case that Danny could have had some other shape, just as Charlie could have

---

[2]. If the organism does not survive these transformations then the counterexample cannot get off the ground, since it rests on the idea that it is a temporal possibility for the organism to change in this way.

[3]. Some (e.g. Peter King (1999)) have argued that it is incoherent to effect temporal comparisons across possible worlds. If so, then the onus is on the opponent of principle (2) to show how else to interpret the counterexample so as to defeat (2).

been butterfly-shaped. While it may be true that a seed that necessarily has a particular shape could not have had a plant-like shape at that stage of its existence, that is (again) not a *de re* modal claim about a particular individual.

3. Does Counterfactual Possibility Imply Temporal?

In the above elaboration of Lowe's first counterexample, a bust of Socrates with a small nose, call it 'Soc', cannot at time t_1 (given its dimensions at that time) be carved so that it has a large nose. So it is not temporally possible for Soc to acquire a certain shape from t_1 onwards, but it would have been possible for Soc to have had that shape in some other possible world. Imagine that Soc is first carved into a bust with a large nose at some time t_0 prior to t_1. Now let us suppose that the sculptor has learned that Socrates had a small snub nose and has further chiseled Soc's nose at time t_1, making it smaller. It is not possible at t_1 for the bust to be transformed in such a way that it becomes large-nosed, but it could have been large-nosed in another possible world (indeed, it had a large nose in the actual world at t_0). This is supposed to refute the claim that counterfactual possibility implies temporal possibility.[4]

The general form of such cases is as follows. They ask us to consider an individual *i* that lacks some contingent property *P* at time t_0, where *P* is such that once it has been acquired it cannot be lost. Then they suppose that *i* acquires *P* at some later time t_1. Since *P* is a contingent property of *i*, *i* might not have possessed *P* in the first place, but once it has acquired *P* at t_1 it cannot be without *P* any time after t_1. Thus, when we consider *i* any time after t_1, it is clear that it might not have had *P* but that it cannot come not to have *P*. This shows that something can be a counterfactual possibility but not a temporal possibility. But are there any such contingent properties *P*, which are such that once acquired, they cannot be lost? Here are a few examples: *adult*, *immune to chicken pox*, *war veteran*, and *afflicted with an incurable disease*. Moreover, it is quite simple to construct such an «irreversible property» from certain ordinary «reversible» properties. Take some contingent property *Q*, such as *riding a bicycle*, and then construct a tensed version of *Q*, call it *P*, *has ridden a bicycle*. Clearly, anything that has had *Q* at some time always has *P* at every subsequent time, so *P* is not a property that *i* can lose from t_1 onwards (even though *i* might lose *Q* itself). Thus, an individual can possess *Q* at one time and not another, but having had *Q*, that individual continues to have *P* no matter what. Losing *P* is therefore not a temporal possibility for that individual from t_1 onwards, even though it is clearly a counterfactual possibility not to possess *P*, simply because *Q* is contingent in the first place.

How does this apply to Lowe's counterexample? The bust of Socrates could have been small-nosed in another possible world, but it cannot be made to become small-nosed after t_1 in this world. However, in this case the property *small-nosed* is not obviously an irreversible property, as defined above. What makes it irreversible is the assumption, to which I alluded in section 1, concerning the medium in which the sculpture has been fashioned (though Lowe himself does not make this assumption explicit). In this case, having acquired a small nose, we are assuming that Soc cannot regain the matter in the right location, in order to acquire a large nose (thereby losing the property *small-nosed*). If the sculptor were working in the medium of clay, he could simply replace the bits of matter that had been discarded, and Soc

4. I take it that it is uncontroversial that the sculpture endures through such relatively minor changes to its nose, remaining the same sculpture throughout. If not, the changes can be made as small as one likes so as not to affect the identity of the artifact.

could become large-nosed again. Hence, the irreversible property in this case is something like, *small-nosed bust carved in marble*.

The above scheme also accounts for Shoemaker's counterexample (which was suggested by Clarke) concerning the «historical property» of being the child of someone who has visited Paris. First, consider a simpler case, the property *has visited Paris*. This is indeed an irreversible property, as defined above: it is contingent but cannot be lost once it has been acquired. Moreover, it is a tensed property that can be formed from a reversible property in the manner suggested above. If *Q* is *is visiting Paris*, then *P* is *has visited Paris*, and it is clear that if *Qi* is actual at some time t_1 for an individual *i*, then *Pi* is actual for all times after t_1. In other words, one cannot lose the property *P* in the actual world [5], even though there is a possible world in which *~Pi*. In Shoemaker's case, the matter is complicated somewhat by the fact that the property is possessed not by the individual under consideration, but by that individual's parents. Presumably, this works because many essentialist writers presuppose some form of origin essentialism for human beings, so having the parents that one actually has is also an essential property of an individual. Thus, one can displace the property *P* onto the individual's parents rather than the individual him or herself, but the conclusion remains the same.

Therefore, principle (1) fails for what I have called «irreversible properties» (including some tensed properties that can be constructed in such a way that they are irreversible). Such peculiar properties might be banished by «sparse» theorists of properties, who argue that not all predicates correspond to properties and that only real properties should be admitted into our ontology. But there is no need to resort to such desperate measures, provided we bear in mind that principle (1) does not apply to the class of irreversible properties.

4. Conclusion

This paper has tried to show, by elaborating on Lowe's counterexamples, that they do not block the inference from temporal to counterfactual possibility, and that although they do block it from counterfactual to temporal possibility, they do so only for a class of non-standard irreversible properties (including some tensed properties). The inference from temporal to counterfactual possibility would seem to be more crucial for modal theorizing than the converse, since one plausible guide to what could happen in other possible worlds is what could happen in the course of time in this world. It is still permissible to reason from what properties an individual might come to possess in the future to what properties that individual might have possessed in some other possible world, at least insofar as it makes sense to speak of *de re* modal properties of individuals in the first place. Whether this would enable us to replace our possible-worlds talk with talk about possible futures is a subject for another paper.

References

Ellis, Brian (2001), *Scientific Essentialism*, Cambridge: Cambridge University Press.

King, Peter J. (1999), «Hereafter, in a Later World Than This?» *Sorites* 10, 74-79.

Lowe, E. J. (2002), *A Survey of Metaphysics*, Oxford: Oxford University Press.

[5]. Hence the phrase, «We'll always have Paris.»

Shoemaker, Sydney (1998), «Causal and Metaphysical Necessity,» *Pacific Philosophical Quarterly* 79, 59-77.

Acknowledgment

I would like to thank Nelly Dib for discussions relating to the topic of this paper and for reading it and providing helpful feedback.

Muhammad Ali Khalidi
Department of Philosophy
York University
4700 Keele Street
Toronto, ON M3J 1P3. Canada
khalidi@yorku.ca

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 43-48
Actualism and the Distinction of Truth over Truth in a World

ACTUALISM AND THE DISTINCTION OF TRUTH OVER TRUTH IN A WORLD

by Edward R. Moad

Robert Adams characterizes actualism regarding possible worlds as «the view that if there are any true statements in which there are said to be nonactual possible worlds, they must be reducible to statements in which the only things there are said to be are things which there are in the actual world, and which are not identical with nonactual possibles.»[1] Both Adams and Alvin Plantinga, two major proponents of actualism, hold that there indeed are possible worlds, and that they are reducible in the way demanded by this condition.

In this paper, I will briefly explain actualism about possible worlds, showing that an essential pillar of the theory is the claim that truth is distinct from, and ontologically prior to, truth in a world. The rest of the paper is premised on the idea that an interesting philosophical defense of this claim calls for an analysis of truth itself, and is not intended as an objection to actualism, but rather to underscore the interest actualists should have in the question of what truth is. First, I will consider the idea, drawn from Adams and Alan McMichael, that truth differs from truth in a world in its being a matter of correspondence between a proposition and an independent object; that object being, in McMichael's words, the 'concrete universe'. Then, I will show that, given such an analysis of truth, the truth conditions for propositions about non-actual possibilities violate the central tenet of actualism, as articulated by Adams, above.

Adams defines a 'world' (or 'world-story') as a maximal consistent set of propositions, and the 'actual world' as the 'true story'; that set including all and only true propositions. Plantinga's possible worlds, on the other hand, are maximal states of affairs, i.e. a state of affairs that, «for every state of affairs S, either includes or precludes S.» Here, «S includes S* if it is not possible that S obtain and S* fail to obtain,» and «S precludes S* if it is not possible that S and S* obtain.»[2] Of these possible worlds, the actual world is a maximal state of affairs such that every state of affairs it includes is actual. States of affairs are abstract objects, all possible worlds are states of affairs, and the actual world is a possible world. Therefore, the actual world is an abstract object, but it is not a proposition. Plantinga's reason for postulating a distinction between the two is that, «propositions have a property — truth or

---

1. Adams 224.

2. Plantinga 172.

falsehood — not had by states of affairs.»[3]

Maximal consistent sets of propositions, on the other hand, are books. On this theory, «the set of propositions true in a given world W is the book on W,» and, «a proposition p is true in a state of affairs S if it is not possible that S be actual and p be false,» and is true in a world «if it is impossible that W obtain and p be false.» These latter are not definitions of truth, but truth in a world. Like Adams, Plantinga makes it clear that, «truth is not be explained in terms of truth in the actual world, or truth in α; the explanation goes the other way around.»[4]

Thus, actuality is to be analyzed in terms of truth. «'In the actual world, p',» writes Adams, «is to be analyzed as, 'The proposition that (p) is true.'»[5] Analysis of actuality in terms of truth is a central feature of actualism, with important implications. «In the true story theory of actuality, the notion of truth is presupposed, if not as primitive, at least as prior to the notion of actuality, since the latter is analyzed in terms of the former.»[6] Since «truth in w» is, for Plantinga, a relation a proposition bears to a state of affairs in w, truth in the actual world is explained as a relation a proposition bears to an actual state of affairs. On pain of circularity, then, actuality cannot be analyzed as a relation to a proposition true in the actual world. Likewise, truth cannot be analyzed as a relation a proposition bears to an actual state of affairs. So, neither truth nor actuality can be analyzed in terms of relations between states of affairs and propositions.

Adams, on the difference between actualist and possibilist views of truth, writes that the actualist «sees the truth of a proposition in a possible world as basically a matter of relations of consistency between propositions, rather than of correspondence with an independent object.»[7] The key here is that the possibilist does see the truth of a proposition in a possible world as correspondence with an independent object, and consequently, a possible world is an independent object. Thus, for the possibilist, there is no difference between truth and truth in a world; truth is just truth in *this* world. Actualism, however, is an analysis of actuality in terms of truth that attempts to answer, in no world-relative terms, the question, as Adams puts it: «In what does the actuality of the actual world consist?»[8]

Then some analysis of truth is necessary to complete the actualist analysis, for its central feature is that actuality is analyzed in terms of truth and not the converse. If such an analysis is to clarify the distinction between the actual and possible worlds, there must be a clear distinction between truth and truth in a world. Conditions for truth in a world, alone, are not sufficient to clarify this distinction. A positive analysis of truth distinct from truth in a world is called for, particularly in light of the fact that the issue over whether there really is such

---

3. Ibid 173.

4. Ibid.

5. Adams 225.

6. Ibid 226.

7. Ibid 227.

8. Ibid 211.

a distinction is the central point of difference between actualism and possibilism. Despite Adams' suggestion, above, that the notion of truth might be taken as a primitive, it is hard to see how anything at all can be said in defense of the proposition that truth is distinct from, and ontologically prior to, truth in a world, if there is nothing to be said about what truth is. In «Theories of Actuality» and «Actualism and Possible Worlds», where Adams and Plantinga, respectively, construct their actualist theories, such an analysis is absent. Only «truth in a world» is treated: by Plantinga, notably, as part of his explanation of essences and truth conditions of modal propositions, and by Adams, as we saw, in denying that truth in a world is a correspondence relation with an independent object.

This denial may be a clue to putting such an analysis together. It seems natural to presume that Adams sees truth as what truth in a world is not — correspondence with an independent object. 'Independent', here, likely means 'independent of the proposition'. If so, then actualism precludes the actual world from being the independent object correspondence to which makes true propositions true. Of course, the actual world consists of just such propositions. So what is this independent object?

Another actualist, Alan McMichael, helps here. «The actuality — non-actuality distinction is based on the true — false duality of propositions: one of the worlds is actual because it contains all the true propositions.»[9] Two pages before, we find, «the actual world is not actual merely in the sense that it exists — all possible worlds exist — but rather in the sense that this concrete universe corresponds to it.»[10] So the actual world is the maximal consistent set of all the true propositions, and its actuality consists in its corresponding to the concrete universe. It looks like a proposition is true iff it corresponds to the concrete universe. If this is true and a world consists of propositions, then actuality just is truth.

For, Plantinga, however, a world consists, not of propositions, but of states of affairs. But he distinguishes them, as we saw, on the basis of the fact that propositions bear the properties of truth or falsehood, while states of affairs bear the properties of actuality or non-actuality. Truth, then, cannot be distinct from actuality simply on the basis of the fact that propositions bear the former, while states of affairs bear the latter. Neither distinction — either that between worlds and books, or that between truth and actuality — can be maintained unless at least one of them can be analyzed independently of the other. So the distinction between worlds and books collapses, along with any distinction between truth and actuality that one may have hoped to base on a distinction between worlds and books, and actuality turns out to be truth after all.

A problem arises for actualism, now, with the question of the truth conditions of modal propositions. A proposition is true iff it corresponds to the concrete universe. A maximally consistent set of propositions are all true (the set is the actual world) iff they all correspond to the concrete universe. For any true proposition, then, some feature of the concrete universe must be the «truth-maker» — that to which it bears the relationship in virtue of which it is true. So propositions about non-actual possibilities, if true, are true in virtue of features of the concrete universe. What features of the concrete universe can be the «truth-makers» for propositions about possible worlds?

---

9. McMichael 52.

10. Ibid 50.

McMichael gives the following truth conditions for modal propositions:

> 'It is possible that A' is true iff there is a possible world w such that w includes the state of affairs expressed by A.
>
> 'It is necessary that A' is true iff every possible world w includes the state of affairs expressed by A.[11]

These conditions, modified in light of our analysis of truth, then, are as follows:

'It is possible that A' is true iff a possible world w, such that w includes the state of affairs expressed by A, is a feature of the concrete universe.

'It is necessary that A' is true iff the inclusion, in every possible world, of the state of affairs expressed by A, is a feature of the concrete universe.

Given the actualist theory as understood here, if propositions about non-actual possibilities can be true, then maximally consistent sets of propositions — some of which are false — are features of the concrete universe. Furthermore, propositions like 'it is possible that A', if true, are members of the actual world in virtue of corresponding to a proposition 'A' that is not a member of the actual world, but is a member of a maximally consistent set of propositions which itself is a feature of the concrete universe. Propositions like 'it is necessary that A', if true, are members of the actual world in virtue of corresponding to a proposition 'A' that is a member of every maximally consistent set of propositions (including the actual world) that is a feature of the concrete universe.

The issue at hand is whether or not the situation into which we have been led is acceptable under the original actualist mission. That mission was to show that «true statements in which there are said to be nonactual possible worlds» are «reducible to statements in which the only things there are said to be are things which are in the actual world and which are not identical with nonactual possibles.» But we have seen that, given the preceding analysis of truth, true propositions about nonactual possible worlds entail that those possible worlds are features of the concrete universe. True propositions about non-actual possible worlds, then, entail that those worlds exist. But while true propositions about possible worlds are members of the actual world (the maximal consistent set of all true propositions), the possible worlds themselves are not. Indeed, no consistent set could contain them all.

An analysis of truth as correspondence between a proposition and the concrete universe, then, renders actualism false, as it entails that true statements about non-actual possible worlds are not reducible to statements in which the only things said to be are things which there are in the actual world. Here, possible worlds (maximally consistent sets of propositions) are said to be, but they cannot be part of the actual world (the set of true propositions).

There seem to be two ways for an actualist to respond. The first is to deny that being a feature of the concrete universe is equivalent to existence, in which case the proposed analysis of truth, as correspondence to the concrete universe, does not render the problematic entailment between true propositions about non-actual possible worlds and the existence of those worlds. But to do so begs the question of the meaning of 'concrete universe'. If being a part of it does not entail existence, then it becomes merely an empty term meant to serve as a point of reference for 'correspondence' in what turns out to be an ad-hoc analysis of truth.

---

[11]. Ibid 52-53.

A second possible response is to deny that correspondence to the ‘concrete universe’ is the correct analysis of truth. In this case, the actualist could either offer a different analysis of truth, or else no analysis at all. As I had previously pointed out, the most natural option — given actualist intuitions — would seem to be some analysis in line with the correspondence theory. But we have already seen that an understanding of truth as a correspondence to the ‘concrete universe’ violates the essential actualist commitment. To what, then, are we to understand truth as a correspondence? The problem with the ‘concrete universe’ was that it was understood as that of which being a part entails existence. Then, whatever the actualist takes to be that with which correspondence constitutes truth, it must be something of which being a part does not entail existence. But then, truth could be correspondence with the non-existent, and this eliminates all that might have appealed to an actualist’s intuitions about a correspondence theory of truth.

In the latter case, perhaps the argument could be made that, as Adams suggests, truth is a primitive notion, and that, therefore, no analysis is possible. However, as previously pointed out, this puts the actualist in a vulnerable position in relation to possibilism, if we are to understand the central controversy between the two views in terms of the question as to whether or not there is a metaphysical distinction between truth and truth in a world. Consider a possibilist charge that, given the actualist contention that there is such a distinction between truth and truth in a world, the actualist is obligated to explain the basis of the distinction between the two, along with that of the ontological priority that the former allegedly bears in relation to the latter. But such an explanation cannot be given unless the salient differences between the two can be identified, and this requires a clear idea of what constitutes truth. Without this, it might be said, the case for actualism seems to amount to the claim that ‘something we know not what’ is distinct from truth in w, and is in fact that in terms of which truth in á is to be understood, and that therefore, actualism is true.

Of course, this is not a knock-down argument for possibilism (nor do I intend to defend possibilism). It remains quite possible that truth is an un-analyzable primitive that is distinct from and ontologically prior to truth in w. But any actualist whose philosophical interest is anything over and above simply defending actualism ought to wonder if, indeed, nothing more can be said as to what makes truth itself distinct from truth in w. This should motivate interest, on the part of actualists, in whether or not any other analysis of truth might be given that explains this distinction, and it introduces the question as to whether, and how, any of the various theories of truth currently on offer might prove compatible with actualist intuitions and commitments.

———————————————

References

Adams, Robert M. «Theories of Actuality,» Nous, 8 (1974), 211-221.

McMichael, Alan, «A Problem for Actualism about Possible Worlds,» *The Philosophical Review*, 92, issue 1, (Jan. 1983), 49-66.

Plantinga, Alvin, «Actualism and Possible Worlds,» reprinted in Loux, Michael, *Metaphysics: Contemporary Readings* (Routledge, 2001).

Edward R. Moad
Department of Philosophy
National University of Singapore
3 Arts Link
Singapore, 117570
phimer@nus.edu.sg

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 49-66
The Constitution Argument Against Conceptualism

THE CONSTITUTION ARGUMENT AGAINST CONCEPTUALISM

by André Abath

---

1. Introduction

In his *Mind and World* (1994) and elsewhere, John McDowell defends the view that the contents of perceptual experience are conceptual. This view came to be known as *Conceptualism*. However, Conceptualism has not proved particularly popular with the philosophical community. The main reason for this lack of popularity is the fact that many of those interested in the relation between concepts and the contents of perceptual experience think that a version of the so-called Fineness of Grain Argument, which supposedly has as its consequence the falsity of Conceptualism, is valid and sound. In rough terms, the argument has as its conclusion that we do not possess concepts for all that is presented to us in perceptual experience at the level of detail at which things are presented to us. For instance, we are presented in perceptual experience with different shades of red, but we would have no concepts for such shades at the level of detail at which they are presented to us. That is, we might possess a concept for the general category to which the shades belong, a concept such as RED, but not for the specific shades of red presented to us.

However, proponents of Conceptualism think they are able to block the Fineness of Grain Argument. According to them, we might not possess *general* concepts for fine-grained properties at the level of detail at which they are presented to us, but we possess *demonstrative* concepts for fine-grained properties at this level of detail.

In this paper, I will grant proponents of Conceptualism that as a matter of fact we do possess demonstrative concepts for fine-grained properties at the level of detail at which they are presented to us. But I will argue that, even if this is the case, we have good reasons to think that Conceptualism is false. For there is a valid and sound argument — that I will call the Constitution Argument — that has as its conclusion the falsity of Conceptualism. I present the Constitution Argument in section 4. In sections 4, 5 and 6 I show that the argument is valid and sound, in which case we have good reasons to think that Conceptualism is false. However, before introducing the Constitution Argument, we need to have a better understanding of what Conceptualism amounts to, of the Fineness of Grain Argument and of the Conceptualist response to the argument. Sections 2 and 3 are dedicated to this task.

2. Introducing Conceptualism and the Fineness of Grain Argument

Now, what, exactly, is Conceptualism? What does it mean to say that the contents of perceptual experience are conceptual? Let us begin to address this question by considering a few quotes from McDowell's *Mind and World*, in which he presents his version of

Conceptualism. He writes things like:

> Experiences in general are states or occurrences in which conceptual capacities are passively drawn into operation (1994:30).

> *That things are thus and so* is the content of the experience, and it can also be the content of a judgement … So it is conceptual content (1994:26/McDowell's italics).

> Experiences have their content by virtue of the fact that conceptual capacities are operative in them (1994:66).

The idea then is that concepts are *applied* in perceptual experience (in McDowell's terms, «conceptual capacities are passively drawn into operation» in perceptual experience), and these concepts are *constitutive* of the contents of such experiences. That is, they give, or determine, the contents of perceptual experience. In terms close to McDowell's, it is in virtue of the fact that concepts are applied in perceptual experiences that experiences have a given content. If, in perceptual experience, I apply the concepts THAT THINGS ARE THUS AND SO, then what is presented to me in perceptual experience — the content of my perceptual experience — is *that things are thus and so*. The concepts I apply in experience are constitutive of the content of my experience.

Now, if this is the case, then Conceptualism can be construed as a thesis regarding what makes two contents of perceptual experience be identical. For if the concepts applied in perceptual experiences are constitutive of the contents of such experiences, then two contents of perceptual experience are identical if and only if the same concepts are applied in them in the same order. More schematically, we have the following thesis:

Conceptualism

> For any two perceptual experiences A and B, the content of perceptual experience A is identical to the content of perceptual experience B if and only if the same concepts that are applied in perceptual experience A are applied in perceptual experience B in the same order.[1]

This, I take it, is the clearest way of stating the thesis of Conceptualism. For, given this formulation, it is easy to see how Conceptualism can be evaluated as being true or false. If it is possible that there are (or if there are as a matter of fact) differences in the contents of two perceptual experiences even though the same concepts are applied in the experiences, then Conceptualism is false. Or if it is possible that there are (or if there are as a matter of fact) no differences in the contents of two perceptual experiences even though different concepts are applied in the experiences, Conceptualism is again false. If it is not possible that such situations occur, then Conceptualism is true. So, given the formulation of Conceptualism as a thesis regarding what makes two contents of perceptual experience identical, one can easily see how Conceptualism can be evaluated as being true or false.

Now, opponents of Conceptualism typically argue against this position by stressing the

---

[1]. Why in the same order? Let us suppose that concepts applied in perceptual experiences are constitutive of the contents of such experiences. Now imagine that I apply, in perceptual experience, the concepts JOHN IS KISSING MARY, and that you apply the concepts MARY IS KISSING JOHN. Clearly, the contents of our experiences differ, even though we apply the same concepts in experience. And they differ because we apply our concepts in a different order. So, the Conceptualist thesis must require that, for two contents of perceptual experience to be identical, the same concepts have to be applied in the same order.

fineness of grain of what is presented to us in perceptual experience. In doing this, they often appeal to a passage in Evans' *Varietes of Reference* (1982), in which he asks, in a rhetorical manner: «Do we really understand the proposal that we have as many concepts as there are shades of colour that we can sensibly discriminate?» (1982:229).

Evans' point was developed, among others, by Heck (2000). He writes:

> Before me, for example, are arranged various objects with various shapes and colours, of which, it might seem, I have no concept. My desk exhibits a whole host of shades of brown, for which I have no names. The speakers to the sides of the computer are not quite flat, but have curved faces; I could not begin to describe their shape in anything like adequate terms. The leaves on the trees outside my window are fluttering back and forth, randomly, as it seems to me, as the wind passes over them.- Yet my experience of these things represents them far more precisely than that, far more distinctively, it would seem, than any characterization I could hope to formulate, for myself or for others, in terms of the concepts I presently possess. The problem is not lack of time, but lack of descriptive resources, that is, lack of the appropriate concepts (2000:489-90).

The idea, then, is that, in perceptual experience, we are presented with fine-grained properties, such as very specific shades of colour, or very specific shapes. A single object, such as a desk, may be perceived as having different shades of brown, for instance. And, according to Heck, we have no names for such shades of colour. In fact, we have no appropriate, or, as I shall put it, no *adequate concepts* for such shades. That is, we have no concepts for such shades at the level of detail at which they are presented to us in perceptual experience. For instance, for a certain shade of brown that is presented to us, we possess a concept for the general colour category it belongs to. We possess the concept BROWN. But, according to Heck, we do not possess a concept for the specific shade of brown presented to us. We possess no concept under which only the specific shade of brown presented to us, and no other property in the neighbourhood, falls. Or, simply put, we possess no adequate concept for the shade of brown.[2]

But why are we supposed not to have adequate concepts for fine-grained properties, such as shades of colour? In Heck's passage, it is suggested that we have no adequate concepts for fine-grained properties because we have no names for them. After all, he takes «lack of concepts» to be paraphrased in terms of «lack of descriptive resources».

Now, it is certainly the case that we do not have words for all the fine-grained properties that are presented to us in perceptual experience. I do not have words for all the shades of black that I see on my laptop. But does it follow from this that I do not have adequate concepts for such shades? Is the possession of concepts dependent on the possession of words that express them? This is certainly controversial. It seems prima facie plausible that one can possess certain concepts without having words to express them. And if this is possible, then why are we supposed not to have adequate concepts for fine-grained properties?

Some opponents of Conceptualism argue that the problem is not that we do not have words for fine-grained properties, but that we cannot *recognize* them. Tye, for instance, in arguing for the idea that we do not possess adequate concepts for fine-grained properties, says that «The ordinary person cannot recognize red_{27} even after having just seen it» (2006:520). Thus, the idea is that, as fine-grained properties are presented to us, we do not take them, and are not capable of taking them, as being of the same type as properties presented to us in the past. And from this it is supposed to follow that we do not possess adequate concepts for such

---

[2]. Epistemologists interested in the justificatory relations between perceptual experience and empirical beliefs often make similar remarks. See, for instance, Williamson (2000:197) and Bonjour (2003:71).

properties. If one cannot recognize red_{27}, then one does not possess the concept RED_{27}. Thus, what is in play here is the following requirement on the possession of adequate concepts for fine-grained properties, such as shades of colour:

Recognitional Requirement

> In order to possess an adequate concept C for a fine-grained property P, one must be able to recognize P — that is, one must be able to take P as being of the same type as properties perceived in the past.

Given this requirement, if we are not able to recognize fine-grained properties, then we do not possess adequate concepts for such properties.

Let us assume for now that this is right, that we cannot recognize fine-grained properties, and hence that we do not possess adequate concepts for fine-grained properties, such as shades of colour. But if it is, then the following argument — let us call it the Fineness of Grain Argument — can be put forward:

Fineness of Grain Argument

(FG1) In perceptual experience, we are presented with fine-grained properties, such as shades of colour.

(FG2) We do not possess adequate concepts for fine-grained properties presented to us in perceptual experience.

(FG3) Therefore, we do not possess adequate concepts for all that is presented to us in perceptual experience.

Now, notice that there are two main ways of reading (FG2), one weaker and one stronger. They are the following:

(FG2*) We do not possess adequate concepts for *some* of the fine-grained properties presented to us in perceptual experience.[3]

(FG2**) We do not possess adequate concepts for *any* fine-grained property presented to us in perceptual experience.

It is not clear if it is (FG2*) or (FG2**) that opponents of Conceptualism are committed to. But I think it is clear that what the opponents of Conceptualism *should* be committed is to (FG2*). For, if this claim is true, it already implies (FG3), and, consequently, the falsity of Conceptualism. I will explain why next. There is no need for the opponents of Conceptualism to commit themselves to the stronger claim made in (FG2**). So, I will be assuming that (FG2) is read in terms of (FG2*), as being the claim that we do not possess adequate concepts for *some* of the fine-grained properties presented to us in perceptual experience.

Now, suppose that the Fineness of Grain Argument works. If it does, why is it supposed to be a threat to Conceptualism? Conceptualism, remember, is the following thesis:

[3]. Whether «some» should be taken as meaning «a few» or «most» is not relevant for the purposes at hand.

Conceptualism

> For any two perceptual experiences A and B, the content of perceptual experience A is identical to the content of perceptual experience B if and only if the same concepts that are applied in perceptual experience A are applied in perceptual experience B in the same order.

Now, suppose that there are some fine-grained properties for which subjects have no adequate concepts. Suppose that two of these fine-grained properties are the shades red_{24} and $red_{25.}$ Suppose, moreover, that a subject is, at time *t*, presented in perceptual experience with $red_{24.}$ And suppose that, at time *t1*, the subject is presented in perceptual experience with red_{25}. Now, if the subject does not possess adequate concepts for these fine-grained properties, then, when she is presented with them in experience, she can only apply a non-adequate concept, such as RED, for the two shades. But there is a clear intuitive difference between the content of the perceptual experience of the subject at *t* and at *t1*. In fact, it seems absurd to deny that there is a difference. After all, at *t* the subject is presented with a given property in experience. At *t1*, the subject is presented with a different property. But this difference is not captured by the concepts applied in the experience by the subject, which are the same. If this is the case, then Conceptualism is false. After all, if Conceptualism were true, the fact that the subject applies the same concepts in two perceptual experiences would imply that their contents are identical. But if the Fineness of Grain Argument works, this is not true. For there will be cases in which the same concepts are applied in different experiences, but their contents will still be different. So, if the Fineness of Grain Argument works, it seems to have as a consequence the falsity of Conceptualism.

But does the Fineness of Grain Argument work?

3. A Conceptualist Response to the Fineness of Grain Argument

A major proponent of Conceptualism, Bill Brewer, has been unconvinced by the Fineness of Grain Argument. He responds to it in the following way:

> There is an unacceptable assumption behind this line of argument, that concepts necessarily correspond with entirely context-independent classification of things, in such a way that they can, in principle at least, be grasped by anyone, anywhere, regardless of their current relations with the semantic values in question. This is what sustains the restriction upon the concepts available to capture subjects' perceptual discriminations of colours and volumes, in the examples given above, just to those associated with verbal expressions, like 'scarlet' and 'four cubic feet', which have content-independent norms of application. This restriction unacceptably rules out any appeal to context-dependent demonstrative concepts, though — concepts associated with expressions like «that shade of red» or «just that large volume», grasp of which essentially depends upon the subject's relations with the actual entities which constitute their semantic value … My reply [to the Fineness of Grain Argument] is that the fineness of grain in perceptual discrimination is matched precisely by the perceptual demonstrative *concepts* which the subject has in virtue of her conscious contact with the items in question. In other words, for any fineness of grain in perceptual content to which my opponent wishes to appeal in making his argument, the subject is capable of making a perceptual demonstrative judgement «that is thus», with just that fineness of grain. (1999:171-2/ Brewer's italics).

So, according to Brewer, the proponents of the Fineness of Grain Argument have wrongly

supposed that the only adequate concepts we might have for the fine-grained properties presented to us in perceptual experience are *general concepts*. They have ignored the possibility that we might possess *demonstrative concepts* for the fine-grained properties presented to us.[4]

Take a concept to be general only if it can be legitimately applied by a subject independently of any contextual relation between the subject and the objects or properties that fall under the concept. Thus, the concept RED is a general one because it can be legitimately applied by a subject independently of any contextual relation between the subject and the colour red. The subject need not be perceiving the colour to apply it, or be in any other relation to it. Unlike general ones, demonstrative concepts can only be legitimately applied given a certain contextual relation between the subject who uses it and the object or property that falls under the concept. It is typically taken to be a necessary condition for a subject to legitimately apply a demonstrative concept that the subject is perceiving the object or property that falls under the concept, and that in fact there is such an object or property before the subject.[5] Thus, in order to apply the concept THAT SHADE for a certain shade of colour, I must be perceiving it, and the shade must in fact be before me.[6]

So, for a shade of colour such as $black_{24}$, for instance, we might not possess the general concept $BLACK_{24}$. From this, proponents of the Fineness of Grain Argument concluded that we do not possess adequate concepts for $black_{24,}$ and for other fine-grained properties. But, according to Brewer, this does not follow. For we might well have non- general, demonstrative concepts, for fine-grained properties. For fine-grained shades of colour, for instance, we might possess the concept THAT SHADE. And, when applied before a given shade, such as $black_{24}$, the concept seems to refer to $black_{24}$ only, and to no other property in the neighbourhood. It seems, then, to be an adequate concept for the fine-grained property. And this should be the case in general. Brewer's point is that, for every fine-grained property presented to us in perceptual experience, we will have an adequate demonstrative concept for it.

---

4. According to Brewer himself, and also to people such as Heck (2000) and Kelly (2001), McDowell also defends the view that we possess demonstrative concepts for fine-grained properties. However, it is not clear that this is so. Nowhere does McDowell say that the concepts we possess for fine-grained properties are demonstrative. As I read him, and also as Peacocke (2001) reads him, he introduces demonstratives as *linguistic expressions* of general concepts based on recognition. He says, for instance, that «what is in play here is a recognitional capacity, possibly quite short-lived, that sets in with the experience» (1994:57), and that «demonstrative *expressions* would have to figure as *linguistic expressions* of them [such capacities]» (1994:58/my italics). The idea is that, as we are presented in perception with fine-grained properties in general, if the property is taken from our view and another presented after a short interval, we are able to judge whether this property is of the same type as the originally presented one or not. That is, according to McDowell, we are able to recognize fine-grained properties for a short while after they are presented to us. If this is the case, then we can be considered as having adequate concepts for fine-grained properties. These concepts, I take it, are, according to McDowell, general ones. But since we typically have no words to express such concepts, we have to express them by means of demonstratives.
The problem with McDowell's proposal, however, is that it is empirically false. There is evidence that we do not have a capacity to recognize fine-grained properties for a short while after they are presented to us and taken from our view. For most shades of colour, we are typically unable to tell if a shade presented five seconds after the presentation of a given shade is of the same type as the originally presented shade or not. If this is the case, then we are unable to recognize fine-grained properties for a short while after they are presented to us. See Burnam & Clark (1955).

5. For an influential defence of this view, see Evans (1982:72-3).

6. This applies only to demonstratives that are meant to refer to objects and properties in the subject's present environment. It does not apply to demonstratives that are meant to refer to objects and properties in the past.

If this is the case, then the second premise of the Fineness of Grain Argument is false. The premise states that

(FG2) We do not possess adequate concepts for fine-grained properties presented to us in perceptual experience.

According to Brewer, we do. We possess adequate demonstrative concepts for every fine-grained property presented to us in perceptual experience. So, if Brewer is right, the Fineness of Grain Argument is unsound.

Let us assume here that Brewer is right, and that the Fineness of Grain Argument can in fact be blocked. But this does not settle the issue. For the fact that we *possess* adequate concepts for all the fine-grained properties and everything else presented to us in experience does not imply that Conceptualism is true. In order for this to be the case, we must not only possess adequate concepts for fine-grained properties and everything else presented to us in experience, but these concepts must be applied in perceptual experience, and be constitutive of the contents of perceptual experience. After all, Conceptualism is a thesis according to which the concepts applied in perceptual experience *determine* the contents of such experiences. Thus, the opponent of Conceptualism could try to argue that the concepts we possess for the fine-grained properties presented to us in perceptual experience are not constitutive of the contents of perceptual experience.

In the next section, I will present an argument along those lines, which I will call the Constitution Argument. If the argument works, as I will show that it does, then we have good reasons to think that Conceptualism is false, even if the Fineness of Grain Argument fails.

4. The Constitution Argument

In this section, I introduce and clarify the Constitution Argument, which has as its conclusion the falsity of Conceptualism. The argument is the following:

Constitution Argument

(C1) If the adequate concepts we possess for all fine-grained properties presented to us in perceptual experience are not constitutive of perceptual contents, then it is possible for the contents of two perceptual experiences A and B to differ even though the same concepts that are applied in A are applied in B.

(C2) The only adequate concepts we possess for all fine-grained properties presented to us in perceptual experience are demonstrative concepts.

(C3) Demonstrative concepts are not constitutive of perceptual contents.

(C4) Thus, the adequate concepts we possess for all fine-grained properties presented to us in perceptual experience are not constitutive of perceptual contents (from C2 and C3).

(C5) Therefore, it is possible for the contents of two perceptual experiences A and B to differ even though the same concepts that are applied in A are applied in B (from C1 and C4).

(C6) If (C5), then Conceptualism is false.

(C7) Therefore, Conceptualism is false.

Let me comment on the premises of the argument, from bottom to top.[7]

I take it that (C6) is clearly true. Conceptualism, remember, is the following thesis:

Conceptualism

> For any two perceptual experiences A and B, the content of perceptual experience A is identical to the content of perceptual experience B if and only if the same concepts that are applied in perceptual experience A are applied in perceptual experience B in the same order.

So, according to Conceptualism, it is not possible for two perceptual contents to differ if the concepts applied in the experiences do not differ. So, if (C5) is true, Conceptualism is false, as stated in (C6).

(C3) is the most controversial premise in the argument. In the following sections, I will try to show that it is true. (C2) is not obvious, but I think that we have good reasons to think it is true. Adequate concepts for fine-grained properties are generally taken to be either general or demonstrative. If a concept is general, it can be legitimately applied by a subject independently of any contextual relation between the subject and the objects or properties that fall under the concept. If a concept is demonstrative, it can only be legitimately applied given a certain contextual relation between the subject who uses it and the object or property that falls under the concept.

Now, in section 2, we presented the following requirement as being one that is usually adopted by those who put forward the Fineness of Grain Argument against Conceptualism:

Recognitional Requirement

> In order to possess an adequate concept C for a fine-grained property P, one must recognize P — that is, take P as being of the same type as properties perceived in the past.

Now, it is a matter of debate whether this requirement holds both for the possession of adequate demonstrative concepts for fine-grained properties and for the possession of adequate general concepts for these same properties. People like Brewer (1999) and Chuard (2006) believe that this is only a requirement for the possession of adequate general concepts for fine-grained properties. Possession of demonstrative concepts for fine-grained properties would not depend on a capacity to recognize the property the falls under a concept. It should suffice that the subject is able to attend to the property, and to track it in space. For the sake of the argument, let us assume that this is the case. But the requirement could be modified in the following way:

Recognitional Requirement for General Concepts

> In order to possess an adequate general concept C for a fine-grained property P, one must recognize P — that is, take P as being of the same type as properties perceived in the past.

Unlike the original Recognitional Requirement, I think the Recognitional Requirement for General Concepts does state a necessary condition for possession of adequate general concepts for fine-grained properties. In fact, it seems to be supposed by all sides of the debate that this

---

[7]. Both Heck (2000) and Peacocke (2001) think that (C3) is true, and that it somehow implies the falsity of Conceptualism. Their remarks on the issue, however, are brief and sketchy. In a way, the Constitution Argument, as presented here, is an attempt to clearly state and develop Heck and Peacocke's intuitions.

is the case. Both opponents of Conceptualism, such as Tye, and Conceptualists, such as Brewer and McDowell, suppose that possession of adequate general concepts for fine-grained properties require a capacity to recognize such properties. The reasons are clear. Take a fine-grained property such as red_{25}. The idea here is that, in order to possess the concept RED_{25}, I must be able to recognize the shade. I must be able to take samples of red_{25} as being of the same type as samples perceived in the past. And this seems to be true. Intuitively, if I possess information about red_{25}, but information that does not enable me to recognize the shade, I will still not possess an adequate general concept for it. For instance, knowing that red_{25} is a colour, that it is a shade of red, that it is Mark's favorite shade of red, none of this seems to make it the case that I possess the concept RED_{25}. In fact, it seems that, unless I possess a capacity to recognize the shade, I do not possess an adequate general concept for it. Thus, it seems that the Recognitional Requirement for General Concepts does state a necessary condition for possession of adequate general concepts for fine-grained properties. And since all parties in the debate agree with this, we can safely assume it for purposes of argument here.

But, as it was mentioned earlier, we are not able to recognize every shade of colour presented to us, not even for a short while after the shades are taken from our view. For most shades, we are not able to tell if a shade presented five seconds after the presentation of the original shade is of the same type as the original shade or not. So, it seems to be an empirical fact that we are not able to recognize all the fine-grained properties presented to us in perceptual experience. If this is the case, then, given the Recognitional Requirement for General Concepts, we do not possess adequate general concepts for all the fine-grained properties presented to us.

This leaves us with demonstrative concepts. In the last section, I granted Conceptualists that we do possess demonstrative concepts for all fine-grained properties presented to us in perceptual experience. Let us go on assuming that this is the case. Now, there seems to be no reason for us to believe that we might possess adequate concepts for fine-grained properties which are neither general or demonstrative. Certainly Conceptualists have given us no reason to believe that there are such concepts. Thus, since we do not possess adequate general concepts for all fine-grained properties presented to us, there are good reasons to believe that the only adequate concepts we possess for all fine-grained properties presented to us are demonstrative concepts (assuming that we do possess such concepts). Thus, I think there are good reasons to think that (C2) is true.

This brings us to (C1). I take (C1) to be clearly true, but it needs to be clarified. Why is it the case that, if the adequate concepts we possess for all fine-grained properties presented to us in perceptual experience are not constitutive of perceptual contents, then it is possible for the contents of two perceptual experiences A and B to differ even though the same concepts that are applied in A are applied in B? To make things simpler, let us assume that the only adequate concepts we possess for all the fine-grained properties presented to us in experience are demonstrative concepts, as stated in (C2). The question then becomes: why is it the case that, if our demonstrative concepts for fine-grained properties are not constitutive of perceptual experiences, then it is possible for the contents of two perceptual experiences A and B to differ even though the same concepts that are applied in A are applied in B? The reason why this is so is fairly simple. Suppose that a subject A has a perceptual experience of a fine-grained property, such as of the shade red_{24}. And suppose that a subject B has a perceptual experience of slightly different shade of red — say, red_{25}. Now suppose that subject

A applies the concept THAT SHADE in experience, which refers to red_{24}. And suppose that subject B applies in experience the concept THAT SHADE, which refers to red_{25}. Now suppose that these concepts, and all other demonstrative concepts, are not constitutive of the contents of the perceptual experiences of subject A and subject B. Thus, these concepts do not determine, or help to determine, the contents of the perceptual experiences of A and B. But one thing is for sure: the contents of the perceptual experiences of A and B differ. But since the concepts they apply in perceptual experiences are not constitutive of the contents of such experiences, what makes the contents different is surely not the fact that they apply different concepts in perceptual experience. Whatever makes it the case that their contents differ (the simplest option is to say that they differ because they perceive different properties), it is not the fact that they apply different demonstrative concepts in experience that makes this the case. For the demonstrative concepts they apply have no influence on which perceptual content the subjects are in.

If this is true, then it would make no difference for the perceptual contents of the subjects if they had applied the same demonstrative concepts in experience, or no concept at all. Since these concepts are not constitutive of perceptual contents, whether they are applied or not is irrelevant to the question of whether two perceptual contents differ or not. So, it might well be the case that, for any two perceptual contents that differ, this difference is accompanied by a difference in the demonstrative concepts applied by the subjects. But this has nothing to do with the fact that the contents differ. It is a mere contingency that every difference in perceptual content is accompanied by a difference in the demonstrative concepts applied by the subjects. For the perceptual contents would differ irrespective of whether demonstrative concepts are applied in experience or not. So, if the demonstrative concepts we possess for fine-grained properties are not constitutive of perceptual contents, then it is certainly possible for two perceptual contents to differ even though there is no difference in the concepts applied by the subjects in their perceptual experiences. If demonstrative concepts are not constitutive of perceptual contents, two perceptual contents can differ irrespective of which demonstrative concepts are applied in perceptual experience, and irrespective of whether demonstrative concepts are applied in perceptual experience at all. It is then possible for two perceptual contents to differ even though the same concepts are applied in the perceptual experiences, as stated in (C1).

So, I take it to be clear that (C1) is true. But in order to show that the Constitution Argument is sound, I need to show that (C3) is true. I need to show that demonstrative concepts, which, we have been assuming, are the adequate concepts we possess for all the fine-grained properties presented to us in experience, are not constitutive of perceptual contents. But before going into this, we have to discuss what demonstrative concepts refer to. This will quickly get us to (C3).

5. The Reference of a Demonstrative Concept

I take it that there are two options, and only two options, regarding what a demonstrative concept refers to. They are:

Option 1

A demonstrative concept refers to an object or property before a subject.

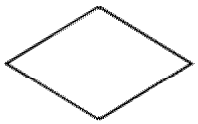
Option 2

A demonstrative concept refers to an object or property before a subject according to the

way this object or property is presented in the subject's perceptual experience.

In this section, I will argue that Conceptualists are forced to adopt Option 2. But, unfortunately for Conceptualists, Option 2 has as a consequence that demonstrative concepts are not constitutive of perceptual contents.[8] This is what I will show in the next section. But before going into any of this, we need to understand what is the distinction between Options 1 and 2.

This distinction can be made clear with an example. Consider the following shape:



Regular Diamond

Shape 1

Now, if we are sitting down with our head straight, this shade is perceived as a regular-diamond. But if we bend our body to the left (or right), the shape can be perceived as a square. And if we look at the shape with our head straight but it is moved to the right (or left), again we will perceive it as a square, as below:

Square

Shape 1

Thus, shape 1 can be perceived in two different ways, depending on the position of our body relative to it, and depending on its orientation relative to us. Now, suppose that, in perceiving the shape, I entertain a thought that can be expressed by the English sentence «Seeing that shape relaxes me». The question is: What does the concept THAT SHAPE, applied in such a thought, refers to? In answering this question we will see how Option 1 differs from Option 2.

According to Option 1, THAT SHAPE will refer simply to the shape before me, no matter

8. Peacocke (2001) also defends the view that Conceptualists, and everyone else, should adopt Option 2, but it is not clear in his work why this should be the case. And he also argues that Option 2 has as a consequence that demonstrative concepts are not constitutive of perceptual contents. So, his conclusions and mine are the same, but I try to provide a clear argument for the adoption of Option 2, while Peacocke does not.

how the shape is presented in my perceptual experience. It will refer to a shape in the world that can be perceptually presented either as a regular-diamond or as a square.

But according to Option 2, THAT SHAPE will refer to the shape as presented in my perceptual experience. Right now, given that my body is straight, the shape is presented in my perceptual experience as a regular-diamond, and not as a square. Thus, according to Option 2, when I entertain the thought that can be expressed by the sentence «Seeing that shape relaxes me», the concept THAT SHAPE, applied in my thought, refers to the shape in the world as presented in my perceptual experience. Thus, THAT SHAPE refers to the shape above presented as a regular-diamond. It does not refer to the shape above presented as a square. Or, to put it another way, the concept THAT SHAPE refers to regular-diamonds and not to squares, although these are only two different ways in which a same shape can be presented in perceptual experience.

Having clarified the distinction between Option 1 and 2, which one should we adopt? I will argue that we are faced with a serious problem if we adopt Option 1, a problem that does not appear if we adopt Option 2. The problem is the following. Let us go on considering the example above. The shape above is presented in my perceptual experience as a regular-diamond, and, while perceiving it, I entertain the thought that can be expressed by the sentence «Seeing that shape relaxes me». In this case, I take my thought to be about the shape as it is presented in my perceptual experience, about a regular-diamond. In fact, my thought may not be true at all of squares. It might be the case that only when presented in my perceptual experience as a regular-diamond the shape above relaxes me. When it is presented as a square it does not. It might be the case that, when the shape is presented in my perceptual experience as a square, I get anxious. Suppose this is in fact the case.

Let us now suppose that Option 1 is true. If it is, then my thought expressed by «Seeing that shape relaxes me» is about *both* regular-diamonds and squares. After all, the concept THAT SHAPE, that I apply in thought, refers to the shape in the world, no matter how it is presented in my perceptual experience. Now, given this, then my thought expressed by «Seeing that shape relaxes me» could *also* be expressed by either of these two sentences:

(1) Seeing square shapes relaxes me.

(2) Seeing regular-diamond shapes relaxes me.

Square shapes and regular-diamond shapes are, of course, the same shape. According to Option 1, my thought that can be expressed by «Seeing that shape relaxes me» is about the shape itself, no matter how it is presented in my perceptual experience. If this is the case, then my thought is about both squares and regular-diamonds, since they are, in fact, the same shape. Thus, given Option 1, my thought could be expressed both in terms of (1) and (2). Now, if this is the case, then although I am not aware of it, I am thinking what is expressed in (1) and (2).

But the problem is that my thought *cannot be expressed* in terms of (1). After all, (1) is false about me. Only when presented in experience as a regular-diamond does the shape in question make me relax. Presented the other way, as a square, the opposite reaction is caused in me: I become anxious. Thus, my thought is true only for the shape presented in experience as a regular-diamond. It is false for the shape presented as a square. This being the case, my thought cannot be expressed in terms of (1).

The problem can be put in general terms. Suppose that Option 1 is true. Then, there will

be certain demonstrative thoughts T that are about a property P in the world that can be perceived in two different ways. There will then be cases in which, according to Option 1, T can be expressed in the ways E_1 and E_2. The problem is that, in some cases, contrary to what should be the case given Option 1, T *cannot be expressed* in terms of E_1 or E_2.

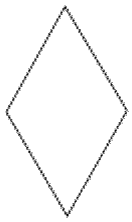
I take this to be a serious problem, one that disfavors the adoption of Option 1 as an account of what demonstrative concepts refer to.

Option 2, of course, has no problem whatsoever with cases such as the one above. According to Option 2, my thought expressed by «Seeing that shape relaxes me» is about the shape as presented in my perceptual experience. It is, then, about regular-diamonds, and not about squares. Thus, according to Option 2, the sentence «Seeing regular-diamond shapes relaxes me» is a way of expressing my thought, since my thought that the shape in question relaxes me is precisely about regular-diamond shapes. In this case, my thought cannot be expressed as (1), as «Seeing square shapes relaxes me». According to Option 2, my thought is not about the shape presented in experience as a square, so my thought cannot be expressed in terms of (1).

So, Option 2 does not face the problem with things that can be presented in perceptual experience in different ways that Option 1 faces. I take this to be a strong reason for us to favor Option 2 instead of Option 1 as an account of what demonstrative concepts refer to. And, thus, I do not think Option 1 is a viable route for Conceptualists to take in giving an account of what demonstrative concepts, the concepts we possess for fine-grained properties, refer to. In order to avoid the problem above, Conceptualists should adopt Option 2.

But there is another reason why Conceptualists should avoid Option 1. Simply put, it seems that, if Option 1 is true, then Conceptualism is false. Let me explain why.

Consider the following shape:



Non-Regular Diamond

Shape 2

Now, in the same way that Shape 1 can be presented in experience either as a square or as a regular-diamond, shape 2 can be presented either as a non regular-diamond or as a parallelogram, as below:

Parallelogram

Shape 2

This shape, a parallelogram, is the same shape as the non regular-diamond above, only orientated in a different way relative to the viewer, which leads to it being presented in experience in a different way. Now, it seems that at least most of us do not possess adequate general concepts for shapes such as these. Most of us do not possess words for them. Moreover, it seems that we would not be able to recognize such shapes. At least not in all their fineness of grain. We might be able to take a given shape as being a parallelogram. But this is not to recognize the shape in all its fineness of grain. For parallelograms can be of many sorts. An equilateral parallelogram is called a «rhombus». A rhombus with acute angles of 45 degrees is called a «lozenge». A rhombus with acute angles of 63.43 degrees is a golden rhombus. So, when we perceive a parallelogram, to recognize it as a parallelogram is not to recognize it in all its fineness of grain. For properties such as the degrees of its acute angles will make the shape in case a parallelogram of a given sort. And it seems unlikely that most of us would be able to identify such properties, and thus to recognize a parallelogram in all its fineness of grain, as being of a given sort. The parallel with the case of colour should be obvious. We are able to recognize a given colour as, say, red, but we are not typically able to recognize the specific shade of red perceived. The same goes to shapes. We might be able to recognize a given shape as a parallelogram, but it seems unlikely that we would be able to recognize which specific sort of parallelogram it is that we perceive.

Thus, it seems that we would not be able to satisfy the recognitional requirement for possession of adequate general concepts for such shapes. The requirement, remember, is:

Recognitional Requirement for General Concepts

> In order to possess an adequate general concept C for a fine-grained property P, one must recognize P — that is, take P as being of the same type as properties perceived in the past.

Since we are probably not able to recognize the shape above in all its fineness of grain, then, according to the Recognitional Requirement, which is widely accepted, then we do not possess adequate general concepts for shape 2.

A Conceptualist might reply, at this point, that we may not possess adequate general concepts for shape 2, but we possess demonstrative concepts. We possess, for instance, the concept THAT SHAPE. Let us assume that this is the case. Let us also assume that such a concept can be applied in perceptual experience. And let us also assume that Option 1 above is true. That is, let us assume that a demonstrative concept refers to an object or property before a subject, no matter how it is perceived. Now, if this is the case, then the only adequate concept we possess, and thus the only we can apply for what we perceive in all its fineness

of grain, when before shape 2, is the concept THAT SHAPE, which refers to the shape no matter how it is perceived.

So far, so good. Now suppose that there are two subjects, A and B, perceiving shape 2. Suppose that A perceives shape 2 as what we call above a non regular-diamond. And suppose that B perceives shape 2 as what we call above a parallelogram. But, like most of us, subjects A and B do not possess words or adequate general concepts for such shape. Thus, when before the shape, both subjects can only apply the concept THAT SHAPE, in order to capture what is perceived in all its fineness of grain. But I take it that it is uncontroversial that subjects A and B have perceptual experiences with different contents.[9] For A, the world is presented as having what we called above a non regular-diamond. For B, the world is presented as having what we called above a parallelogram. Thus, the contents of their perceptual experiences differ. But, and this is the important point, if Option 1 is true, then subjects A and B apply *the same concept* (type, not token) in perceptual experience. Both subjects A and B apply, when before the shape, the concept THAT SHAPE. Given Option 1, THAT SHAPE refers to the shape in the world, no matter how it is presented in the subject's experience. Thus, THAT SHAPE, when applied by subject A, refers to shape 2, not as a non regular-diamond, but to shape 2, no matter how it is presented in the subject's experience. And the same is the case with the concept THAT SHAPE when applied by subject B. It refers to shape 2, not as a parallelogram, but to shape 2, no matter how it is presented in the subject's experience. Thus, THAT SHAPE, when applied by subjects A and B refers to shape 2, no matter how it is presented in the subject's experience. So, the concept THAT SHAPE, applied by subject A, is the same concept (type) as THAT SHAPE, applied by subject B. So, subjects A and B, when before shape 2, apply the same concept. And there is no other concept at their disposal that could be applied instead for what is perceived in all its fineness of grain. Moreover, most of us would be in the situation of A and B, with no concept at our disposal but THAT SHAPE to apply when before shape 2.

However, if this is the case, then *Conceptualism is false*. For subjects A and B have perceptual experiences with different contents. But, given the truth of Option 1, they both apply the same concept (type) for shape 2. And they have no other concept at their disposal for what is perceived in all its fineness of grain. Thus, the content of the perceptual experiences of A and B are different, with no difference in concepts applied. But Conceptualism is the thesis according to which

Conceptualism

> For any two perceptual experiences A and B, the content of perceptual experience A is identical to the content of perceptual experience B if and only if the same concepts that are applied in perceptual experience A are applied in perceptual experience B in the same order.

However, if Option 1 is true, there can be (and there will be, given that the shapes such as 2 can be found in the world) perceptual experiences with different contents but with no

---

9. There is a debate between Peacocke and Tye regarding what makes the contents of the experiences of subjects A and B differ. According to Peacocke, they differ simply because the same shape is represented differently in the experiences of A and B. According to Tye, they differ because A and B perceive different properties. According to him, when you perceive the shape as a non-regular diamond, certain properties are perceived. When you perceive the shape as a parallelogram, other properties are perceived. But all sides seem to agree that the contents of the experiences of A and B differ. See Peacocke (2001) and Tye (2006).

difference in concepts applied. The problem is, of course, general, and not relative to shape 2. For any shape that can be presented in experience in different ways and for which we possess no adequate general concepts, we can only apply a demonstrative such as THAT SHAPE to them, if we are to apply an adequate concept to what is perceived. But, given Option 1, the concept THAT SHAPE will always refer to the shape in the world, no matter how it is presented in experience. If this is the case, then the same concept will be applied when a given shape is presented in experience to any two different subjects in two different ways. Thus, the contents of the perceptual experiences of the subjects will differ, with no difference in concept applied. If this is the case, then Conceptualism is false.

So, given that there are shapes that can be (and are) presented in experience in different ways, if Option 1 is true, then Conceptualism is false. So, not only Option 1 has the problem of obliging us to say that certain thoughts that we have can be expressed in certain ways, when they actually cannot be expressed in these ways, but also it has as a consequence that Conceptualism is false. So, it is not an option for Conceptualists to adopt Option 1 as an account of what demonstrative concepts refer to.

So, I take it that Conceptualists have to adopt Option 2 as an account of what demonstrative concepts refer to. Given Option 2, subjects A and B apply different concepts when before shape 2. THAT SHAPE, applied by subject A, refers to the shape presented as what we have called a non regular-diamond. THAT SHAPE, applied by subject B, refers to the shape presented as what we have called a parallelogram. Thus, THAT SHAPE, applied by subject A, refers to something different from what THAT SHAPE, applied by subject B, refers to. Thus, subjects A and B apply different concepts. And this is precisely the result that Conceptualists want. Different concepts are applied for perceptual experiences with different contents.

Conceptualists, then, have to take Option 2 as being their account of what demonstrative concepts refer to. However, if Option 2 is true, then demonstrative concepts are not constitutive of perceptual contents. This is what I will show next.

6. Demonstrative Concepts and Perceptual Experience

So, let us now assume that Option 2 is true. According to it,

Option 2

> A demonstrative concept refers to an object or property before a subject according to the way this object or property is presented in the subject's perceptual experience.

Given Option 2, a demonstrative concept THAT SHAPE, applied by us before a given shape, refers to the shape in the world as presented in our perceptual experience, and not to the shape in the world no matter how it is presented to us. But, if this is the case, then the demonstrative concept *is dependent on the content of perceptual experience*. The demonstrative concept gets to be about something, and thus gets to be a given demonstrative concept, only given the content of perceptual experience.

I take this to be an undeniable consequence of Option 2. If Option 2 is true, then demonstrative concepts are dependent on perceptual contents, in the sense that a demonstrative concept only gets to be about something given the content of perceptual experience.

Now, this being the case, then demonstrative concepts *cannot be constitutive of the content of perceptual experiences*. If they were, then they would be at least part of what makes a

given perceptual content the content that it is. A concept such as THAT SHADE, applied by me in perceptual experience, would be at least part of what makes the content of my perceptual experience the content that it is. In fact, if this were the only concept I apply in experience, then it would fully determine the content of the experience. It would be what makes the content of my experience the content that it is. So, if demonstrative concepts were constitutive of perceptual contents, then perceptual contents would be dependent on demonstrative concepts. But we have seen that demonstrative concepts are themselves dependent on perceptual contents. So, if demonstrative concepts were constitutive of perceptual contents, we would have that demonstrative concepts are dependent on perceptual contents and that perceptual contents are dependent on demonstrative concepts. This is, of course, viciously circular. It cannot be the case that demonstrative concepts are dependent on perceptual contents and that perceptual contents are dependent on demonstrative concepts. Something must give.

Since we have good reasons to think that Option 2 is true, and since Option 2 has as a consequence that demonstrative concepts are dependent on perceptual contents, then what must give is the idea that demonstrative concepts are constitutive of perceptual contents. Since demonstrative concepts are dependent on perceptual contents, they cannot also be constitutive of perceptual contents. As Peacocke puts it, since demonstrative concepts are «themselves individuated in part in terms of ways in which properties, magnitudes and relations are given in experience», they cannot «be used to elucidate the nature of such ways of being experienced» (2001:250).

So, I take it that demonstrative concepts cannot be constitutive of perceptual contents, as stated in (C3) of the Constitution Argument. The Constitution Argument, remember, is:

(C1) If the adequate concepts we possess for all fine-grained properties presented to us in perceptual experience are not constitutive of perceptual contents, then it is possible for the contents of two perceptual experiences A and B to differ even though the same concepts that are applied in A are applied in B.

(C2) The only adequate concepts we possess for all fine-grained properties presented to us in perceptual experience are demonstrative concepts.

(C3) Demonstrative concepts are not constitutive of perceptual contents.

(C4) Thus, the adequate concepts we possess for all fine-grained properties presented to us in perceptual experience are not constitutive of perceptual contents (from C2 and C3).

(C5) Therefore, it is possible for the contents of two perceptual experiences A and B to differ even though the same concepts that are applied in A are applied in B (from C1 and C4).

(C6) If (C5), then Conceptualism is false.

(C7) Therefore, Conceptualism is false.

The argument is clearly valid. I have tried to show that it is also sound. If am right about this, then we have good reasons to think that Conceptualism is false.

7. Conclusion

In this paper, I have argued that, even if Conceptualists are able to block the Fineness of Grain Argument, Conceptualism still seems to be false. For, even if this is the case, Conceptualism is still subject to the Constitution Argument, which has as its conclusion the falsity of Conceptualism. I have argued that the argument is both valid and sound. If that is the case, we have good reasons to think that Conceptualism is false.

Acknowledgments

I wish to thank Stephen Laurence, Rob Hopkins, David Papineau and Dominic Gregory for comments on an earlier version of this paper. Thanks are also due to Capes Foundation, for funding this research.

Bibliography

Bonjour, L. 2003. «A Version of Internalist Foundationalism». In L. Bonjour & E. Sosa, *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. London: Blackwell.

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

Burnham. R.W. & J.R. Clark, 1955. «A Test of Hue Memory». *The Journal of Applied Psychology*, 39:164-172.

Chuard, P. 2006. «Demonstrative Concepts Without Re-Identification». *Philosophical Studies*, 130, 153-201.

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

Heck, R. 2000. «Nonconceptual Content and the Space of Reasons». *Philosophical Review*, 109:483-523.

Kelly, S. 2001. «Demonstrative Concepts and Experience». *Philosophical Review* 110, 3: 397-420.

McDowell, J. 1994. *Mind and World*. Cambridge (MA): Harvard University Press.

Peacocke, C. 2001. «Does Perception Have a Nonconceptual Content?» *Journal of Philosophy*, 98, 5: 239-264.

Tye, M. 2006. «Nonconceptual Content, Richness, and Fineness of Grain». In T. Gendler & J. Hawthorne (eds.), *Perceptual Experience*. Oxford: Oxford University Press.

Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

André Abath

The University of Sheffield

Sheffield — UK

andreabath@gmail.com

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 67-73
Free-will and determinism: A debate in sociology

FREE-WILL AND DETERMINISM: A DEBATE IN SOCIOLOGY

by Jorge Gibert-Galassi

Introduction

The endless debate around the crisis in sociology and in social sciences in general, can be interpreted as the reissue of the discussion on their scientific status, in particular on the topic about the capacity to generate «legalform» propositions, as well as the plausibility of the explanation and the prediction based on the resulting products of knowledge of such a capacity.

Historically, this discussion was developed through antinomical positions as «understanding or comprehensive sciences versus natural ones»; «nomological versus ideografical sciences»; «determinism versus probabilism», and so on. In sum, who postulated that the social sciences are scientific and who postulated that they are not (or that they are «in their own way»).

We will outline our problematization about the epistemological hypothesis, which are on the base of the discussion.

I

The consequences for the generation of laws, about the fact of social contingency and individual freedom, could be visualized with the support of the resolution between determinism and indeterminism of the sociological laws.

The focuses to approach the topic of determinism are: a) as a property or characteristic and b) as a constant and univocal connection.

Let us exemplify a) like a propositional function ‘Fx’, that is to say the group of individuals of a specific class (‘x’) that satisfies a certain predicate (‘F’) and that, in consequence, by means of a classified operation indicates us that the property or characteristic determines the members of a certain class.

Let us exemplify b), analogically to the equation x-1 = 0; where the unique solution is x = 1 (where the universe is IR).

What is supposed in both cases is a kind of invariance that has been used as argument to whom postulates that science deals with the «objective order» or the underlying real structure of the apparent reality. However, strictu sensu, is not more than the philosophical hypothesis that facilitates the scientific work. Nevertheless, the problem arises when diverse stream of thoughts confuse the epistemological statute of the legal propositions and, confusing it with the univocal description, they state that in the case of the social realities, such «objective order» does not exist, because, if it were, we would deny the human freedom. Taken to the extreme, such a position indicates that exists a radical separation between social realities

investigations and the no-social ones, since in the first case the phenomena emerge endowed with intentionality and will, while in the second one, not. This would reveal the incapability of the pretense of apprehending the social reality, which would be ontologically slippery and not objective (because it is free). The will and the intentionality would only produce rather amorphous structures, unstable and ephemeral, elusive to a legal treatment.

But a brief exam indicates us the spurious nature of such a radical separation.

In the case of a), it would be necessary to state that all determination supposes an operation of «free» classification. The relevant character of the classificatory principle is given by some available theory that has been freely accepted by the scientific community.

In the case of b), we could make it complex by means of another equation such as x + y-1 = 0; what tells us that, although it expressed a rigid interdependence among x and y, another equation is needed to solve the problem; being indeterminate (where x,y e to IR).

Perhaps what the antinomical thinker would tell us it is that, in the case of the social realities, the classifications are not governed by the objectivity: no social phenomena is identical to another and the observer's reading always depends on his point of observation. But such an argument forgets the principle of indiscernibility of the identicals, which is worth in a no-empty way for the conceptual objects and worth in an empty way for the material objects. Regarding the material or concrete objects, modern science has arrived to a different proposition to that of Leibniz: it is only acceptable a «partial» identity or an «approximate» identity between two or more concrete objects. Regarding the observation point, we will only remember that at playing ping-pong in a train, one would find that the ball obeys Newton laws in the same way that it would do it in a table placed next to the railroad; which supposes a solution and a problem: the solution is that the theory describes the movement of the ball no matter if the railroad or the train are in state of rest.

But it outlines a problem: our ping-pong ball is bouncing and hitting the table in the same place with an interval of one second; but the two boats for an observer located next to the railroad will seems to take place with a separation of about 40 meters. What implies that we need a theory to join space and time concepts in one (space-time) which includes the explanation of this new problem but, in any case, it implies that we «opt» for the observer's description that is inside the train or for whom is on the railroad's side.

The general objective is to insist in the thesis on the scientific statute of the sociology and social sciences in general, affirming the pretense of objectivity, unity of the scientific method and refuting the separation and/or distinction between social sciences and nature sciences.

The general theoretical foundation is made by means of the following argument: If does not exist social regularities, any at all, we could not drive our individual life nor even our collective life, making impossible the exercise of freedom, since this supposes the visualization of the objective selection.

The work hypothesis that I present is the following: The fact of social contingency and individual freedom deserves a specific treatment for the determination category within the framework of the determinants existence hypothesis.

II

One of the obstacles for the generation of laws is the plausibility of «the self-fulfilling prophecy» and «the double contingency» theorems.

Although the logical definition of «theorem» is that of a demonstrated proposition which starts from non defined concepts (» p; p v q) and axioms; the term has been imported to the sociology and it is used with smaller rigor. Therefore, in regards to our problem, we have two theorems in which varied focuses lean on to sustain the irreductibility of social behavior to laws: the self-fulfilling prophecy and the double contingency. Indeed, those theorems debate the anticipatory capacity of the sociological knowledge, the prediction based on theories and data.

The prediction, for the purpose of this brief exposition is an anticipatory enunciated about the occurrence of an event given certain conditions. For example, where (x,y) e IR, the equation of a straight line: Y= mX; where m represents a constant (here: m = 2/3); X to a given condition (here: X = 6); and Y the proyectandum or projected value of Y (here: 4). These would be the data (or circumstances) and the theory is the analytic geometry. The logical structure is {Law (s), Circumstance (s)} $\rightarrow$ Proyectandum.

What happens in social sciences?. It is declared: «Michael will buy a tie tomorrow.» Within the following context 1) Michael belongs to the high financial executive class, 2) one of the daily events in the routine of this class members is the formal business meeting, and, 3) tomorrow is a work day. It is also added the circumstance that Michael has been victim of a robbery of all his ties yesterday. Leaving aside extreme contingencies, such as Michael were run over and he had to stay a day in a city hospital or that it were suddenly instituted the habit of going to formal business meetings without tie, I believe that we could think that the prediction is reasonable and good. But, even so, it could happen what in sociology is denominated the self-fulfilling prophecy theorem, which consists, in this example, in that Michael prophesies «tomorrow I will forget to buy a tie». And maybe happen, although it should not happen. That is to say, in all social prediction, the individual (or a group or a society) could influence in the predicted result, since he is implied in the «tomorrow» circumstance. That is to say, it is as if we induced to Y to say «I am worth 8» because we want that X to be equal 12; because if we wanted that X were 24, then we would try to achieve that Y could adopt the value of 16. The prediction is contingent regarding who has formulated it. It is the phenomenon of «The self-fulfilling prophecy».

Thus, in an given interaction between an individual A and another B, every action either originated in A or B acts as a boomerang over that same A or B. An action towards certain goals, for example, from A to B, could generate unanticipated consequences to this action, and therefore would achieve other goals. A courtesy greeting probably generates a similar answer or gratefulness, but it could also generate a claim or anything and, therefore, modify the course of original action, which started on that greeting. It is a case of a hunted hunter or the case of Don Juan who falls in love, being caught in his own web. Also in this case, the contingency is present, but in a peculiar way: it is what has been denominated social sciences the theorem of the double contingency, which indicates that the contingency of my behavior options is altered by the contingency of the behavior options of the other one and vice versa, unstabilizing probable interaction courses.

An analytic linking could exist between both theorems, in particular if it is acceptable to derive one from the other or if both are implied somehow.

Theoretically, we sustain the plausibility of this achievement with the help of the common property for both propositions, that is, the individual free will.

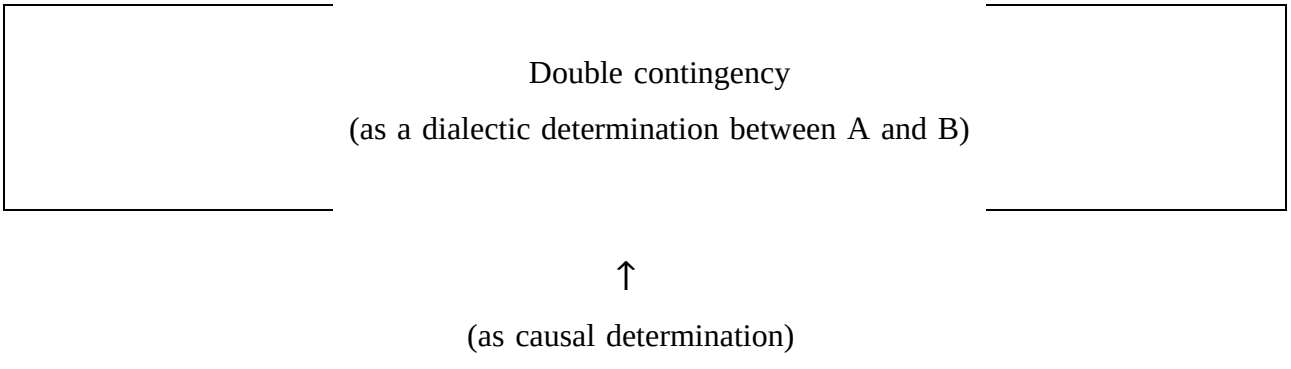
The relevance of such a connection (if some exercise illustrated appropriate plausibility) would allow, in my opinion, to develop a group of difficulties and paradoxes, especially the one that links explanation and prediction. The argument is the following: Being the prediction a proposition not immediately contrastable (when I foresee, in this moment, «tomorrow it will rain», I delay the contrast for tomorrow): is this true or false as proposition?. And if that proposition does not assume some of these values: how is it possible that certain legal propositions engender something that is not a proposition?. That is to say: being both deductive «inferences starting from laws and data», and obviously, both also (from this analytic point of view) components of the scientific investigation; how is it that the predictive proposition escapes to the truth — falsehood bivalent logic?

The «whim» of freedom, uncertain by definition, makes that the existence of both theorems implies the property of being uncontrollable to the public opinion and predisposed to deny fallibility; that is to say, it closes the possibility of the bivalent logic (truth and falsehood), not allowing falsehood to have an occurrence probability. This means, what happens is that all prediction is evaluated with posteriori and, in the case of the human actions, we have the possibility to intervene the facts to refute it. A consequence of the above-mentioned is the defense of the thesis of freedom, demonstrated by means of the following mental experiment. Supposing that we could know future events with certainty: would not we maybe try to change these events to adapt them to our particular interests?.

But neither we can deny the existence of determinants. Also, if we use the formula of causation as a necessary production, we see that the required «conditions» for the existence of a causal bond are contingent with regard to the cause, this would refute the formula. The problem lies upon the conditions: if they cannot be determined in a conclusive way (that is to say that could eventually exist «other» undetermined conditions, that would produce the same effect, but with some theoretically no relevant variation in magnitude), then, we are in presence of a non causal determination but an statistic one (freedom as a no-explained variance).

Our generic hypothesis, in a minimum social model (A and B), is that the analytic treatment of both theorems is the following: The self-fulfilling prophecy of A and / or B as long as statistical regularity, it is a causal component of the entirely self-determined qualitatively (the interaction of A and B), which, in consequence, would be determined dialectically.

The model is the following:



Self-fulfilling prophecy

(as a statistic determination of A and B behavior)

III

The epistemic status of the category denominated «subject» in sociological theory constitutes, without a doubt, a central axis of the discussion.

The focus by means of which I will treat the epistemic status of «subject» category, according to the exam of this exercise, I will denominate it paradoxical. It is summarized affirming that who postulate «subject» as the central category of the sociological theory are indebted of a maximalist thesis regarding the subject relevance in the societal evolution, and which, in consequence, are not able to settle the problem of social contingency. This problem can be cleared, if we conceive society as millions of interactions, it becomes evident that the pretense to coordinate social life, negotiating senses, it is completely unlikely (therefore, to build a better world is not possible, since it is not reducible to a good will matter). On the contrary, the maximalist thesis of those who subtract importance to the subject, only identifies the determinations or conditions of social life, this brings as a result the improbability of human freedom. That society is ontologically a mix of both theses, it is something well known by every sociologist. What interests us to emphasize is that the relationship of incompatibility between pretense and postulate has not been coded as paradox: who intend to liberate the man of their chains (the very well-known thesis which postulates that men are the ones who produce and reproduce the conditions of their existence), postulate the coordination of the individual actions, that it is clearly unlikely though not impossible; on the contrary, who do not have this pretense postulate the study of the conditions that it would be necessary to overcome so that this pretense could be acceptable.

Epistemology has historically had as main object of investigation the relationship between an observer and a cognoscible object. The hypothesis that matter cannot know itself has been the base of the very fruitful separability between subject and object, as requirement of all objective observation. It has been the angular stone of the development of the natural sciences and, according to the interpretive sociologists, it is what allows to sustain what differs essentially this from social sciences. The subject's implication in the prophetic enunciation makes spurious the content of the scientific propositions in social sciences. The content of sociological inquiry returns to its creator, becoming a factor that contaminates, more than it elucidates, the plot it tries to explain: it is more an «actor» or an «action». On the other side, the scientific «ethos», this separability is unavoidable, at least as a rigid methodological organization that should accompany all scientific investigation (in particular, the exclusion of logical contradictions). Historically, the solutions have been two: to conceptualize individual and his action as «thing», «body», «behavior» or another entity category or property observable directly; or, to include it in such a way that it is the sociological discipline the one that has been excluded from science or reconverted in social philosophy when not in mere ideology. But some intermediate positions also exist, where subject is outside sociological theory as long as it pretends to explain society, but it is recognized as an existent condition of society the existence of individuals or subjects connected to it. However, our position is

epistemological and not theoretical.

A work hypothesis could be that the outlined problem is not a problem, since postulating the thesis of freedom, we affirm subject and reject determinism; and the other way around, if we postulate the determinist thesis, we affirm science and reject freedom. In both cases, what we make is to contradict the same definition of social sciences as long as they are sciences based on subjects with free will.

IV

Nevertheless, the consequences of a legal architecture in sociological theory are summarized in if it is, maybe, possible to explain and to predict social events.

The habitual focus to describe what an explanation is, it is what constitutes the inference deductive, where starting from at least two true propositions or explanans, a third one is generated or explanandum, as the legal consequence of the information articulation of explanans. From this point of view, the explanation is never empiric, but logic. The explanans should possess some defined characteristics, maybe the most important one it is that both propositions contain law formulations of the same level of reality.

But, in the case of the social sciences, most of legal propositions are not universal in a exclusively way, but rather hypothesis that confirm certain rules or statistical regularities, local existentials, singular and pseudosingulars, etc. Thus, their reach is usually limited.

On the other hand, the sociological reflection has abused of structural or holistic determination outlines, regarding the parts for the everything (any event is conditioned or determined by the society as a group, what «explains» everything and nothing at the same time), question that has subtracted power to the explanation, besides it has generated a shortage of credibility inside scientific community, trivializing many of its no-trivial sociological investigation results.

In the case of the prediction, most of the analysts have opted for the framework theorem: As bigger the width of the frame, bigger will be the predictability of the event. Some examples of this reasoning are: «Tomorrow the NY Stock Exchange will work» or «In some place of the world an earthquake will happen». In both cases the foresight has big occurrence possibilities, since it is not specified how the stock exchange will work (with an up or down tendency), or because it is not specified the place where the earthquake will happen. However, in both cases, those analysts confuse common sense prognosis with scientific prediction.

Every epistemological position in sociology, I affirm, should discard the antinomy determinism and freedom, being superfluous and futile.

Thus, in philosophical reflection about contemporary sociology, the problem of explanation could be guided to answer queries such as the following ones: do sociological laws exist or only social regularities exist? which would be the laws of the double contingency? until what point and under what base some constants discovered by statistical modeling could be validated ? do law or social regularities exist at margin of law or historical regularities? do the contents impact on the explanans in the deductive inferences?.

While in the case of prediction: what consequences could be extracted for the prediction, the existence of our two theorems (self-fulfilling prophecy and double contingency)? will it be possible to regulate or to control their incidence in the formulation of predictive propositions?.

In the meanwhile, the empiric investigation should insist in the elaboration of statistical models and the compared international studies, besides small scale social experiments.

The general objective of an investigation towards that direction should be in insisting in that the unified science way, the one that tries to explain and to predict, is the correct one; and to exam as many as theories in this respect could it be possible. If the individual freedom and double contingency are social particular properties, it will be necessary to exam what special characteristic or restrictions we can impute to the explanation and prediction in the case of sociology, instead of rejecting the requirements of the explanation and scientific predictions in general.

Jorge Gibert-Galassi
University of Viña del Mar (Chile)
jgibert@uvm.cl

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 74-79
Free Agency and Self-Esteem

FREE AGENCY AND SELF-ESTEEM

by Robert F. Allen

The reflective acceptance approach (RAA) to free agency states that an agent has acted freely if and only if she would reflectively accept the process by which her motive was formed. Its adherents eschew the notion that the *content* of an agent's motive determines whether or not she acts freely, believing instead that any volition she forms, no matter how self-destructive or bizarre it seems to others, is an impetus to a free action just in case she would positively appraise its genesis.[1] Here, I define the role of self-esteem in promoting free agency, in order to meet the following objections to RAA's content neutrality.

The first objection, developed by Paul Benson,[2] is based on a case of «gaslighting,» that is, a case in which a person comes to *mistakenly* believe that she is unable to make socially acceptable decisions, others having misled her into thinking that she is mentally incompetent (657). The case of gaslighting presented in the 1944 film *Gaslight* is not a counterexample to RAA. There, a husband employs various machinations whose effect is his wife's belief that she is 'losing her mind'. (This loss of self-confidence keeps her from realizing that he is after something valuable that is hidden in her house.) Benson, however, devises a case of this type for which RAA yields counterintuitive results. It involves inadvertent misleading. To wit, a respected Victorian psychologist attributes his wife's hysteria to cognitive deficiencies, rather than emotional trauma. Such a diagnosis being then generally accepted, the woman loses confidence in her decision-making. Feeling «unworthy» to act in the presence of others, she, then, avoids socializing (656). According to Benson, that the former case involves deception while the latter does not entails that the latter victim, but not the former, would reflectively accept the deliberations that proceeded her resolving to forego taking certain decisions. Thus, because of its content neutrality, RAA seems to produce the counterintuitive result that the woman in Benson's case acts freely, her false belief that she is cognitively deficient notwithstanding.

This argument, however, invites the following response. Would the woman in his example reflectively accept her deliberations were she made aware of the fact that she has been misdiagnosed? That her husband's misinforming was inadvertent, rather than intentional, does not seem to make her decision any less regrettable than the film victim's. In both cases, believing a falsehood leads to an unwarranted and severe loss of self-confidence. If a rational agent would disavow any instance of decision-making involving misinformation, then

---

[1]. This view is developed by John Christman's in «Autonomy and Personal History,» (*Canadian Journal of Philosophy*, XXI, 1, 1991, pp. 1-24).

[2]. Paul Benson, «Free Agency and Self-Worth,» *The Journal of Philosophy* 91: 650-668. Page numbers in the text and notes refer to this article.

Benson's case of gaslighting is no more a counterexample to RAA than the film's version. In neither case would there be reflective acceptance of the process by which the victim's motive is determined.

This response, in turn, raises the following question. Does exercising one's deliberative skills upon falsehoods *always* entail a loss of liberty? That is, would a rational agent reject any set of deliberations that involved misinformation? Answering here in the affirmative yields counterintuitive results. One obviously does not suffer a loss of liberty merely because one decides to take an umbrella upon hearing an inaccurate weather forecast. At this point an epistemic constraint must be placed upon reflective acceptance that would provide for the above rebuttal without turning all decisions based upon misinformation into losses of liberty.

Benson would have us see the gaslighting victim losing freedom because of the content of her motive. The proponent of RAA, he contends, cannot offer her preferred explanation: a flaw in (the process of) decision-making the reflective realization of which would justify its being abjured. But if that process is supposed to occur under certain conditions, which are not met in a case of gaslighting, then there is a way of handling such a case that preserves RAA's content neutrality. The defender of RAA, I maintain, should distinguish here between an *obstacle* to the exercise of one's deliberative skills and their *exercise* upon misinformation. In general, one may distinguish between being unable to exercise skills under certain circumstances and exercising them under less than ideal circumstances. In a case of gaslighting, the victim's deliberative skills are virtually useless, insofar as their proper exercise, which requires a certain amount of self-confidence, is precluded by the impairment of that which generates it: «primary self-esteem» (PSE), that is, the ability to evaluate one's (other) abilities, character, and conduct.[3] The extreme mistrust that Benson describes seems best understood as stemming from the reluctance to 'take a good look at oneself' — that is, damage to one's PSE. This impairment means, on the view sketched above, that they will unfreely in foregoing certain choices, lacking enough self-confidence to gainfully deliberate. One remains a free agent, on the other hand, if one resolves upon a piece of misinformation whose reception leaves one's PSE intact, even if one would reflectively abdicate the reasons upon which the decision was taken. Authoritatively issued false statements to the effect that one is either emotionally disturbed or cognitively impaired might be expected to greatly lessen one's PSE. However, believing untruths of other sorts is not generally similarly deleterious. E.g., having one's job performance unfairly critiqued may cause a temporary loss of self-confidence. But, leaving one's PSE intact, it would not render one incapable of recovering from such a 'blow to one's ego'.

Thus, there is a way of handling Benson's objection to RAA without establishing an excessively strict standard for free agency. The reflective acceptance in question must be understood as having been reached after one has been apprised of all the facts pertinent to the operation of one's deliberative skills. In particular, as highlighted by gaslighting cases, it is important that one be provided with the truth concerning the employment of one's PSE. Each gaslighting case involves an agent one of whose beliefs has diminished that ability whose exercise is a necessary pre-condition of gainful deliberation: one must realize that one is a skillful deliberator, capable of making sound decisions, before one is willing to engage in it. Upon realizing this impairment, each would reject the way in which her motive had been formed: seeing that the proper conditions for the exercise of her deliberative skills did not

---

[3]. Robert J. Yanal, «Self-Esteem,» *Noûs* 21, (1987): 363-79.

obtain. A person who had been misled by a weather forecast, on the other hand, would not reject the *process* by which her volition was formed, realizing that it was an application of those skills to the best available information. To reasonably disclaim the way in which a motive was formed, an agent would need evidence supporting either the belief that she lacked no such skills or the belief that their exercise was somehow impeded. A gaslighting victim apprised of her plight would be justified in believing the latter.

Benson responds here by contending that the victim in his gaslighting case does *not* suffer from an impairment the recognition of which would warrant renouncing the way in which she developed her attitude towards decision-making. Despite «ceas(ing) to *trust* herself to exercise (them) competently,» she retains deliberative skills.[4] His contention appears false, however, when we consider how PSE affects those skills. While it may be true that the now «isolated» wife is yet able to attend to her personal needs, there are many socially important decisions she is no longer willing to make, having lost not only her store of self-confidence, but the ability to regain it. She, thus, retains her deliberative skills, as Benson contends, but is incapable of exercising them in all but the most inconsequential situations. E.g., one can imagine her being unwilling to decide with whom to socialize. Being thus «disassociated,» (Benson's term) from her deliberative skills, they *are* less effectual than they would otherwise be: there are applications of them that she would be able to make but for her diminished PSE. Recognition of this fact would provide a good reason for rejecting the way in which her motive was formed. She would realize that she had given up on herself for no good reason.

Benson's other putative counterexamples to RAA may be similarly handled. *Shame*, he tells us, may «(undercut a) person's sense of self-worth (leaving) him wanting to hide, even to disappear, in order conceal his shameful weakness or inadequacy from view. ... (I)t tends to be disorienting, disrupting behavior and producing confusion in thought» (658). The loss of freedom here may again be seen as stemming from damaged PSE. I would argue that the «sense of being worthy to act,» which Benson posits as essential to free agency, is a part of self-confidence, and, as such, required to exercise one's practical reasoning skills. Regaining this sense is a function of one's PSE. Thus, as long as this capacity is left inoperable by shame, one is unable to act freely. How could shame have such an effect? As Benson notes, it makes one want to «hide,» presumably from oneself as well as others. That is, shame diminishes PSE by fostering the reluctance to exercise it.

It should be noted that feelings of shame do not always entail a loss of freedom. Benson should thus be taken to be commenting upon a «morbid» or unhealthy sense of failure, one that is disproportionate to the suspected offense. Indeed, self-abasement often expresses one's autonomy, as one realizes a difficult truth about oneself and sustains one's disapproval independently of how others may feel. Even the choice of a life of servitude, should it be reflectively accepted as one's calling — as in the case of Jesus Christ and his followers — must be taken as free.[5] Here we see a virtue of RAA: it does not force us to treat a person who has acted from such a motive, which is very difficult for some, maybe most, to understand, as no less unfree than those who are held against their will in bondage.

Slavery may also be seen as inimical to free agency because of its deleterious affect

---

4. In correspondence.

5. Gerald Dworkin also discusses such a case in «Paternalism: Some Second Thoughts,» in Rolf Sartorius ed., *Paternalism* (Minneapolis: University of Minnesota Press, 1983) pp. 105-11.

upon its victim's PSE. A slave, one imagines, would have a hard time retaining her sense of being a competent decision-maker, capable of directing her affairs, as the result of oppressive assaults upon her self-confidence. This loss is the (almost) predictable result of the dehumanizing treatment to which she is continuously subjected. That her humiliation *keeps* her estranged from her deliberative capacity is a sign of her debilitated PSE. For fear of viewing her degraded self-image, she cannot effect its rehabilitation and, thus, recover her sense of being able to determine how to act. It was only by overcoming fear of this sort that freed American slaves were able to become soldiers in the war against their former masters, leaving powerful testimony not just to their courage, but also to the indominability of their wills (as dramatically portrayed in the film *Glory*).

My rejection below of a «social component» of free agency, however, allows me to think of *some* slaves as having acted freely. Were not being enslaved a necessary condition of being a free agent, as Benson maintains, then no slave would be capable of acting freely. But some slaves did perform free acts: witness those who escaped along the Underground Railroad or Frederick Douglas turning on his master. Here we see examples of unbroken, nay, indomitable wills. Such cases, however, being rare, bondage can then be seen as evil precisely because it typically destroys a vital part of the slave's mind, leaving her incapable of *deciding* her own destiny.

Flattery or false praise, which Benson does not discuss, may also be seen as a potential threat to free agency for the same reason as are shame and humiliation: it can incapacitate one's PSE. The person 'taken in' by flattery is convincingly misinformed concerning her character, appearance, or abilities. Part of her self-image then becomes based on the proffered untruths. So long as she remains capable of revising the resulting picture of herself her PSE is intact, and, thus, *cæteris paribus*, she can act freely. However, her dependency upon hearing the false praise may increase to the point where she becomes incapable of taking 'a good hard look' at the misbegotten aspect of her self-image. One may speak then of the «atrophying» of her PSE (commonly known as smugness). Think of a supervisor who secures the performance of a menial task by regularly singling out a charge lacking in self-confidence as «the only member of the department who can do the job *right*.» The victim here, as in the gaslighting cases, would reflectively reject the process leading up to the carrying out of the misinformant's wishes, realizing that his misguidance had incapacitated a part of her will, making her unfree.

It should not be thought, however, that flattery immediately causes a complete loss of PSE. Rather, it initially deprives its victim of a prerequisite of that ability's exercise, viz., the desire to rethink one's self-image. Thus, flattery can leave one for a time *virtually* incapable of evaluating oneself. Moreover, as with any other capacity, prolonged lack of exercise of one's PSE spells it attenuation, if not complete diminishment. Therefore, being gaslit may be both a short-term and a long-term effect of flattery. Hubris, insofar as it stems from «self-flattery,» may, thus, be seen as a product of diminished PSE. It reflects the loss of the ability to correctly gauge one's self-worth, as that is determined by the extent to which one realizes one's ideals. Here it is one who is paying oneself false compliments. As long as one lacks the sense that such impairment exists, the desire to rehabilitate one's PSE will be lacking, perhaps leading to megalomania.

Finally, to close our discussion of the ways in which gaslighting is tied to impaired PSE, there is the case of someone being tormented with the truth. For it is not only misinformation concerning oneself that is potentially psychologically damaging but also

undeniable faults, if they are presented as immutable, something over which one has no control. I can just as easily damage an agent's PSE by constantly pointing out the character flaws she herself acknowledges, suggesting that they are beyond repair, as I can by repeatedly boosting her ego. In either event, she may be left incapable of evaluating herself. I debilitate no less than does the cajoler and to the same effect.

As noted, Benson further contends that gaslighting cases demonstrate that freedom has a «social component»: that one must trust one's ability to *publicly* explain one's conduct in order to act freely (660-63). That is, were one to believe oneself incapable of giving an account of one's doings to others, one would cease to be a free agent. It was demonstrated above that free agency requires PSE. The question now becomes, does having PSE entail being confident of publicly explaining oneself?

It does not. Consider the case of Friedrich, who has come to believe that he is a hopeless misfit, unworthy of social interaction of any sort. Does he thereby suffer a loss of liberty in deciding to become a hermit? Is this a choice he would reflectively reject? Not unless he realized that the exercise of his deliberative skills whereby he arrived at that decision was in some way impaired. But his believing that he is incapable of explaining his decision to others would not necessarily have the same effect as gaslighting. He might yet retain his PSE. A victim of gaslighting is devoid of PSE, the loss of which renders her unable not only to respond to others' questioning of her conduct, but to also her own inquiries into her motives. Friedrich, on the other hand, might still possess the ability to measure his conduct/character against *his* ideals, despite being incapable of persuading others that he is acting rationally. Thus, this inability does not entail a loss of liberty on his part. No matter how low he thinks he has sunk in the eyes of others, it is, to modify Shakespeare's dictum, only to himself that he must be able to answer in order to exercise a free will. Thus, it has not been demonstrated that free agency has a social component.

In Friedrich's case, we need only imagine that he at one time disposed himself to withdraw from society should his *mitwelt* become intolerable. This rule may be something others cannot understand; they may deem his response to his situation to be wholly inappropriate. He himself may realize that no one could understand his decision, even if they appreciated his plight. The gap in others' understanding here does not entail, however, a lack of rationality on his part. As long as it has been true of him that he would have reacted to his present circumstances by completely avoiding contact with others, he is doing what *he* once meant to do in the event that they obtained. He is expressing himself, a part of his character, albeit in a way others find incomprehensible.[6]

[6]. This view, of course, conflicts with the conclusion of Wittgenstein's private language argument — that it is impossible for an individual to follow a publicly inexplicable rule. (Cf. his *Philosophical Investigations*, New York: Macmillan Publishing Co. 1968, pp. 80-2.) I would challenge, however, that argument's key premise, which states that within a private practice correctness must be subjective, it can be nothing more than what an individual thinks is correct. The antidote to such skepticism is the individual disposition theory. (Cf. my «Rule-following, Dispositions, and Infinity,» unpublished manuscript.) According to this view, one is following a rule iff one is applying it as one would have had one been asked when establishing it to extend it to the case at hand. Cf. Saul Kripke, *Wittgenstein on Rules and Private Language*, Cambridge, Mass: Harvard University Press, 1982, pp. 23-37.

Conclusion

Proponents of RAA may allow that gaslighting, shame, oppression, flattery, and torment entail a loss of liberty without conceding that free agency involves more than the unimpeded exercise of deliberative skills. Given that PSE monitors this faculty, its diminishment by such affects renders it impaired. The motives formed under their influence would, thus, be reflectively rejected as the products of an unhealthy mind. Thus, Benson has failed to produce a counterexample to RAA. PSE, moreover, does not require the ability to explain one's actions to others. It is enough to be capable of evaluating oneself, assessing, according to one's personal standards, the reasoning behind one's choices. Freedom lacks a social component; an alienated person may yet be free.[7]

Robert F. Allen
allen1rf@cmich.edu

[7]. Thanks to Joel Anderson, Paul Benson, Jay Campbell, Joshua Gert, Drew Hinderer, Paul Hughes, Karen Jones, Robert Kane, Andrew Melnyk, Marina Oshana, John Pauley, Lawrence Powers, Michael Reed, Ralph Forsberg, and Bruce Umbaugh for helpful comments on earlier drafts of this paper.

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 80-93
The Interpretive Mind

THE INTERPRETIVE MIND

by Peter Francis Colbourne

The Perceptual Double Barrier

We are part of the universe, as much a part as the planets, the stars, the nebulae and the galaxies, while at the same time experiencing two degrees of separation from it. On the one hand, our physical bodies are made of the same fundamental particles as all matter formed into atoms and molecules. On the other, the processes by which we perceive the universe — including indeed our own bodies as part of that universe — place a double barrier between our minds and that universe.

Firstly, we rely on distinctive physical pathways for our only perceptions of that universe: we call them senses. Thus, energy waves and friction act on our physical receptors and signals are transmitted via networks of nerves to various areas of our brains where an impression of the extant physical world creates an experience for the mind — a sense of touch, sight, sound or smell. Of that process of transportation between the external physical realm and the mind we have no experience whatsoever. The phenomena occur without any volition on our part and we do not possess a sense that 'feels' the passage of the impulses through our nerve fibers. Everything we experience through our senses occurs within our brains, and our minds — however they may be conceived — are conscious of the end product, the firing of the neurons that create the sensations for us to experience. We are thus always and forever locked within the confines of our skulls, experiencing the illusion of touch, smell, sight and hearing in a mini-universe that is devoid of tactility, odour, light and sound.

Our minds have no direct contact with the external physical universe but rather we are aware of the response of a tiny portion of that universe — our brains — to electrical and chemical signals triggered in response to the body's remote sensing system. This is a defining factor of our experience of the external world that the seventeenth century English philosopher Thomas Hobbes was entirely aware of. These are his four explanatory points on the «conception» of objects:

> That the subject wherein colour and image are inherent, is not the object or the thing seen.
>
> That that is nothing without us really which we call an image or colour.
>
> That the said image or colour is but an apparition unto us of that motion, agitation, or alteration, which the object worketh in the brain or spirits, or some internal substance of the head.
>
> That as in conception by vision, so also in the conceptions that arise from other senses, the subject of their inherence is not the object, but the sentient. (c1641)

The first barrier to our experience of any objective reality we are part of is therefore a physical one — it is the very body we inhabit and which mediates sensation for us[1]. Furthermore, the senses that create what appears to us as an incredibly rich experiential world of smells, sights, sounds and touch are in themselves highly limited faculties. Many animals have a sense of smell that is a million times more sensitive than ours. Bats and dolphins have sonar. Eagles have much more acute vision than we do, and goldfish can see into both the infra red and the ultraviolet. Wonderful as it is, our world of sensation is a highly restricted one.

Secondly, any processing system must interpret the data received so that meaning can be inferred. A face has to be recognised as a face; the call of a bird as birdsong. Our brains interpret the raw data — the nerve impulses — and create sense and order. Without this processing system[2] there would be no perception of a coherent external reality, a point also made by Dan Sperber (1997). This is starkly highlighted by people in whom this process has been damaged in some way and who hear colours and see sounds, for example, or who suffer from hallucinations.

This process is therefore the second barrier to our perception of the universe, for the process is an interpretive one and requires both something to interpret experiences against, for comparative purposes, and through, for the assignment of meaning[3]. This latter issue may be illustrated with two linked examples. Picture yourself watching a sunrise. The horizon lightens and eventually the sun rises above it. Morning has arrived. In this case, photons have impacted on your retinas and nerve impulses have been decoded or interpreted to represent the experience of a sunrise. Further, being a twenty-first century person, you perceive the movement of the sun to represent the spinning of a roughly spherical earth in space around a hydrogen-burning star. Thus, the horizon you see is for you the curved edge of our round planet.

Now transport yourself back into the body and mind of an individual in medieval Europe who is observing a sunrise there. Again, the architecture of this person's brain will create the experience of a sunrise. However, what she sees is a heavenly sphere circling the earth. Moreover, the horizon for her is the edge of the earth over which the unwary explorer is likely to fall. For both yourself and our medieval acquaintance, it is not the objective reality that imparts meaning to the experience but the belief system. For human beings, believing is seeing.

This phenomenon of the subjectivity of human perception has been graphically illustrated in a number of research projects on the unreliability of eyewitness reports.

---

1. This position coincides with Varela, Thompson and Rosch's (1991) argument that knowledge coevolves with the knower and not as an outside, objective representation. Varela et al go on to propose that this knowledge is a function of the whole cellular organism that is the human individual rather than any one subset of it as I have discussed here. Bruno Latour (2005) has developed a modification of this model, proposing that lived meaning informed by experience negates the apparent barrier between the object observed and the subject observer. My alternative argument here is that we also have the lived experience of being separate from the object.

2. For an overview of the neurodynamical and phenomenological models for this processing system, see Gallagher and Varela (2001).

3. David Bohm and F. David Peat (2000) have discussed the related concepts of selection and collection under the heading of «category formation» (p.112).

Consistently, individuals will report seeing activities that did not happen or people who were not there (Bizzell et al 2000; Wright, Gail and Justice 2000; Poole and Lindsay 2001; Haber and Haber 2001; Wright, Loftus and Hall 2001). It has been estimated that currently 5000 people are incarcerated in the Unites States of America on the sworn and incorrect testimony of eyewitnesses.

The physical limitations of our perceptual systems combined with the interpretive functions of the brain make these errors inevitable. We shall tend to 'see' what we are predisposed to see and will probably be convinced that our interpretation represents the only 'objective' reality.

The impact of the double perceptual barrier on our processes of inquiry is significant. An object (or phenomenon) acts as a stimulus on our sense receptors and signals are sent to the brain. To use a visual phenomenon as an example, the brain then reconstitutes an impression of the object. We thus do not «see» the object but rather its reconstituted image. That image may be poorly or wrongly realised because of the limitations of our perceptual systems and the processes of interpretation we impose on the image. For example, we may decide that the red car parked next door is, as usual, our neighbour's Ford when it is in fact a visitor's Toyota. We have «seen» what we expected to see.

Our relationship with the external universe is therefore both second hand and problematic. Our minds, however they may be defined, do not and can never directly perceive the 'objective' universe. We certainly construct interpretations and representations of that universe — or at least parts of it — but those phenomena are not the universe itself, as David Hume noted over two hundred and fifty years ago.

> … nothing can ever be present to the mind but an image or perception, and that the senses are only the inlets, through which these images are conveyed, without being able to produce any immediate intercourse between the mind and the object. (1748, p. 174)

In relation to our formal inquiries into the universe of which we are a part, scientific research may be conceived of as a vehicle for moving our perceptions and beliefs closer to an accurate description of objective phenomena — or as Simon Blackburn (1999) has put it: «Science … contains within itself the devices for correcting the illusions of science» (p. 232). Also, scientific theories are tested against observation, or research, and may be discarded or modified accordingly. The process of science may therefore be seen as a codified attempt to align the reconstituted, internal conception of the universe, and its recorded descriptions in writing and mathematics, with the objective extant universe[4]. (In terms of the hypothetical example above, a further examination of the car in question may reveal what make it is and to whom it belongs.)

However, Karl Popper (1966) has identified the problematic nature of this process:

> … the concept of truth plays mainly the role of a regulative idea. It helps us in our search for truth that we know there is something like truth or correspondence. It does not give us the means of finding truth, or of being sure that we have found it even if we have found it. (p 28)

To explore this issue further, it is certainly the case that scientific inquiry is concerned primarily with the search for truth, or the search for accurate descriptions of phenomena, and

---

4. Clive Beck (1993) has expressed a similar view: «Knowledge is the product of an *interaction* between our ideas about the world and our experience of the world.»(p. 3)

science, in general, does overtly address the problems of subjectivity and perceptual plurality through both its methodologies and an evolved process of peer review[5]. Popper (1953) referred to this process as *falsificationism* — that is, the applicability of scientific theories is assessed against how inaccurate they are rather than how accurate. To take the above example again, it is an identification of the inaccuracies of the 'theory' concerning the neighbour's car that would lead to adjustments and not any confirmation of its accuracy. The modern system of scientific inquiry is by far the most effective yet developed for establishing descriptions of the physical properties of phenomena, devising technologies to manipulate those phenomena and creating physical outcomes. The system may be just as much belief driven as any other human endeavour and always open to the inconsistencies that are inherent in the human perceptual system, but a process of critical analysis, or belief modification, is inherently part of it. Thus, neither the effectiveness of science in forming and sustaining our technological society nor its role in developing relevant representational theoretical structures is questioned here.

What is questioned is the objectivity of these processes, both in their inception and their execution, and the extent to which they may be applied to phenomena. Robert Maxwell Young (2000) has expressed this view as follows:

> … science, technology and medicine — far from being value neutral — are the *embodiments of values* in theories, things and therapies, in facts and artefacts, in procedures and programs. I also believe that all facts are theory-laden, all theories are value-laden and all values occur within an ideology or worldview. (p. 2)

In other words, the quest for scientific truth is always to some extent vulnerable to the values and perceptions of the scientists involved as well as to the problematic nature of truth itself.

Dualism and Monism

This subjective/objective paradox has a long and dynamic history. The second perceptual barrier I have identified above is a re-statement of the «dualism» first proposed by Plato[6] and established as a principle of human psychology in the Western scientific tradition by Descartes (1649). This view of human personality suggests that mind and body are not only separate, but can also exist separately from each other. Descartes believed that the mind and the body were composed of different substances that interacted through the pineal gland. Other theorists, and in particular Malebranche (1674) and Leibniz (c1696), developed variations on this concept, but for all three men God was an essential element of their hypotheses, either as the creator of mind, or as the medium through which the mind and body interacted. Eventually, the concept of dualism was replaced by that of «monism» — or the belief that mind and body were aspects of the one phenomenon, initially introduced by Spinoza (c1677). Although Spinoza saw God as the unitary force in this model, the view of human beings as *machines* (La Mettrie, 1749) soon became central to this hypothesis, and came to focus on the mind as being a product of chemical or neurological processes in the brain (Wozniak, 1992). This latter form of the model denied the metaphysical aspect of dualism and became the orthodoxy as theistic explanations for phenomena became less accepted by scientists.

---

5. Or as Willard Quine (1951) put it, «…our statements about the external world face the tribunal of sense experience not individually but only as a corporate body.» (p. 36)

6. In the *Phaedo* (360 BCE).

Variations on these rationalist theories remained current until well into the nineteen-eighties when both Roger Sperry (1987) and Michael Gazzaniga (1993) reopened the debate. The basic premise of both neurobiologists was that we *experience* mind as separate from body and that, therefore, they may be considered conceptually separate. Further, both support Descartes' *interactionist* model that assumes a mutually influential relationship between mind and body.

> Scientific theory has become squared finally with the impression of common experience; we do in fact use the mind to initiate and control physical actions. (Sperry, 1987, p. 166)

However, both Gazzaniga and Sperry, and later Searle (1999), also supported a monist view of consciousness that saw this phenomenon as an emergent property of the functioning of the brain — a view compatible with complex systems theory. That is, mind and body (or the brain in this case) might be experientially separate, but nonetheless are part of the same phenomenon — the biological entity that is the human brain[7].

The model I have proposed takes this one step further. In the *Interactive Mind*, Sperry, Gazzaniga and Searle's concept of the experientially separate mind is assumed — that is, mind and body may be experienced as separate but may not be able to function without each other. However, the dimension of the affective/subjective observer that I have proposed significantly alters the outcomes that might be considered to arise from the functioning of this model. For example, Descartes' model presumes not only God as *agent*, but also that mind can objectively observe body — or, in its universalist form, all matter. Thus, all that is separate from mind becomes an object to be observed and used. Descartes' combination of the two concepts of the separation of mind and matter and the objective observer entails objectifying all that is non-mind. For Descartes, mind is God created and God given and, because God is incapable of error, mind can objectively observe God's creation — the extant universe. Descartes (1649) explicitly expressed this view of man the controller of the natural world in the *Meditations*.

This may be represented as follows:

'Mind' the objective observer

*****

Interaction between mind and matter direct and objective

*****

Matter consists of objects to be observed

*****

Therefore matter consists of objects to be described and used

Figure 1: Cartesian Dualism

However, if the objective dimension is removed, the consequence is very different:

---

7. John Dewey was ahead of his time here and discussed a similar model in *Democracy and Education* (1916).

'Mind' an affective/subjective phenomenon

*****

Matter probably extant but not directly observable

*****

Therefore perception of matter is a function of mind to be analysed and questioned

Figure 2: Postmodernist Monism

In this case, no direct relationship between mind and matter is assumed. Rather, the physical media and the perceptual and intellectual processes through which mind pursues and interprets a relationship with matter are assumed to be problematic and possibly objectively impenetrable. Thus, the focus is removed from an objectified universe to an affective/subjective interface between mind and matter. In this model, the conception of the extant universe as an object to be manipulated is removed at least one dimension, and unless the objective/subjective paradox is resolved, may indeed continue to be excluded from it.

The consequences for the Descartes worldview are profound. Far from a God given ability and right to control the rest of nature, a consideration of the rationales for any interventions becomes an issue. In other words, the features of the relationship between the inquiring mind and the universe can be seen to put into question all motivations and justifications for actions that impact on that universe. Albert Einstein (1930) made this point as follows:

> A human being is part of the whole called by us *universe*, a part limited in time and space. We experience ourselves, our thoughts and feelings as something separate from the rest. A kind of optical delusion of consciousness. This delusion is a kind of prison for us, restricting us to our personal desires and to affection for a few persons nearest to us. Our task must be to free ourselves from the prison by widening our circle of compassion to embrace all living creatures and the whole of nature in its beauty … We shall require a substantially new manner of thinking if mankind is to survive. (reported in *Einstein's wisdoms*, 2002)

Essentially, the concept of the subjective observer places ethical and moral considerations at the heart of the debate on human activities, and that must be particularly true of science and technology as such overwhelmingly potent forms of those activities, as Roger Sperry (1987b) has noted:

> Instead of maintaining the traditional separation of science and values, cognitive theory says the two come together in brain function. If we are correct in saying that our conscious mental values not only arise from, but also influence brain processing, then it becomes possible to integrate values with the physical world on a scientific rather than supernatural basis …. On these new terms, science no longer upholds a value-empty existence, in which everything, including the human mind, is driven entirely by strictly physical forces of the most elemental kind. (p. 3)

The emergence of a recent form of monism, which I have called *postmodernist monism* to reflect the relativistic nature of the model, thus highlights a crucial dimension of the relationship between the inquiring mind and the extant universe. The assumption that the scientist is in some way transformed into an objective agent — some sort of machine — for analysing phenomena, is not supported by this model. Rather, the conception of the scientist as an affective agent exploring his or her relationship with the universe seems unavoidable. Indeed, as Karl Popper (1966), Paul Feyerabend (1975) and J.D. Trout (2002) have detailed, assuming an unquestioned objectivity has had significant consequences in the history of

scientific inquiry, resulting in a litany of false conclusions that were considered to be evidently true.

For illustrative purposes, I have included two examples here that demonstrate the overt influence that an assumption of objectivity, or of the objective observer, has on all aspects of the processes of inquiry from the formation of the hypothesis, to the choice of methodologies and to the analysis of results.

The Peppered Moth Affair

Biston betularia, or the peppered moth, was the subject of a series of classic experiments by English botanist Bernard Kettlewell (1955, 1956) that seemed at the time to confirm the processes of natural selection detailed by Charles Darwin in The Origin of the Species (1859). The peppered moth exists in various shades of grey from a near white to a near black, with individuals normally exhibiting lighter shades interspersed with a speckling of darker scales. It had been noted by Tutt (1896) that darker, or melanic, varieties were replacing lighter ones near heavily industrialised areas. He hypothesised that the lighter varieties were more easily located and predated upon by birds when the moths rested on the dark trunks of trees exposed when industrial pollution killed the lighter coloured lichen that had covered them. It was assumed that natural selection then resulted in a progressive darkening of the population.

Kettlewell's experiments were designed to test Tutt's hypothesis. He first established that the moths were predated upon by birds. He then, through a process of releasing and recapturing moths in polluted and non-polluted woodlands, established that melanic individuals were twice as likely to survive in polluted environments, with the opposite being the case for lighter individuals. He therefore concluded that the theory of evolutionary change, and particularly natural selection, was supported.

Kettlewell's work was widely lauded by botanists and geneticists (Wells, 1999) and his experiments on the peppered moth became, and still are, essential study for biology students as a classic example of natural selection in operation.

However, Kettlewell made two crucial and unfounded assumptions when framing his hypothesis and the research to explore that hypothesis. First, he assumed that evolutionary change was driven by simple cause and effect relationships — in other words, that a simplistic model for natural selection could be observed in nature — and secondly, and most importantly, he also assumed that an empirical research methodology could accurately and effectively be utilised to explore and describe that process. Kettlewell then went on to complete the paradigmatic circle — he found what he expected to find.

Subsequent research has thrown significant doubt on both Kettlewell's methodologies and his conclusions.

> The geographic distribution of melanic peppered moths did not fit the theory: the frequency of melanics was not as high as it should have been in some places, and higher than it should have been in others. Furthermore, melanism is not correlated with lichen cover; in the U.K., it declined before lichens returned to the trees, while in the U.S., it occurred despite the presence of lichens and declined without any perceptible changes in lichen cover. Finally, peppered moths do not normally rest on tree trunks: instead, they normally rest under horizontal branches high in the canopy, not where Kettlewell and his successors had carried out their experiments on selective predation. (Wells, 1999)

Exactly why peppered moth populations exhibit significant changes in pigmentation between generations is not yet clear. Hypotheses include air born pollutants such as sulphur dioxide, genetic drift and thermal melanism (Wells, 1999). Whatever the underlying causes are, however, it is clear that Kettlewell's conclusions are unsupported. Although his experiments clearly demonstrated evolutionary change between generations, both the choice of methodologies to explore the causes for those changes and the conclusions drawn from the empirical results are founded on unproven assumptions.

The essential point to be made here is that if the classical objectivist paradigm is considered the way to do science, then Kettlewell's work is to be admired:

> Bernard Kettlewell was a good scientist. Even now, almost half a century after his initial experiments, Kettlewell's scientific papers make exciting reading. (Wells, 1999)

Kettlewell conducted his research in an era when the classical paradigm for research methodology was relatively unchallenged. He was simply following the almost universally accepted and expected processes for exploring phenomena. The challenges to this hegemony from the complexity sciences[8] and postmodernism were yet to come. Nevertheless, it was the beliefs, the assumptions, that Kettlewell carried into his research that informed the hypothesis, structured the research and presupposed the analysis of the results. Kettlewell therefore found what he was looking for — what he believed he would find. However, beliefs cannot automatically be translated into facts, or truths, about the objective universe external to our minds. Kettlewell got it wrong in this instance not because he was a poor scientist, but because he was enculturated into the norms and mores of his community. He — together with myself and all those who adopted the objectivist paradigm for research — was misled by a subconscious assumption about the nature of phenomena and how those phenomena might be explored.

Wells (1999) has neatly summed up the issue raised by the debate over Kettlewell's work:

> The classical story, elegant and appealing though it may be, should no longer be presented as a textbook example of evolution in action. If the purpose of science education is to teach students how to do good science, then instead of re-telling the classical story textbooks would do better to focus on how science revealed its flaws.

In the context of my argument, one of the greatest 'flaws' revealed by a focus on science are the often unconscious beliefs and assumptions each of us carry into our work. Bernard Kettlewell had little choice but to do so in his research — at the time there was only one major paradigm in existence that could inform and guide his work. However, the same excuse should not be accepted today. The same series of experiments repeated now with the same hypotheses and methodologies would be highly questionable. Furthermore, the conclusions should certainly be very different as well. Kettlewell's research projects were classics of their time and indeed I studied and appreciated them as well. The point here is not a criticism of Kettlewell, but rather that changes in the assumptions applied to any outcome of the processes of scientific inquiry can change completely the value that may be placed on that outcome and the methodologies utilised which led to that outcome. In Wells' terms, this work may now be considered an example of a flaw in the way science is conducted rather

---

8. Although Warren Weaver (1948) had already heralded its birth: «Science must, over the next 50 years, learn to deal with these problems of organized complexity.»

than an example of best practice. The automatic assumption of an objective observer who experiences a direct and unalloyed perception of the external world can no longer be supported.

The Flight of Archaeoraptor

In 1999 a fossil discovery was announced by the National Geographic Magazine in its November edition (Sloan, 1999) that was hailed at the time as one of the most significant in the history of palaeontology. Unearthed in the Liaoning Province of China and bought at an Arizona minerals show in the USA, the fractured pieces of shale seemed to contain the fossil of a missing link between dinosaurs and birds. The fossil exhibited the arms of a primitive bird, and possibly feathers, and the tail of a theropod — a small dinosaur. The fossil was dated at about 125 million years old.

The missing link between the two genuses had been sought by palaeontologists since the discovery of the first archaeopteryx fossil in 1860 near Solnhofen in Germany. Archaeopteryx was a primitive bird-like species with feathers that the scientist who named it, Hermann von Meyer, suggested had evolved from a dinosaur. Until archaeoraptor, no evidence of intermediary links between the genuses had been discovered.

Within a year, archaeoraptor was shown to be a fake. By an incredible stroke of chance, Chinese palaeontologist Xu Xing obtained what turned out to be the counter slab of part of the archaeoraptor fossil from a farmer and sometime amateur fossil hunter in Liaoning Province. The deposits in the area had been formed in layers that often split apart when removed. Thus a fossil can be split into two complimentary halves, known as the slab and the counter slab. Xu Xing compared his fossil with pictures of archaeoraptor and realised that the latter must be a composite, for not only were completely different pelvises exhibited on each slab, but also the rear legs bones in the photographs were mirror images of each other — that is, they were yet another slab and counter slab of a single fossil leg arranged in archaeoraptor as if they were separate limbs. His email to the National Geographic's Christopher Sloan was a bombshell. The error was National Geographic's most embarrassing in its 110 year history.

Lewis Simons was commissioned by National Geographic to investigate the error. His (October 2000) report details what he called:

> … a tale of misguided secrecy and misplaced confidence, of rampant egos clashing, self-aggrandizement, wishful thinking, naïve assumptions, human error, stubbornness, manipulation, backbiting, lying, corruption, and, most of all, abysmal communication. (p. 1)

He eventually established that archaeoraptor was created by a Chinese farmer from fragments glued together of at least two separate and completely unconnected fossils. Further analyses by Timothy Rowe et al (2003), utilising high-resolution x-ray analysis, has demonstrated how the fake was created, including the use of grouting to hold it together. Archaeoraptor was nothing more than a clever combination of two stone jigsaws.

As it has turned out, neither of the two identifiable fossils that made up the archaeoraptor composite were of species previously known to science. They have now been classified and named as, respectively, a fish-eating bird Yanornis martini and a small bipedal meat-eating dinosaur Microraptor zhaoianus (Mayell, 2002). However, neither specimen may be considered an intermediary link between dinosaurs and birds.

In relation to the context of this paper, the most important finding to come out of Simon's investigation was that nearly everybody involved, amateurs and expert scientists alike,

saw what they expected, or perhaps wanted, to find — a missing link between species presupposed by the neo-Wallace (1858) and neo-Darwinian (1859) theories of natural selection. After all, if the theories are accurate reflections of a real and measurable effect in the natural world then such links must exist and finding the relevant fossils may be simply a matter of time. Nevertheless, the important point to be made here is that it was a set of assumptions that drove the conclusions. The objectivist paradigm, embodied in this case in the application of the current theories on speciation, predisposed the result. Indeed, Simon makes the point that without the amazingly unlikely finding of the counter slab by Xu Xing then archaeoraptor may well still be considered one of the most important discoveries in palaeontology. In other words, the story of the finding of archaeoraptor may have been considered an example of the modern scientific method revealing an important truth or fact about the history of our world. As it is, archaeoraptor has existed only for a short time, and then, like Pegasus, has flown only in the imagination.

Conclusion

As human beings we have an experience of mind that appears to separate us from the universe of which we are a part — an existential dualism that engenders and supports the illusion of the objective observer. The assumptions and beliefs implicit in the adoption of the conventions attached to the objectivist paradigm tend to remove the individual as an affective parameter in these processes. Rather, the relationship the individual has with the world — for example as a scientist — may be considered an impersonal one. The ethical dimension of our interventions in the universe then becomes secondary to objectifying the phenomena we investigate and act upon, or as Denzin and Lincoln (1998) have noted ethics is extrinsic to this form of inquiry process. The emergence of a recent form of monism arising out of neurobiological research challenges this conception and instead highlights the subjective interface between the inquiring mind and all phenomena. As a result, ethics is placed at the centre of all considerations of scientific inquiry. As Roger Sperry (1987b) noted in one of his last papers:

> We're beginning to learn the hard way that today's global ills are not cured by more and more science and technology. Technical solutions … only tend, over time, to escalate the problem. What is needed to break the vicious spiral is a worldwide change in attitudes, values, and social policy. (p. 3)

I would argue that only by recognising and taking into account, both as individuals and collectives, the impact of objectified dualism on our processes of inquiry can the subjectivism of the interpretive mind be embraced and these changes potentially achieved. The concept of an existence separate from physical reality in some way — whether this be in terms of Plato's 'intellect' (360 BCE), Descarte's 'mind' (1649) or Aquinas's (1266-71) 'soul', for example — allows a detachment from the world we inhabit that underpins most forms of inquiry that we take part in, including scientific inquiry. That detachment can no longer be supported. Rather, as Benz and Shapiro (1998) state:

> … methodology is parasitic on epistemology and ontology, and we believe that an individual who uses a particular research method without being able to articulate its epistemological and ontological assumptions and preconditions is not a fully human, fully responsible researcher. (p. 34)

It seems clear that it is time to accept being fully human and embrace the complete range of our intellectual gifts — for interpretation and subjectivity as well as for rational analysis — so that we may increase our attempts to address the ills of the world rather than continue to add to them through an insupportable pursuance of the objectivist paradigm. This

is not, however, a denigration of the modern scientific method — it is simply a call to extend it. Indeed, it is important to note here that it may be that no form of inquiry would be possible without our ability to be an observer — without the illusion of objective dualism. Through this ability, the inquiring mind can reflect on all phenomena it encounters, including itself. It is one of our defining and most profound gifts. It is nevertheless also problematic and double-edged and, unleavened by an integral consideration of the subjective and interpretive nature of our processes of inquiry, can engender outcomes that are not only false, but potentially divorced from ethicality. We cannot afford scientific endeavours that assume an objective detachment from the world we live in.

References

Aquinas, Thomas (1266-71) *Summa Theologica*. Fathers of the English Dominican Province, trans. 2d, rev. ed. 22 vols. London: Burns, Oates & Washbourne, 1912-36; reprinted in 5 vols., Westminster, MD: Christian Classics, 1981.

Beck, Clive. 1993. «Postmodernism, Pedagogy, and Philosophy of Education». *Philosophy of Education*. http://www.ed.uiuc.edu?EPS/PES-Yearbook/93_docs/BECK.HTM

Benz, V.M. and Shapiro, J.J. 1998. *Mindful Inquiry in Social Research.* Thousand Oaks: Sage Publications.

Bizzell, Beth; Morgan, Meredith; Povill, Dave. 19 April 2000. *Eyewitness Testimony in the Real World.* http://www.unc.edu/~jdumas/projects/eyewitness.htm.

Bohm, D. and Peat, F. D. 1987. *Science, Order and Creativity*. London: Routledge.

Blackburn, Simon. 1999. *Think.* Oxford: Oxford University Press.

Burdett, Alfred. 2002. «Subjectivity in Science». *Natural Science.* On-line journal accessed 8/01/2003. http://naturalscience.com/ns/articles/comment/ns_com.html

Darwin, Charles. 1859. *On the origin of the species by natural selection.* London: John Murray. The Writings of Charles Darwin on the Web. http://pages.britishlibrary.net/charles.darwin/texts.htm

Denzin, N. and Lincoln, Y. 1998. *The Landscape of Qualitative Research: Theories and Issues.* Thousand Oaks: Sage Publications.

Descartes, Rene (1649) 1864. *Les Passions De L'Ame.* Paris: The Napoleon Chaix Edition.

Dewey, John. 1916. *Democracy and Education.* Accessed on the web 6/05/2003. http://www.columbia.edu/publications/Projects/digitexts/dewey/d_chapter25.html

Feyerabend, Paul. 1975. *Against Method.* Source: *Analytic Index.* http://www.marxists.org/reference/subject/philosophy/works/ge/feyerabe.htm. Accessed 16/05/2003.

Einstein, Albert. (1930) 2002. Quoted in *Einstein's wisdoms.* http://www.ludwig.ucl.ac.uk/st/einst.html

Gallagher, S, and Varela, F. 2001. *Redrawing the Map and Resetting the Time: Phenomenology and the Cognitive Scxiences.* In Crowell, S., Embree, L., Julian, Samuel J., (eds), The Reach of Reflection: Issues for Phenomenology's Second Century. Centre for Advanced Research in Phenomenology Inc. Electronically

Published at www.electronpress.com

Gazzaniga, Michael S. 1993. *Brain mechanisms and conscious experience.* In: Ciba Foundation Symposium 174. Experimental and Theoretical Studies of Consciousness (pp. 247-257). New York: Wiley.

Haber, Ralph Norman and Haber, Lyn. 16 November 2001. *Meta-Analysis of Research on Eyewitness Lineup Identification Accuracy*. Paper presented at the Annual Convention of the Psychonomics Society, Orlando, Florida.

Hobbes, Thomas. (c1641) 2002. *The Elements of Law Natural and Politic* (c1641) in *Hobbes* Kemerling, Garth http://philosophypages.com/ph/hobb.htm

Hume, David. 1748. *An Enquiry Concerning Human Understanding.*The Internet Wiretap online edition. http://www.knuten.liu.se/~bjoch509/works/hume/human_underst.txt.

Kettlewell H.B.D. 1955. «Selection experiments on industrial melanism in the Lepidoptera». *Heredity* 9:323-342.

Kettlewell H.B.D. 1956. «Further selection experiments on industrial melanism in the Lepidoptera». *Heredity* 10:287-301.

Latour, B. 2005. *What is the style of matters of concern? Two lectures in empirical philosophy.* University of Amsterdam. Published on the Bruno Latour website. www.ensmp.fr/~latour/

Lloyd, Peter. 1994. «The Physical World is a Fiction». *Philosophy Now* 11.

Gottfried Wilhelm Leibniz (c1696) 1998. in *Philosophical Texts*, ed. by Richard Francks and R. S. Woolhouse. Oxford.

Malebranche, Nicolas. 1674-1675 *De la recherche de la vérité.* Paris: Chez André Pralard.

La Mettrie, Julien Offray de. 1749. *Man a Machine.* Translated from the French of the Marquiss D'Argens. London: Printed for W. Owen.

Mayell, Hillary. November 2002. *Dino Hoax Was Mainly Made of Ancient Bird.* National Geographic News. http://news.nationalgeographic.com/news/2002/11/1120_021120_raptor.html

Plato (360 BCE) *Phaedo*, translated by Benjamin Jowett, in The Internet Classics Archive, *http://classics.mit.edu/Plato/phaedo.html*

Poole, Debra Ann and Lindsay, D. Stephen. March 2000. «Children's Eyewitness Reports After Exposure to Misinformation From Parents». *Journal of Experimental Psychology.* 7(1): 27-50.

Popper, Karl. 1953. *Conjectures and Refutations.* Lecture given at Cambridge. http://cla.calpoly.edu/~fotoole/321.1/popper.html. Accessed 15/05/2003.

Popper, Karl. (1966) 1972. *Objective Knowledge* (1966) in *Objective Knowledge*, Clarendon Press. http://www.marxists.org/reference/subject/philosophy/works/at/popper.htm

Quine, Willard Van Orman. 1951. «The Two Dogmas of Empiricism». *The Philosophical Review* 60: 20-43.

Rowe, Timothy, Ketcham, Richard, Denison, Cambria, Colbert, Matthew, Xu, Xing and

Currie, Philip J. 2000. *High-Resolution X-ray Analysis of the Archaeoraptor Forgery.* http://www.ctlab.geo.utexas.edu/pubs/nature2000

Simons, Lewis M. October 2000. *Archaeoraptor Fossil Trail.* National Geographic Magazine. http://ceirp.cornell.edu/Review/NGArtHlt.html

Sperber, Dan. 1997. «Intuitive and reflective beliefs». *Mind and Language* 12(1): 67-68.

Sperry, Roger W. 1987. *Consciousness and Causality.* In R.L. Gregory, Ed., Oxford Companion to the Mind. Oxford: Oxford University Press, pp. 164-166.

Sperry, Roger W. October 1987b. Interviewed in *On Campus* (4:2).

Searle, John R. October 1999. *Consciousness.* Published on the web. http://ist-socrates.berkely.edu/~jsearle/rtf/Consciousness1.rtf

Sloan, Christopher P. 1999. «Feathers for T. rex?». *National Geographic* 196(5): 98-107.

Spinoza, Benedictus de. (c1677) 1870. Benedict de Spinoza; His Life, Correspondence, and Ethics. By R. Willis, M.D. London: Trübner & Co.

Trout, J.D. June 2002. «Scientific Explanation And The Sense Of Understanding». *Philosophy of Science* 69: 212-233.

Tutt JW. 1896. *British Moths*. London: George Routledge.

Varela, F. Thompson, E. and Rosch, E. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: The MIT Press.

Wallace, Alfred Russel. 1858. *On the Tendency of Varieties to Depart Indefinitely from the Original Type.* S43. The Alfred Russel Wallace Page. http://www.wku.edu/~smithch/index1.htm, Accessed 3/04/2003.

Wells, Jonathan, 1999. *Second thoughts on peppered moths*. http://www.arn.org/docs/wells/jw_pepmoth.htm

Weaver, Warren. 1948. «Science and Complexity». *American Scientist* 36: 536.

Wozniak, Robert H. 1992. *Mind and Body: René Descartes to William James.* Published on the web http://serendip.brynmawr.edu/Mind/Table.html. Accessed 24/04/2003.

Wright, Daniel B.; Self, Gail; Justice, Chris. 2000. «Memory conformity: Exploring misinformation when presented by another person». *British Journal of Psychology* 91: 189-202.

Young, Robert M. June/July 2000. *Science and the humanities in the understanding of Human Nature.* On Line Opinion. http://www.onlineopinion.com.au/June00/Young,%20R.htm

---

Peter Francis Colbourne
5 South Arm Road
Rokeby. Tasmania
Australia 7019
peter.colbourne@tqa.tas.gov.au

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 93-97
On the Empirical Reality of Pure Mathematics

ON THE EMPIRICAL REALITY OF PURE MATHEMATICS

by Jonathan C. Sampson

By sensibility is meant the things we take for granted when we think about a mathematical statement. It will be the same, we reckon, the next time we think about it, and it cannot be altered by an arbitrary agency. The most renowned challenge to this kind of supposition appears in the Meditations (Descartes 1641). So it is ironic that the Cartesian system of reals, a seamless complementation of geometry and number theory, is where that sensibility breaks down. This metaphysical wobble was detected by Freiling (1986), who supposed that it constitutes a probabilistic argument against the Continuum Hypothesis. To catch the philosophical enormity of Freiling's discovery, we must elaborate a setting wherein both, the logical necessity of probability and the generalisation of truth to fact, arise.

1.

Imagine a rational being, ignorant of mathematics, who one day inherits a functioning calculator. Eventually, it is discovered that the principles connecting the numbers are related with operators, etc. Because the being is ignorant, he never knows what the calculator will do in advance. That the machine always does the same thing is at the level of natural certitude, like the sun rising each morning. The very next level of sureness assumes that a perfect intuition can take hold. That sureness is challenged for the real number system.

Probability will be defined in the aleatory sense, as describing what is rationally foreseeable to a consciousness innately ignorant of the future. By admitting this interpretation, we allow that mathematics has some frail connection to the natural world. That the sun will not rise tomorrow might be termed a *preternaturally problematic projection* (PPP), one with which our species has a long affiliation (e.g., the Mexica). Although this projection has a probability of 0, it is not formally false, like ½ + 1 = 2. By the same token, the law of gravity is not formally true, but an empirical fact, with probability 1. Although usually implied, probability (or its equivalent) is logically necessary to make statements of fact. The singular feature of our enquiry is that it rejects ordinary equivocations about this minimal humility.

If one's attention is restricted to facts and PPPs, probability is equivalent to any measure of ignorance, so long as it is logically consistent with respect to statements and conjunctions of statements. For example, a deterministic machine might contain some hidden variable, of which the observer has no existential certainty. Then events corresponding to singularities in the variable are observed with probability zero. See Dürr, Goldstein, and Zanghì 1992. Equally relevant: a very truthlike intuition might only deviate from the truth by a measure of 0. In this case, the measure of its truthlikeness, 1, corresponds to the probability that the intuition is correct (Niniluoto 1986).

We have invoked a scenario wherein a rational being, lacking perfect intuition, senses

mathematical facts. An opinion materialises that any violation of these facts would constitute a PPP. Nonetheless, the machine, as his external source, has the final say on all matters. If the sun doesn't rise, it doesn't rise.

Suspending disbelief, let us allow that his calculator can access all relations of the real numbers. One day, he stumbles upon a preprogrammed function that sends any pair of reals to {0, 1}. Denoting the function by I(*x*, *y*), the calculator presents our being a choice of which variable to enter first (in time). Irrespective of his thoroughly random mental selection, every time he enters an *x* before his *y*, the calculator returns an I of 1; and each time he reverses the order, I equals 0. With enough repetitions (and remembering that everything else was a fact), confirmation of the mystery function assumes the level of natural certitude. Either the machine has an index of time, or it gives him the power to control this function, or facts and PPPs can trade places arbitrarily. Under the last alternative, his calculator is the plaything of a Cartesian demon that puts all PPPs in the shade. Phenomena of this type can be dubbed *insensible* (instead of counterfactual), because they violate the premise that the whole system is factual.

That is the short version of what you will experience. To organise our work, it will be useful to transcribe this model to the following theses:

i. A sensible mathematical phenomenon that does not explicitly index time is time-invariant.

ii. A sensible mathematical phenomenon that does not explicitly index a subject does not vary by subject.

iii. No phenomenon of a system that embeds an insensible phenomenon is reliably sensible.

Insensibility results from neither self-reference (like incompleteness) nor logical contradiction. It cannot be isolated or neglected as a special case. On the contrary, the phenomenon fully manifests either as a fact or PPP for each combination of time and subjectivity. Therefore any related parts of the system are infected.

These theses are uniformly weaker than the analogous qualifiers of an absolute truth, that is, eternity (Meyer 2002), objectivity, and consistency. So sensibility is a more general qualifier that speaks directly to the predicate. Classically existing and nonexisting formal truths and falsehoods can all be sensible (as it is commonly assumed they are).

In this study, probability theory is used to test the time-invariance of Freiling's *F* (below). Our test focuses on the combined necessary consequence of two uniform experiments. In one experiment, a value of *x* is generated, and in the other, a value of *y*. If *F* is time-invariant, then the sampling order of *x* and *y* should not affect whether *x* is an element of *F*(*y*). Thesis ii is covered by the extraneous determination of sampling order.

There is no known predicate for which sampling order affects outcome in the way described. On the contrary, time-invariance is a working hypothesis of mathematics. Section 3 contains the proof itself.

2.

If Gödelian Platonism is right, then beyond the perimeter of precedent there exists a true axiom which finally demarks the truth or falsehood of the Continuum Hypothesis. Among the attempts to unearth this axiom, Freiling 1986 is conspicuous in the breadth of its

popularity. Freiling proved that the Continuum Hypothesis is inconsistent with an axiom that can be written in the form:

$$\neg\exists F:\mathbb{R}\to\mathbb{R}_{\aleph_0} \mid \forall x\in\mathbb{R},y\in\mathbb{R}:x\notin F(y)\Rightarrow y\in F(x)$$

henceforth the Axiom of Symmetry.[1] The axiom embeds an alluring and unconventional connection to probability theory. Consider a specific value of y in (0, 1), for example y = .123 … (Champernowne's Constant). Now imagine throwing a dart at the whole open segment. The cardinality of F(.123 …) is $\aleph_0$, but the range of the throw has a cardinality of c. So the probability of hitting some value x_0 such that x_0 belongs to F(.123 …) is exactly equal to zero. For each random throw, $x_0 \notin F$(.123 …) Kolmogorov-'almost surely'. Therefore, under the definition of F, .123 … *is* a member of $F(x_0)$ for all randomly selected x_0 almost surely. However, the cardinality of each $F(x_0)$ is only $\aleph_0$; so what is true for Champernowne's Constant cannot be true for all values of y. It therefore appears reasonable to postulate that F does not exist. Understandably, the Axiom of Symmetry was presented as a *mathematical* argument against the Continuum Hypothesis.

On its own terms, the argument holds with respect to a joint experiment if F has a probability measure on (0, 1) × (0, 1). However, there is no joint uniform distribution which applies to each of the elements in the range of F.[2]

The critique herein is based on a dimensional reduction. Measurable subsets of (0, 1) are atomized by *time*. By making time explicit, the dynamics of the conditionality concept are unfettered, and the distinction between probability and 'mere' measure theory, illuminated.

As a pedigree, the Gödelian Platonist premise that all consistent mathematical predicates are eternal falls under a test of the time-invariance of F.

3.

$\forall\, x \in (0, 1)$, $y \in (0, 1)$, either $x \in F(y)$ or $x \notin F(y)$. Allow I(x, y),

I : (0, 1) × (0, 1) → {0, 1}, to denote the indicator function $1_{x \in F(y)}$.

For each of two sampling orders, A and B, a sequence of two uniform experiments is conducted. In this application, we can define a *uniform experiment* (UE) succinctly by:

$\alpha.\rightarrow$ One singleton, $\{s\} \neq \varnothing$, is selected | $s \in (0, 1)$.

$\beta.\rightarrow$ For each Carathéodory-measurable subset E of (0, 1), $\Pr(s \in E) = \mu(E)$, where $\mu(\cdot)$ denotes the Lebesgue measure.

The probability operator $\Pr(\cdot)$ is defined in the aleatory sense, that is: the chance of an occurrence before the UE is finalised. s will be referred to as the sample.

---

[1]. Both the form of the axiom and its elucidation differ from the original work; this is done for continuity. Note that the first arrow denotes 'maps to,' whilst the second denotes 'implies under the definition of *F*'.

[2]. For example, a uniform distribution on the natural numbers requires that each has the same probability of $p \in [0, 1]$. If p = 0, then the sum over the infinite series of naturals is 0. If p > 0, then the sum is infinite.

Under α, a sequence of two UEs will generate a sample size of two. These samples are generated at times 0 and 2, respectively. Allow s_0 to refer to the first sample generated, and s_2 to the second. In this test, we will position ourselves after the first sample has already been realised and before the second UE is conducted. $\Pr_A(\cdot)$ and $\Pr_B(\cdot)$ denote the probabilities at time 1 of events at time 2 under diverse sampling orders A and B respectively.

Under sampling order A, denote s_0 by y_0 and the unrealised s_2 by x_2. Then

$$\Pr_A(I(x_2, y_0) = 1) = \Pr(x_2 \in F(y) \mid y = y_0).$$

Since the cardinality of the range of F is $\aleph_0$,

$$A\ y_0 : |\{x_2 \mid x_2 \in F(y_0)\}| \leq \aleph_0.$$

Therefore

$$A\ y_0 : \mu(\{x_2 \mid x_2 \in F(y_0)\}) = 0.$$

See, for example, Rao 2004. Moreover, any set of Lebesgue measure 0 is Carathéodory-measurable (ibid.), therefore

(1) $$\Pr_A(I = 1) = 0.$$

Under sampling order B, denote s_0 by x_0 and the unrealised s_2 by y_2; then we know that

$$A\ x_0 : |\{y_2 \mid y_2 \in F(x_0)\}| \leq \aleph_0.$$

Since $\forall\ x, y$, either $x \in F(y)$ or $x \notin F(y)$, this is equivalent to

$$A\ x_0 : \{y_2 \mid y_2 \notin F(x_0)\} = (0, 1) \cap U^c$$

for some set $U \subset (0, 1)$ of maximum cardinality $\aleph_0$. Under the definition of F,

$$y_2 \notin F(x_0) \Rightarrow x_0 \in F(y_2);$$

so

$$A\ x_0 : \{y_2 \mid y_2 \notin F(x_0)\} \subseteq \{y_2 \mid x_0 \in F(y_2)\}.$$

Since $A\ x_0 : \{y_2 \mid y_2 \notin F(x_0)\} = (0, 1) \cap U^c$,

$$A\ x_0 : \{y_2 \mid x_0 \in F(y_2)\} = (0, 1) \cap V^c,$$

where $V \subset (0, 1)$ is also a set of maximum cardinality $\aleph_0$.

$\mu(V) = 0$ and is Carathéodory-measurable (ibid.).

Since $A\ x_0 : \{y_2 \mid x_0 \in F(y_2)\}$ is complementary to a measurable set, it is Carathéodory-measurable by definition, and its measure is equal to 1. Therefore

$$A\ x_0 : \mu(\{y_2 \mid y_2 \in F^{-1}(x_0)\}) = 1,$$

where F^{-1} denotes the image of F.

And since y_2 is conditional on x_0 given,

(2) $$\Pr_B(I = 1) = \Pr(y_2 \in F^{-1}(x) \mid x = x_0) = 1.$$

Comparing equations (1) and (2), it is plain that sampling order determines outcome almost surely. Thus the predicate F is not time-invariant.

The real number system embeds at least one insensible phenomenon. Under thesis iii, the system is not reliably sensible.

References

Descartes, R., 1641. *Meditationes de prima philosophia*. www.thelatinlibrary.com/des.html.

Dürr, D., Goldstein, S., and Zanghì, N., 1992. «Quantum chaos, classical randomness, and Bohmian mechanics». *Journal of Statistical Physics* **68**, 259-270.

Freiling, C., 1986. «Axioms of symmetry: throwing darts at the real number line». *Journal of Symbolic Logic* 51(1), 190-200.

Meyer, U., 2002. «Prior and the Platonist». *Analysis* 62.3, 211-216.

Niniluoto, I., 1986. «Truthlikeness and Bayesian estimation». *Synthese* 67, 321-346.

Rao, M. M., 2004. *Measure theory and integration*. New York: Marcel Dekker; 2nd edition.

Jonathan C. Sampson
Sungkyunkwan University
Seoul, Korea
jonathancsampson@yahoo.com

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 98-116
Nozick, Parfit, and Platonic Glasses

NOZICK, PARFIT, AND PLATONIC GLASSES

by Wesley Cooper

---

1 Schema and Theory

The Closest-Continuer schema is a theory of identity according to which identity through time is a function of appropriate weighted dimensions. A at time 1 and B at time 2 are the same just in case B is the closest continuer of A, according to a metric determined by continuity of the appropriate weighted dimensions. For example, the metric for a functional artifact such as a *ship* arguably is equally weighted between the dimension of physical make-up and the dimension of functional structure. So the Ship of Theseus creates a dilemma. Its physical parts are gradually removed at sea and deposited on shore, but they are replaced by new parts that maintain the same functional structure. If the old parts are reconstituted into the same functional structure, is the ship with the new parts the original Ship of Theseus, or does the ship with the old parts warrant that designation? Continuity of physical make-up favors the old-parts ship, whereas continuity of functional structure favors the new-parts ship. There is no closest continuer, given equal weighting, so neither is the original ship.

There might or might not be motivation to apply the concept of identity in the Ship of Theseus, but Nozick thinks there would be ample motivation to apply it if one were a person liable to branching of equally close continuers. This would not be tantamount to death, he proposes, because one would see them through Platonic glasses as the best realized instantiation of oneself. In the *Closest Realized Instantiation* mode of conceptualization that is applicable to branching, no other instantiated relation of continuity comes closer to the concept of personal identity than branching does. There is no closest continuer. So one should adopt the Closest Instantiated Relation view instead of the Closest Continuer view. The branches would *be* the trunk person, although the concept of identity would not strictly apply. The motivation, which gets a grip on human beings but not ships, is that by wearing Platonic glasses one does not sink into oblivion.

Nozick's Closest-Continuer schema should be understood as integrated with a theory of specifically *personal* identity, his theory of the reflexively self-aware self. The Closest-Continuer schema becomes a theory of personal identity through an account of the self as self-creating: through an act of reflexive self-awareness that «entifies» the self, and of self-defining through the self's management of its dimensions, as in weighting them and adding new dimensions. Nozick holds that identity through time is a process of constructive self-synthesis. The self-synthesis begins early in a human life with the development of the capacity for reflexive self-reference, the sort of reference to a given self that only that self can achieve. (Not «that one» or even «this one» but rather something like «this very one».) There is no preexisting I. Rather the I is synthesized or delineated around an early act of reflexive self-referring, an act without a doer in which a doer first comes to exist, a thought without a

thinker in which a thinker comes to exist. This process is constructive in that dimensions may then be added. Additions or deletions of dimensions, changes of weightings, and changes in the metric are possible throughout life; there is no requirement that they be the same for any two lives. Each of us is a self-definer.

The Closest-Continuer schema is meant to apply to the identity of all continuants, whereas its application to persons requires recognition of their self-defining nature. Buddhist no-self doctrine has force against the soul-pellet view (Descartes' Thinking Substance, etc.) but not against a person as an ongoing unification of diverse dimensions, as theorized by the closest-continuer theory in *Philosophical Explanations*, for instance. It is reflexive self-consciousness that constitutes and organizes this unification. He tells a genealogical story about many «bits of consciousness», such as a memory of an earlier conscious event, among which is a bit which is «very special», because it is an awareness of many other bits of «experience and thought, plus an awareness of itself, a reflexive self-awareness». In this way the self begins. Somehow the step gets made from being aware of these bits of consciousness to having or possessing them. They come to belong to the self. The self is built upon reflexivity and appropriation.

Nozick realizes that the Closest-Continuer schema doesn't focus especially on personal identity but rather on a general framework for identity through time. To remedy this defect he offers a theory of the self as arising and defining itself through acts of reflexive self-awareness. This agency is a *deep fact* about the self, beyond the dimensions that the self selects and weighs. The Nozickean self — he calls it, with some relish, a Fichtetious object — does not reduce to these weighted dimensions, for it is the cause of their being its dimensions and being weighted so. Its being a deep fact does not mean that it shares anything more in common with Descartes's Res Cogitans. For instance, Nozick's self is mutable and indeed protean in its possibilities for self-definition; and its reflexive self-awareness, though essential to it, does not imply that it is non-physical. That selfhood for Nozick is a deep fact will become important later, where it will be argued that Parfit incorrectly assumes that Nozick shares Parfit's Reductionism and the attendant denial that personal identity is a deep fact.

Reflexive Self-Awareness

Nozick tackles the special features of the self, notably its reflexive self-awareness, beginning at the linguistic level in the use of indexicals like 'I' as opposed to proper names or definite descriptions. His account of reflexive self-reference leads to the conclusion that sentences containing 'I', 'me', 'my', or 'mine' (I-statements) are derivable from non-I-statements, specifically in terms of the more general reflexive self-referring phrase 'this very': 'the producer of this very token.' Dismissing as inadequate Kripke-style rigidity or referring to the same thing in all possible worlds, he opts for a feature that a subject of reference has «that is bestowed in that very act of reference». One refers «from the inside» in reflexive self-reference. He proposes that the most adequate linguistic formulation is, 'this very reflexive self-referrer'. Nozick's linguistic approach to reflexive self-awareness is neutral with regard to the contentious issue about whether it is a «subjective fact» as implied by the accounts of Thomas Nagel and John Searle, or whether it is an objective physical fact as implied by the various materialist theories. However, Nozick makes explicit his preference for functionalism in *Invariances* (Nozick, 2001), so the tentative neutrality of his linguistic approach is upgraded into the metaphysical neutrality of functionalism about the mind in his last book.

To be an I or self is to have the capacity for reflexive self-reference. Nozick

hypothesizes (it doesn't follow from the linguistic points he's made about how the term 'I' refers) that selves are essentially selves, «that anything which is a self could not have existed yet been otherwise.» He explores this hypothesis by asking how reflexive self-knowledge is possible, dispensing with the suggestion that this is a special mode of relating to ourselves as objects, or a dispositional account, or a brute-fact account, or an account in which the self places itself into its reflexive self-referrings. He stalks a better answer by proposing that entification, the classifying which produces entities, «takes place in one fell swoop, rather than in the stages of transverse followed by longitudinal». An informative entification brings together a diversity that it unifies, maximizing the sum of the degrees of organic unity over the entities it classifies. With these remarks about classification in hand, and with some trepidation, he speculates that «the I is delineated, is synthesized» around an act of reflexive self-referring. The entity I comes to exist in the act of synthesis. (He asks at this point, «Can the rabbit be pulled out of the rabbit? …. Can the self really be a Fichtetious object?») A current synthesis does not determine the precise character of a later synthesis, but it can affect what happens later as a (non-binding) precedent, and «thereby syntheses at different times can mesh into a larger continuing entity,» a currently synthesized self including past self-stages in accordance with the closest continuer and closest predecessor schema. The idea that reference is to an independently preexisting and bounded entity is an illusion. Nozick's Fichtetious theory explains why selves are essentially selves: they are synthesized by reflexive self-reference «and around it qua something having it». He acknowledges the counter-intuitiveness of speaking of acts without independently existing agents, but he juxtaposes this with our willingness to hold that Descartes can reach only «thinking is going on» and not «I think».

Platonic glasses

Nozick embeds the Closest-Continuer theory in a broader view of how philosophical concepts like personal identity can be structured.

1. **Intrinsic Abstract Structural**. A concept C's holding means that an abstract structural description «involving only monadic predicates» holds. Example: Personal identity is is an intrinsic feature of a person, the soul.

2. **Relational**. X falls under concept C just in case X stands in a certain relationship R to «another, sometimes earlier, thing of a specified sort». Example: Personal identity as analyzed in terms of spatio-temporal continuity, or psychological continuity. (Relational rather than monadic predicates are required by 2.)

3. **Closest Relative**. To the relational view is added the condition that nothing else is as closely related under R to that other thing. Example: The closest continuer theory of personal identity. (Quantification is required by 3.)

4. **Global**. Something satisfies C only if it stands closest in R to a specified y, and also is a part of any wider thing that stands closer in R to y than do other comparably wide things. Example: «Thus, one might hold that any acceptable theory not only must fit the evidence as well as any alternative theory of the same phenomena, but also must be part of any wider theory of more inclusive phenomena that fits the evidence more closely than any other theory alternative to it.» (Quantification over wider entities is required by 4.)

5. **Closest Instantiated Relation R'**. This adds to 1-4 that there is no other instantiated

relation R' which comes closer to the concept C than R does. Example: In cases where there is no closest continuer R but rather a tie R' such that the concept of personal identity does not strictly apply, one should (in the Platonic mode; see above) adopt the Closest Instantiated Relation view 5 instead of the Closest Continuer view 3. (Quantification over relations is required by 5.)

The integrated Closest-Continuer theory does *not* imply that hypothetical cases of cloning or branching, familiar in the literature of personal identity, are tantamount to death when the branches are equally close to the originating or «trunk» person. This is because one can view one's continuation through «Platonic glasses» as the best instantiated realization of one's identity, even if the concept of identity does not strictly apply, in order to keep oneself from sliding into oblivion. This same motivation is at work in the Closest-Continuer theory to bind reflexive self-awareness at different times into a continuant person. In this latter case the concept of personal identity does «strictly apply» according to the Closest-Relative mode of conceptualization, in contrast to cases of Parfitian branching. But in both kinds of case Platonic glasses are at work. In normal human life Platonic glasses justify knitting the momentary or episodic self-aware selves into a continuing person, and in the branching cases the original person's caring about the branches just as he cares about himself, in such a way as to see himself as continued in the branches, is a rational response to an extraordinary circumstance.

2 A deep fact?

The integration of the Closest-Continuer theory with Nozick's theory of selfhood is overlooked when Parfit, for instance, classifies Nozick as a Reductionist and presents both Nozick and himself as Closest-Continuer theorists for normal human lives. He views the two of them as ranged against the traditional theorists about personal identity who regarded personal identity as a «deep fact». Parfit doesn't think so. All that matters in personal identity is present in the survival that occurs in hypothetical cases of branching or cloning, in which the concept of identity does not apply, since the branches are definitely different people and therefore cannot, by the transitivity of identity, be one and the same with the person from whom they branched. If a=b and a=c then b=c, by transitivity. But branch 1 (b) and branch 2 (c) are not the same person, so the original person (a) can't be identical to either.

For Parfit these hypothetical cases reveal the triviality of personal identity. Everything that matters about a person's experience can be fully described without referring to the person at all. He is critical of Nozick's attempt to salvage the concept of personal identity in these hard cases by seeing the forked continuers through «Platonic glasses,» as close-enough real-world instantiations of the original person to *be* (all of them) that person, despite the fact that the concept of personal identity does not strictly apply. Parfit deems this salvaging to be a conflation of theoretical and practical rationality. It may be practically rational for the original person to believe this in order to escape oblivion, since otherwise forking would mean death, but it is not theoretically rational. The truth is, on his view, that the original person no longer exists after forking or branching (and yet everything that matters is there).

Parfit misleads when he classifies himself as a Closest-Continuer theorist like Nozick, except for the hypothetical cases that reveal the triviality of personal identity. Nozick is not a Reductionist in Parfit's sense alluded to above; he does not hold that a person's experience can be fully described without reference to the person at all. This is evident when the Closest-Continuer schema is integrated, as it should be, with the theory of the reflexively self-aware

self. The self *guides* the application of the schema to himself, so to speak, assigning weighted dimensions to himself. Selves are different from other continuants in this respect. A ship for instance is assigned the weighted dimensions appropriate to it by us, those who created and apply the concept *ship*. Although a self operates with a concept of a *person* that has been externally created and has rules of application that forbid a self to self-define as, say, a number, the imposed framework leaves much leeway for partial self-definition in the assignment of dimensions and weights.

Also Parfit's use of the distinction between theoretical and practical rationality is invidious, given the theory of rational belief that Nozick developed in *The Nature of Rationality* after publication of the Closest-Continuer theory in *Philosophical Explanations*, in order to address problems like this. It introduces a practical element, symbolic utility, in its decision-value account of rational belief. The decision-value of believing that the branched persons are oneself may be sufficiently high to be rational, given the symbolic utility of that belief. This is what wearing Platonic glasses in such hard cases would amount to. On this reading, Platonic glasses are also worn in normal cases, in order to forge a continuing personal *entity* between temporally spread-out personal *acts*. One rationally believes oneself to continue between acts of reflexive self-awareness because decision-value validates this belief. That is, the belief is justified by the weight of the belief's symbolic utility to the reflexively self-aware self.

What Parfit dismissively calls the «deep fact» about personal identity is the existence of an irreducible subject of experience, a fact about the 'I' that has experience, a reference to a person that cannot be eliminated in a full description of that experience. This is what Parfit's Reductionism denies. Historically the Deep Self has been understood to be transcendent to experience, as in Kant, or imminent to experience, as in Descartes. It has been observed however that Descartes' Self — the *I* of his *cogito, «I think therefore I am»* — is not a valid inference. From «I think» he is only entitled to infer «Thinking is going on.» On this reading Descartes' Self is transcendent from experience, just as Kant's transcendental unity of apperception.

However, it is arguable that an interestingly stronger inference can be drawn from «I think», namely, « *This thinking* is going on.» After all, the *cogito* can be performed by two or more people at the same time, but it is meant to pick out only one thought. Moreover the demonstrative is meant to pick out not the thought of the second or third person who performs the *cogito*, but *this very thought*. So the interesting replacement for the *cogito* becomes: «*This very thought* is going on.» Let this thought be termed an act of reflexive self-awareness. The thought is aware of itself in the intimate way implied by *this very thought*. Now suppose that this act somehow bootstraps itself into a subject of thought and other experiences. It *entifies* itself, as Nozick says. This subject is imminent to experience in acts of reflexive self-awareness. It is a reflexively self-aware self. This very self, WC, is writing this very sentence, which began with the word tokens «This very self».

Let this Nozickean self as developed so far be termed *the Real Self*. The Real Self may be picked out by a physical description, such as a description of the state of the nervous system that subserves the Real Self and its experiences. But reference to the self is not eliminable in favor of such a description, any more than a sculpture of a man riding on a horse is replaceable by a molecular description of the marble from which it is sculpted. The Closest-Continuer theory, when it applies to a person, is a theory about the identity conditions of a Real Self. It follows that Nozick is not a Parfitian Reductionist about personal identity.

The Real Self is a «deep self», though not of Descartes' or Kant's varieties. There is rather a kinship with William James's seminal theorizing about the self in *The Principles of Psychology*.

It follows furthermore that Parfit cannot adhere to the Closest-Continuer theory for normal cases of personal identity, at least in the sense that Nozick intends. For Nozick intends the Closest-Continuer theory, when it is applied to persons, to be a theory of the identity conditions of the Real Self. More particularly, the Real Self is self-defining. The self-definition begins with the self-as-act entifying itself as a continuant and continues with its ascribing to itself dimensions, such as having a human body, to which it assigns weight. Personal identity cannot be reduced to such-and-such weighted dimensions, not only because the dimensions and their weights are liable to change as the process of self-definition continues, but also because a listing of the weighted dimensions would omit the crucial fact that they are precisely self-ascribed and self-weighted.

3 Platonic glasses again

Parfit notes that Nozick's pattern of concern in cases of branching might be defended by reference to the Platonic mode of caring about something, in which «we see the world in its aspect of realizing what is beyond it, we see and can respond to its glimmerings of something finer which shine through.» As Nozick admits, viewing clones in this way involves «an unrealistic overestimate of actuality, a seeing of it through Platonic glasses», as opposed to the alternative of making «a more realistic assessment of things, seeing things as they are in themselves.» Nozick objects to this alternative that it «makes one a prisoner or a victim of the actual world, limited by the ways in which it falls short, by how it happens to be …». (Nozick 1981, 67)

Parfit rejects this defense. He grants that, given the distinction between theoretical and practical rationality, Nozick's pattern of concern is defensible, but «theoretically irrational».

> If Nozick reacts to reality not as it is, but as he would like it to be, this is theoretically irrational. But if this kind of wishful thinking is more deeply satisfying, it can be practically rational for him to try to make himself, in this way, theoretically irrational. A more extreme case would be that of someone who wants to be deluded by Nozick's 'experience machine'. (Parfit 1984, 479)

Parfit's distinction between theoretical and practical rationality is inadequate to articulate what is rational about believing that one's branching continuants are, each of them, oneself. Nozick's decision-value account of belief in *The Nature of Rationality*, which includes symbolic utility as well as credibility value in a decision-value formula for rational belief, justifies this belief about the branches, viewing them through Platonic glasses. The credibility value of projecting oneself into the branches is sufficiently high, given the great similarity that each branch has to oneself in terms of their weighted dimensions, including processes of self-definition that are continuous with the original person's.

Nozick supports the side-constraint view against classical utilitarianism and the idea that only felt experience matters, by introducing the famous Experience Machine thought experiment. It induces whatever illusory experience one might wish, but it prevents the subject from doing anything or making contact with anything. There is only pre-programmed neural stimulation sufficient for the illusion. Nozick pumps the intuition that each of us has a reason to avoid plugging into the Experience Machine forever. This not to say that «plugging in» might not be the best all-things-considered choice for some who are terminally ill and in great pain, or an amusing pastime for an afternoon. The point of the thought experiment is to

articulate a weighty reason not to plug in permanently, a reason that should not be there if all that matters were felt experience. Nozick's Experience Machine is importantly different from David Deutsch's more attractive conception of a virtual-reality generator, which leaves intact «internal experience» such as belief, desire, and choice, altering only «external experience» such as input to the various sensory modalities. (Deutsch 1997) Nozick's Experience Machine programs both internal and external experiences, so that free will and even simple self-directed thought and action are out of the question.

One's agency ceases with the act of plugging in to the experience machine, and nothing like that agency and the consequent experience goes on from there. Not only is the content of experience determined by the programmer of the machine, but also the process of self-definition is usurped. Having chosen to plug in to the «politician» experience machine, the virtual politician's choices afterwards are simply those that were designed by the programmer. The erstwhile self-definer is slumped in a chair, a blob serving as a vehicle for simulated experiences. Clearly, wearing Platonic glasses is quite dissimilar to plugging in to the Experience Machine.

The Platonic-glasses diagnosis is that one ought to care about each of the equally close clones as one cares for one's single future self in the real (non-branching) world, because the clones would represent the closest real-world instantiation of the concept of personal identity, close enough to be oneself despite the violation of the logic of personal identity. This care might express itself in arranging that one's clones share equally in one's wealth, for example. The imperatival 'ought' here should probably give way to the permissive 'may': one is entitled to think of one's branches as continuations of oneself, other things being equal, to the extent that they enhance the meaning of one's life, prevent one's sinking into oblivion, enable one to transcend the limits of the current self, etc. The decision value of believing in this continuation may be high enough to make it rational, because of the symbolic utility of the enhanced meaning. This permits as consistent with self-interest the choice, when confronted with the prospect of branching, to donate one's wealth to a favorite charity. One weighs one's uniqueness so heavily that one's care can't extend to the branches.

4 The Real Self

The Real Self is episodic, because episodes of reflexive self-awareness are episodic, broken up by sleep and indeed, under the probing of skepticism, broken up by the passage of time from one «now» to another. Entifying these episodes into a continuing person is a step that is justified by viewing the episodes through Platonic glasses. Just as there is symbolic utility in believing that one is continued by branches, rather than one's sinking into oblivion when branching occurs, so too the reflexively self-aware self at time *t* prefers to see itself as continuous with earlier and later reflexively self-aware selves. Believing in the real self as an entified continuant has symbolic utility for the self-at-a-time, enough to make it rational, by the lights of a decision-value account, to believe in the Real Self. We do this as a matter of course, and the DV account theorizes it. Pathological cases are imaginable in which prudence fails to extend beyond a day. This is not because a soul-pellet disappears at the end of the day, but because the subject's utility profile today is not linked to a subject tomorrow, as it is in the normal case by the symbolic utility of seeing oneself today as continuant with a subject tomorrow, and conversely. This «conversely» clause may not be applicable in born-again psychologies, in which there is strong motivation to see oneself as discontinuous with a previous self.

5 A comparison to William James

Nozick's view has affinities in many respects with William James's discussion of the self a century previous, in *The Principles of Psychology*.

> The unity into which the Thought — as I shall for a time proceed to call, with a capital T, the present mental state — binds the individual past facts with each other and with itself, does not exist until the Thought is there. It is as if wild cattle were lassoed by a newly-created settler and then owned for the first time. But the essence of the matter to common-sense is that the past thoughts never were wild cattle, they were always owned. The Thought does not capture them, but as soon as it comes into existence it finds them already its own. How is this possible unless the Thought have a substantial identity with a former owner, — not a mere continuity or a resemblance, as in our account, but a real unity? …. For how would it be if the Thought, the present judging Thought, instead of being in any way substantially or transcendentally identical with the former owner of the past self, merely inherited his 'title,' and thus stood as his legal representative now? It would then, if its birth coincided exactly with the death of another owner, find the past self already its own as soon as it found it at all, and the past self would thus never be wild, but always owned, by a title that never lapsed. We can imagine a long succession of herdsmen coming rapidly into possession of the same cattle by transmission of an original title by bequest. May not the 'title' of a collective self be passed from one Thought to another in some analogous way? (James 1981, 321)

Less reliant on metaphor, Nozick's account of entification, including the decision value refinement of it, is doing the same work as James's «legal title» handed from one herdsman to the next. James also relies on a metaphor of a herdsman branding cattle to talk about the present self's relationship to its past, and explains the metaphor with another, the idea of a perceived «warmth and intimacy» that the self feels.

> Our recent simile of the herd of cattle will help us. It will be remembered that the beasts were brought together into one herd because their owner found on each of them his brand. The 'owner' symbolized here that 'section' of consciousness, or pulse of thought, which we have all along represented as the vehicle of the judgment of identity; and the 'brand' symbolizes the characters of warmth and continuity, by reason of which the judgment is made. (James 1981, 319)

In a similar vein Nozick introduces principles that govern entification. One such principle, which integrates his theory of personal identity with his theory of value, recommends defining yourself so as to increase your value, which he takes to be marked by organic unity. This principle could account for the born-again murderer, who despairs of creating an organic unity with his former self, preferring instead to «start over» in order to track value in his selfhood. The desire to rise above the moment, to avoid oblivion, expresses a related principle: Entify oneself so as to avoid oblivion, ceteris paribus. The person who weighs his uniqueness so heavily that he cannot wear Platonic glasses with reference to his future branches, in the Parfitian scenarios, takes advantage of the ceteris paribus clause. Nozick would not do so, and he presumes that his readers as well would wear Platonic glasses in such extraordinary cases, doing such things as dividing their wealth equally among the branches rather than giving it away.

A principle of rational prudence stems from the organic-unity principle, since a life united by treating each moment of one's life as of equal weight has more value-as-organic-unity than a life lived for the moment. What Rawls calls the Aristotelian principle also derives from tracking value in one's life. There is a tendency for human beings to seek to develop their more complex abilities. A principle emphasized by Nozick is precedent, which guides one towards choices that cohere with one's previous choices, which set parameters which set limits to what one can choose currently, without causing any particular choice. The model for this principle is *stare decisis* in the common law.

6 Parfit's Combined Spectrum

Refusing to wear Platonic glasses, Parfit reports that before he had formulated his Reductionist view he had «seemed imprisoned in myself»; his life seemed like a «glass tunnel».

> When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air. There is still a difference between my life and the lives of other people. But the difference is less. Other people are closer. I am less concerned about the rest of my own life, and more concerned about the lives of others. (Parfit 1984, 281)

Discounting for some rhetoric, this is a fair statement of difference between Parfit, refusing to wear Platonic glasses, whose life goes better by believing trivializing Reductionism, and other people, who believe in significant personal identity through time, and who would believe that they were continued by multiple equally-close branches, were branching to occur.

In a thought-experiment about «the Combined Spectrum», in which there are gradual physical and psychological changes in a person until he becomes more like Greta Garbo than himself, Parfit writes,

> In the cases in the middle of this Spectrum, there would be a resulting person who would be in some ways psychologically continuous with me as I am now. But this resulting person would also have many new and dissimilar cells, and he or she would also be in many ways psychologically continuous with Greta Garbo. At the far end of this Spectrum, this future person would be in no way related to me. If I accept Nozick's view, I care equally about such a future person, provided that he or she is closely enough related to me. I regard all of the cases in the first part of this Spectrum as being just as good as ordinary survival. As we move along this Spectrum, the future person would be less and less closely related to me. But I am equally concerned about this person, provided that the degree of closeness is close *enough*. (Parfit 1984, 237)

Parfit continues,

> On this view, I must decide just what degree of closeness counts as close enough. I am [sic] must again draw a sharp line on this Spectrum. If my relation to some future person is just on the near side of this line, this relation is as good as ordinary survival. If my relation to some future person is just beyond this line, I should be less concerned. But the future person in the second of these cases would differ hardly at all from the person in the first case. The differences would be only that a few more cells would be replaced, and there would be some small psychological change, such as a desire to be alone. Though these are the only differences, I should care less about what happens to this second person. (Parfit 1984, 478)

Parfit concludes his critique.

> This pattern of concern seems to me irrational. How can it have such importance whether just a few more cells would be replaced, or whether there would be one more small psychological change? Nozick's view treats this Spectrum as though it involves, at some point, a discontinuity. But this is false. Since the Spectrum is smooth, involving all of the degrees of continuity, why care equally in all the cases in the first part of the Spectrum, and then suddenly care less? This would be rational only if identity is some further fact which holds completely in the first part of the Spectrum, and then suddenly fails to hold. But Nozick does not believe that there is any such further fact. (Parfit 1984, 478-9)

Although there is no *general* answer to the sorites question raised by the Combined Spectrum, individuals could give their own answers, possibly different from person to person, about where the cutoff point is. In this a person would be exercising his right, as a self-defining Real Self, to determine the metric of «close enough» in judging whether a continuer is close enough to be him. Self-definition is an addition to the closest-continuer theory that Nozick

employs when that theory is applied to persons. Persons are self-defining entities, unlike other continuants through time, such as clubs and chairs. That the cut-off point might seem arbitrary from an external point of view is neither here nor there. The individual has a pressing reason to draw the line, namely the subjective utility of continued existence, which motivates a choice that would otherwise be arbitrary. Drawing the line here can be compared to choosing whether to provide for one's branches or give one's wealth away when the prospect of branching is nigh. How much weight must one attach to uniqueness to give the gloss of rationality to the choice the latter choice? There is no hard line. The subject has to make the decision by his own lights. The sorites problem reduces to an existential one.

7 Decision value and Incline-the-Beam

Emotions and personal relations affect the decision value of a candidate hypothesis for belief. On the basis of the same evidence that is available to others, a mother whose son is on trial for murder arrives at a different belief from them, that her son is innocent. Nozick defends the rationality of the woman's belief on the grounds of his decision-value account, which allows the symbolic utility of believing her son innocent to have weight and, given that the credibility value of the the hypothesis of his innocence is sufficiently high, to lead her rationally to the belief in her son's innocence. He employs the following rule: «Believe a statement *h* if there is no alternative statement incompatible with *h* that has a higher credibility value than *h* does, and the credibility value of *h* is high enough, given the kind of statement that *h* is, and the expected utility of believing *h* is at least as great as the expected utility of having no belief about *h*.» (Nozick 1993, 89) He translates this into his decision-value account, which maximizes decision-value as the weighted sum of causal, evidential,and symbolic utility — as follows: «Believe (an admissible) *h* only if the decision-value of believing *h* is at least as great as the decision-value of having no belief about *h*.» (Nozick 1993, 89) Nozick's rule for rational belief is closely related to a strategy employed by William James, to which I now turn.

Nozick's example fine-tunes the Incline-the-Beam strategy employed by James in his defense of contra-causal free will in the chapter «Attention» from *The Principles of Psychology*. What he calls there the «effect theory» is the view that attention is always the effect of an antecedent cause; it represents a deterministic outlook. He opposes to it the 'cause theory' according to which attention is an 'original force'; it represents contra-causal free will. He argues that the evidence for determinism is not so compelling that it blocks his right to believe in contra-causal free will, a belief that makes better sense of his life than the effect theory. Dismissing Occam's Razor (parsimony) as adequate to reject the 'cause theory', he characterizes the 'original force' deepening and prolonging the stay in consciousness of innumerable ideas which would otherwise fade away more quickly, continuing:

> But the whole feeling of reality, the whole sting and excitement of our voluntary life, depends on our sense that in it things are really being decided from one moment to another, and that it is not the dull rattling off of a chain that was forged innumerable ages ago. This appearance, which makes life and history tingle with such a tragic zest, may not be an illusion. As we grant to the advocate of the mechanical theory that it may be one, so he must grant to us that it may not. And the result is two conceptions of possibility face to face with no facts definitely enough known to stand as arbiter between them.
>
> Under these circumstances, one can leave the question open whilst waiting for light, or one can do what most speculative minds do, that is, look to one's general philosophy to *incline the beam* [emphasis added]. The believers in mechanism do so without hesitation, and they ought not to refuse a similar privilege to the believers in a spiritual force. I count myself among the latter, but

> as my reasons are ethical they are hardly suited for introduction into a psychological work. The last word of psychology here is ignorance, for the 'forces' engaged are certainly too delicate and numerous to be followed in detail. (James 1981, 429)

When a person is continued uniquely by one person in the continuum who is «close enough» by the original person's metric, Incline-the-Beam favors continued personal identity. When there is branching and many continuants are close enough, the original person should wear Platonic glasses and divide his care equally over them. The law might give such continuants a special legal status, as *bersons* who are entitled to equal shares in the original person's property. (Never mind for now the issues that would arise about property that is not divisible.) When such branching includes continuants who are not close enough, the original person's care is not extended to them. This might be reflected for instance in wills that are designed to cope with such eventualities, according to which continuants deficient on some heavily weighted dimension are not to be regarded as bersons. As for those continuants who are on the fringe of the distinction between bersons and non-berson continuants, such cases could be decided by the courts. By definition, the original person would not know what to say about such continuants. So the decision could be made for social reasons in favor of regarding the marginals as bersons, or not. A line could be drawn by the courts, with the degree of arbitrariness that affects judicial discretion, sorting continuants on one side or the other of the classes of *b*erson and non-*b*erson. The legitimate scope of judicial discretion is not unlimited. Where there are only two *b*ersons, for instance, and one has all the weighted dimensions of the person's identity and the other has very few, a judicial decision in favor of the the latter would be a mistake. To avoid such mistakes persons would have an extra reason to create wills that filter *b*ersons from non-persons according to their metric.

8 Branching copies in parallel worlds

Philosophy in Nozick's explanatory mode may hitch its wagon to a scientific star and see where that goes. Nothing like a proof will come of it, but an interesting possibility will have been explored. That is the intent in what follows. Some implications of no-collapse interpretations of quantum physics are explored, especially with reference to the question whether Parfit's or Nozick's account of personal identity is better suited to describe what happens when a person branches into parallel worlds.

Bersons exist in one world, the world in which they branch from a person, whereas many-worlds copies of a person exist in parallel worlds. They are similar however in that the case for wearing Platonic glasses with respect to *bersons* also applies to a person's copies in other worlds. Many-worlds dispenses with the «collapse of the wave function» in the orthodox interpretation, in which the various probabilities for a quantum superposition are unrealized except for the one that survives the collapse: there is just one future *you*, for instance. On many-worlds however there is no collapse. the various probabilities are actually realized. There is many-worlds branching, creating many copies of you in other worlds, in proportion to the prior-to-branching probability that your future would take that direction. One proponent of the «no-collapse» theory of quantum physics (as David Lewis dubbed it(Lewis 2004)) is quantum physicist David Deutsch, who seems to favor Nozick's «Platonic glasses» approach to branching. There are «multiple identical copies» of me in the multiverse. Which one am I? Deutsch answers, «I am, of course, all of them.»

The no-collapse theorist Lewis assumes too that one continues through one's branches. That is why he finds the prospect of quantum immortality so terrifying.

> Eternal life on such terms amounts to a life of eternal torment. It is not to be welcomed but feared. You should fervently hope that a collapse will cut it short. You who bid good riddance to collapse laws, you quantum cosmologists, you enthusiasts of quantum computing, should shake in your shoes. Everett's idea is elegant, but heaven forfend it should be true! Sad to say, a reason to wish it false is no reason to believe it false.
>
> So, how many lives has Schrödinger's cat? If there are no collapses, life everlasting. But soon, life not at all worth living. That, and not the risk of sudden death, is the real reason to pity Schrödinger's kitty. (Lewis 2004, 21)

The Parfitian answer to the question about the branches would be, «The concept of personal identity doesn't apply.» This answer seems to be favored by several no-collapse theorists. (Wallace 2002, 22) But this creates an inconsistency, because they continue to refer to the branches as the same person, and they take issues like quantum immortality seriously, as Lewis does, when on Parfit's view there is no cause for concern, since the «immortal» won't be you.

Lev Vaidman writes,

> «I» am an object, such as Earth, cat, etc. «I» is defined at a particular time by a complete (classical) description of the state of my body and of my brain. «I» and «Lev» do not name the same things (even though my name is Lev). At the present moment there are many different «Lev»s in different worlds (not more than one in each world), but it is meaningless to say that now there is another «I». I have a particular, well defined past: I correspond to a particular «Lev» in 2002, but I do not have a well defined future: I correspond to a multitude of «Lev»s in 2010. In the framework of the MWI it is meaningless to ask: Which Lev in 2010 will I be? I will correspond to them all. Every time I perform a quantum experiment (with several possible results) it only seems to me that I obtain a single definite result. Indeed, Lev who obtains this particular result thinks this way. However, this Lev cannot be identified as the only Lev after the experiment. Lev before the experiment corresponds to all «Lev»s obtaining all possible results. Although this approach to the concept of personal identity seems somewhat unusual, it is plausible in the light of the critique of personal identity by Parfit 1986. Parfit considers some artificial situations in which a person splits into several copies, and argues that there is no good answer to the question: Which copy is me? He concludes that personal identity is not what matters when I divide. (Vaidman 2002)

Note that Vaidman in this passage is committed to both

1.There is more than one person Lev after a given quantum experiment,

and

2.There is no good answer to the question, «How many Levs are there after a given quantum experiment».

This seems flatly contradictory. Vaidman can have 1 on Nozick's «Platonic-glasses» account, or 2 on Parfit's reductionist account. This passage is instructive because it reflects common practice among MWI theorists, of paying lip service to Parfit's position, presumably because they accept his argument that the concept of identity does not apply in cases of branching, while speaking as though the concept of personal identity continued to apply when talking about the branched copies of oneself.

Copies need not be strictly identical in the sense of the identity of indiscernibles relativized to universes: All of my copies see a coin spinning in a coin toss, but an instant later half my copies see 'heads' come up, the other half see 'tails'. A distinction between copies, versions, and variants is at work here. Variants of me need not see the spinning coin. Versions of me see it though some of them see 'heads' and some 'tails'. Multiple copies of

me all see the spinning coin. So if I toss heads and believe that I could have flipped tails, some version of me actually tosses the coin and sees tails. Variants of me don't flip the coin at all, underwriting the subjunctive, 'I could have refrained from tossing the coin'. The generic sense of 'copy' that includes copies in the narrow sense but also versions and variants may be singled out by capitalizing: my me-Copies comprise me-copies, me-versions, and me-variants. Not included are 'me-outliers' such as the person who, instead of flipping the coin or refraining from doing so, shoots everyone in sight. That is not something I could have done. The transworld structure that is Me — the set of me-Copies — excludes the me—outliers.

Among philosophers Lewis is notable for associating no-collapse theory with the prospect of unwelcome immortality, writing «You who bid good riddance to collapse laws should shake in your shoes. Everett's idea is elegant, but heaven forfend it should be true! Sad to say, a reason to wish that it is false is no reason to believe that it is false.» (Lewis 2004, 21) He argues that in in life-and-death cases (like Schroedinger's Cat) we need to adjust the intensity rule, which departs from orthodox collapse interpretations by giving up *chances* of possible outcomes of a collapse in favor of the *intensities* of the possible outcomes. Chances correspond to the pre-collapse squared amplitudes of the branches, only one of which will become real. No-collapse theorists on the other hand think of their non-collapsing branches as having differing 'intensities', again corresponding to the squared amplitudes of those branches.

The adjustment in the intensity rule that Lewis proposes is to apportion all our expectation to the branches where we survive. When we have life-and-death branching, first discard all the death branches, because there are no minds and no experiences associated with death branches. Only then divide expectations of experience between the remaining branches in proportion to their intensities. Since all causes of death are probabilistic, there will always be a branch in which you survive. Inevitably however you will end up a lonely geriatric wreck. Note that Lewis wears Platonic glasses in his description of the modified intensity rule. You are on all those branches that link your current self to the future wreck. That is why quantum immortality is a bad prospect for you.

David Papineau is notable for his confidence about the no-collapse theory. He sees no need to adjust the intensity rule. We can simply proportion expectations about the future directly to the intensities. He is serene about giving up chances in favor of intensities, though he prefers to think of chances as being theoretically reduced to intensities rather than eliminated and replaced by intensities. He doesn't share Lewis's concern that the intensity rule is unjustified as well as relatively new, observing that the chance rule can't be justified either, as shown by the history of attempts to justify induction since Hume.

Lewis thinks there is another reason why the lack of justification counts against the intensity rule more than the chance rule, namely its adjusting concern about branches according to their intensities, instead of treating them all equally: «All your future selves, on all your branches, are equally real, and equally yours. You will have experiences of all of them. Do they not deserve equal weight regardless of their intensities?» (Lewis 2004, 15))

Papineau replies,

> Still, does orthodoxy make uncertain choices any less discriminatory? According to orthodox metaphysics, in any chancy situation I will have a number of possible successors. Yet these successors do not weigh equally in orthodox choices either, since the chance rule analogously advises me to choose those actions that benefit my high-chance possible successors over those that

> benefit my low-chance possible successors. (Papineau 2004, 159)

And although both collapse- and no-collapse theories track probabilities of branchings, only the no-collapse theorist cares about all the branches, in Papineau's view, whereas the collapse-theorist cares only about the branch that becomes actual. If we really want what benefits one's post-collapse actual successor, isn't it odd that the collapse-theorist would adopt a principle that counsels maximizing benefit over all possible outcomes? As Papineau writes,

> Since the chance rule recommends that we perform actions with one feature (probable success) when we really desire another feature (actual success), it seems as if there ought to be some non-question-begging way of connecting the chance rule's recommendation with what we really desire. But there isn't. (Papineau 2004, 160)

By contrast the no-collapse theorist has no problem on this score, since one will be succeeded by all one's possible successors, weighted by their intensities. Among one's successors is the miserable immortal, but this prospect is vanishingly small on either no-collapse or orthodoxy. There are far greater expectations of a normal life. Papineau defends this parity by rejecting Lewis's adjustment to the intensity rule, so that no-collapse can follow orthodoxy in assigning normal expectations to futures in which one is dead, in proportion to their intensities whether or not those futures contain one's live successors. He favors formulating the intensity rule in terms of expectations of what will happen whether experienced or not, rather than in terms of expectations of experience. So he does not follow Lewis in pruning dead branches, so to speak. He puts the point like this.

> Theoretical confirmation is one purpose for which we need expectations. And for this purpose expectations of experience are all we need. But we also need expectations to guide our rational choices, and here an intensity rule formulated solely in terms of expectations of experience will lead us astray. In particular, such an intensity rule will stop us attaching expectations to branches in which we will have no experience, and so will fail to persuade us to avoid dangers of death. (Papineau 2004, 166)

Papineau concludes that no-collapse theory is as «healthy as could be» with regard to empirical confirmation, and that experimental evidence leaves «no room for anything except highly ad hoc alternatives.» In all circumstances simple enough for the interference effects that demonstrate non-collapse, like the two-slit experiment, those effects are detected. Furthermore the ad hoc accounts that posit collapse, under conditions too complex to test for interference, are «strikingly less elegant than Everett» and also require rejection of special relativity and the conservation of energy. (Papineau 2004, 168)

Lev Vaidman recommends what he calls the Behavior Principle, «We care about all our successive worlds in proportion to their measures of existence,» which avoids Lewis's problem: «I should not agree to play quantum Russian roulette because the measure of existence of worlds with Lev dead will be much larger than the measure of existence of the worlds with rich Lev alive.» (Vaidman 2002) However, he accepts at face value, unlike Papineau, the collapse theorist's care about possible future selves according to the probability of their occurrence. He doesn't see an advantage for MWI on this score, only equal care differently explained. Like Papineau though he is bullish about MWI.

> The reason for adopting the MWI is that it avoids the collapse of the quantum wave. (Other non-collapse theories are not better than MWI for various reasons, e.g., nonlocality of Bohmian mechanics; and the disadvantage of all of them is that they have some additional structure.) The collapse postulate is a physical law that differs from all known physics in two aspects: it is genuinely random and it involves some kind of action at a distance. According to the collapse postulate the outcome of a quantum experiment is not determined by the initial conditions of the

> Universe prior to the experiment: only the probabilities are governed by the initial state. Moreover, Bell 1964 has shown that there cannot be a compatible local-variables theory that will make deterministic predictions. There is no experimental evidence in favor of collapse and against the MWI. We need not assume that Nature plays dice. The MWI is a deterministic theory for a physical Universe and it explains why a world appears to be indeterministic for human observers. (Vaidman 2002)

9 The Distributed Self, and Free Will

When human beings become agents through reflexive self-awareness, they express their agency by having reasons for acting, to which they assign weights. Choosing the dimensions of one's identity is a special case, in which the assigning of weight to a dimension is partly self-constitutive. But all acting for reasons is constitutive of the self in a broader sense, namely, by its shaping one's character and personality in a manner analogous to the shaping that law undergoes through the precedent set by earlier court decisions. Just as a judge does not merely apply the law but to some degree makes it through judicial discretion, so too a person does not merely discover weights but assigns them; one not only weighs reasons but also weights them.

Set in train is a process of building a «framework» for future decisions that we are «tentatively committed to». The life-long process of self-definition in this broader sense is construed indeterministically by Nozick. The weighting is «up to us» in the sense that it is undetermined by antecedent causal factors, even though subsequent action is fully caused by the reasons one has accepted. He compares assigning weights in this deterministic sense to «the currently orthodox interpretation of quantum mechanics»,(Nozick 1981) following von Neumann in understanding a quantum mechanical system as in a superposition or probability mixture of states, which changes continuously in accordance with quantum mechanical equations of motion and discontinuously via measurement or observation that «collapses the wave packet» from a superposition to a particular state. Analogously, a person before decision has reasons without fixed weights: he is in a superposition of weights. The process of decision reduces the superposition to a particular state that causes action.

This picture might better be understood as about unpredictability rather than indeterminism, since the currently much-discussed many-worlds interpretation of quantum mechanics is deterministic. (At any rate it's probably an overstatement to regard von Neumann's wave-packet-collapse view as the current orthodoxy.) Then the unpredictability might be due to the fact that the background field of more or less inchoate reasons that the agent fixes is insufficient by itself, without that fixing, to determine action. But the process as a whole would be deterministic. The telling analogy, as suggested above, would not be to the indeterminacy of quantum mechanics but rather to judicial discretion: A judge does not in «hard cases» discover the law but rather makes it, according to Hart's theory of judicial discretion; similarly the agent in character-forming decisions does not discover his reasons but rather fixes their weight. Neither judicial nor personal agency need be construed indeterministically. (Bratman 2002) On the other hand, the judicial model at the psychological level and quantum indeterminacy at the lower level of the brain's operations could combine to provide a contra-deterministic theory of free will that isn't liable to the objection that it renders free will random. This was Nozick's intent.

Many of us think that we could have done otherwise, just as circumstances were when we chose to act. This doesn't consort well with another belief that we are likely to have, that determinism (at some level, such as the neurophysiological) is true. One way of relieving the

tension between the two beliefs is offered by the Everett or many-worlds interpretation of quantum physics, introduced in the previous section. The self is distributed over many worlds, in virtue of which the subjunctive 'I could have done otherwise' is true. It's true because I actually do otherwise in another world. David Lewis's modal realism makes a similar claim and is viewed by Deutsch as akin to the many-worlds hypothesis. However, Papineau puts this comparison in perspective.

> If you have heard of Everett because of his association with the 'many-worlds' interpretation of quantum mechanics popularised by Bryce Dewitt, you might suppose that there is some affinity between the no-collapse interpretation and Lewis's philosophical realism about possible worlds. But this would be a mistake. While Everett's interpretation does add extra 'branches' to the reality recognized by common sense, these additions fall far short of Lewis's multiplication of worlds. For a start, the extra 'branches' that Everett adds to reality all lie within the actual world that evolves from the actual initial conditions in line with the actual laws of physics — these branches by no means include all possibilities. Moreover, Everett's branches are best conceived, not as sunderings of the whole universe, but rather as entities that spread out causally at finite speeds, 'like ripples on a pond', as Lewis puts it. For example, in the Schrodinger's Cat experiment, *first* the photon branches into a deflected and undeflected version when it passes through the half-silvered mirror; then the detector branches into a triggered and untriggered state when it interacts with the photon; then the poison bottle branches into a smashed bottle and an unsmashed bottle under the influence of the detector; and so on, culminating in the cat branching into a live and dead cat, and the human observer branching into a self who sees a live cat and a self who sees a dead cat …. It is precisely this causal proliferation of branches that makes the no-collapse interpretation so theoretically attractive. (Papineau 2004, 153)

Return to Parfit's Combined Spectrum of changes in me. Insofar as the changes are physically possible, they are present in the multiverse. In addition to copies of me in worlds close or similar to this one, there are versions of me in remote worlds. In addition to the me-copy that writes this sentence, there is the me-version that is a thinking beetle (if packing my psychology into a beetle body is a physical possibility). I belong to the Distributed Self that includes the various Real Selves in the sundry worlds as copies or instantiations. The Distributed Self is a trans-world set of Real Selves, whereas a Real Self is a particular in a particular world. As for which actual changes in me in *this* world are consistent with my continuation, that is for me (my Real Self) to decide. Different Real Selves in my Distributed Self may decide differently. After all, I am free.

Nozick favors a version of James's «cause» theory of attention, which he calls originative free will, as opposed to the compatibilist free will favored by determinists (which James labeled «soft determinism»). Nozick speculates that the brain is a quantum computer, in effect, enabling originatively free choices according to a «precedent» model of the relationship between current and past choices, as described above. Bratman has shown how the appeal of this model remains on a compatibilist or «soft determinist» view. What doesn't survive in Bratman's transfer is the intuition that many people have, that they could have done otherwise when they chose such-and-such, just as circumstances were. This seems to require something like Nozick's reference to quantum phenomena. John Searle for instance, in order to save the intuition, has recently converted to a defense of free will as underlain by quantum phenomena in the brain.

However, there is a another way of saving the intuition which is compatible with a deterministic universe, and it is arguably required in general, in order to justify counterfactual inferences about oneself and others. This is the Everett or many-worlds or multiverse interpretation of quantum physics, which justifies the subjunctive 'I could have done otherwise' by reference to a parallel world in which I actually do otherwise. The 'I' in

question is a kind or set rather than a particular, and the particulars of that kind are called «copies» of the person. If this is correct, then Reductionism is doubly repudiated. Not only is there a Deep Self, namely the reflexively self-aware self and its construct, the Real Self, but also there is a Distributed Self, the self as distributed over many worlds.

The Deep Self rescues after a fashion the intuition about free will, namely, that, on everyday occasions of free action, just as circumstances were, one could have chosen otherwise. For one did do otherwise in other universes; members of your copy consort did do what you-copy did not. This is different from the rescue anticipated by the idea that one's brain is a quantum machine that creates moments of indeterminism that enable free choice. On the many-worlds view the brain may be a classical computer processing information according to deterministic laws that hold complete sway in the multiverse (the many worlds as a whole). But one is still free; one could have done otherwise on many an occasion; for one did do so, thanks to one's copies in other universes. One's Distributed Self was free, or one's Real Self was free qua member of a Distributed Self.

The Distributed Self is somewhat appealing as a solution to the problem of rescuing the free-will intuition in a deterministic universe. Those who accept compatibilism or «soft determinism» don't take seriously the intuition, so they won't recognize a problem. They will «analyze» *I could have done otherwise* in a compatibilist formula like *I would have done otherwise if I had so chosen*, or perhaps abandon analysis as doomed to failure and simply reject the intuition. Those who see a problem will reject the formula, insisting that it leaves out one's freedom, just as circumstances were when one chose as one did, to choose otherwise. A many-worlds metaphysics offers a solution. The many worlds as a whole make up a system deterministically governed by natural law, but in any given universe a person is probabilistically related to his or her copies in other worlds. Some of them do otherwise, supporting the intuition that one could have done otherwise.

In addition to appealing to the free-will intuition, the multiverse metaphysics also appeals to what might be called the distributed-care intuition: when you are deliberating over whether to do *a* or *b*, you care about your future whether you decide one way or another. But on a single-world metaphysics the course not taken is a mere unrealized possibility. It would counsel therefore that one's care should be focused exclusively on one's actual self, the possibility that is realized by the collapse of the wave function. If this is wrong, as it is on Papineau's account summarized above, it is because the action one chose has a reality greater than mere «unrealized possibility». The Everett interpretation articulates this reality: one cares about a copy of you that performs that action, the action you passed up. Otherwise your pattern of care would be irrational, as would your belief that you categorically could have done otherwise on occasions of putatively free action.

Given a multiverse or many-worlds outlook, basic ideas about the universe are either vindicated or undermined by the multiverse hypothesis. For instance, counterfactual conditionals refer to nearby parallel worlds when they stipulate what a thing would do under conditions that do not actually obtain; one-worlders implicitly collapse what things can do into what they actually do. Consider a coin toss. The identical worlds in which I (copies of me) see it spinning become branched; in fifty-percent of those worlds versions of me see 'heads', and in fifty percent they see 'tails'. This actual distribution of worlds is what licenses the inference, about this world, that if the coin hadn't turned up 'heads' it would have turned up 'tails'. Instead of its being a basic fact that my observing 'heads' collapses probabilities into an actual outcome of the coin toss, those probabilities are grounded in actual universes in

which both outcomes are represented.

Knowledge is a trans-universe structure, as one might expect because knowledge supports counterfactual implications, as revealed for instance in Nozick's tracking account of knowledge. Nearby parallel worlds are united by a common history of knowledge acquisition, which may be spelled out in broadly Popperian, conjecture-and-refutation terms. The resulting epistemological niche lends stability and reliability to knowledge in each universe. Moral knowledge is a part of this niche. Life is a similar trans-universe structure, molded by natural selection rather than rational criticism. What distinguishes genuine replicating DNA from junk DNA is that the former but not the latter is representative of a niche of replicators that extends across worlds. Indeed personal identity is inseparable from such a niche, which Deutsch picks out with the word «copies». A person is a set of copies in nearby parallel worlds. This comes out in his analysis of free will: 'I could have chosen otherwise' is analyzed as 'Other copies of me chose otherwise'. And in the denouement to a dramatic chapter that rehearses interference experiments from a multiverse viewpoint, he writes of his copies, «Many of those Davids are at this moment writing these very words. Some are putting it better. Others have gone for a cup of tea.» (Deutsch 1997)

Not only are persons spread out through worlds, but they, like everything else, are quantized through time in any given world. Time is a series of moments, and a person who exists at a moment exists there forever in four-dimensional space-time, rather than being transformed continuously through the flow of time. Such change and flow are mythical, Deutsch argues. The argument doesn't strictly require the multiverse hypothesis, but rather space-time physics since Newton has implied that the openness of the future is an illusion, and consequently causation and free will are illusions. What the multiverse adds is a reduced account of common sense's ideas of causation and free will. Although an effect can't be changed by its cause, the counterfactuals that causal statements support are true. If the cause hadn't occurred, the effect would not have occurred. For the multiverse, which is «to a first approximation» a very large number of co-existing and slightly interacting space-times, includes universes in which the cause doesn't occur and its effect doesn't occur. And although the «me-copy» in this space-time could not have done otherwise, there are me-copies in other worlds that actually do otherwise. There is a branching of these me-copies that validates my sense that my future is open, in contrast to space-time physics. However, the open future of common sense is a myth. There is no flow of time dividing the actualities of the past from the unactualized potentialities of the future.

Since «other times are just special cases of other universes», the temporal granularity of personhood through time is a special case of being spread out through worlds. In addition to one's identically time-stamped copies at a moment across parallel worlds transversely, there are the differently time-stamped copies across parallel worlds longitudinally, linked by natural law so as to give the individual's experience of one world and a continuous self. The implications for the theory of personal identity are not yet clear, but I have argued that endorsing Parfit's Reductionist view would be premature. Your 'copies' in parallel worlds, like branching bersons in this world, are close enough to you to be you, even though the concept of personal identity does not strictly apply.

———————————————

References

Bratman, Michael (2002), 'Nozick on Free Will', in *Robert Nozick*, ed. by David Schmidtz, Cambridge U.P., 2002.

Deutsch, David (1997), *The Fabric of Reality*, Penguin, New York.

James, William (1981), *The Principles of Psychology*, Harvard University Press, Cambridge, MA.

Lewis, David (2004), 'How Many Lives has Schroedinger's Cat?', *Australasian Journal of Philosophy* **82**.

Nozick, Robert (1981), *Philosophical Explanations*, Harvard University Press, Cambridge, MA.

Nozick, Robert (1993), *The Nature of Rationality*, Princeton University Press, Princeton.

Nozick, Robert (2001), *Invariances: The Structure of the Objective World*, Cambridge, MA: Belknap Press.

Papineau, David (2004), 'David Lewis and Schroedinger's Cat', *Australasian Journal of Philosophy* **82**.

Parfit, Derek (1984), *Reasons and Persons*, Oxford University Press, Oxford.

Vaidman, Lev (2002), 'Many-worlds Interpretation of Quantum Mechanics', *The Stanford Encyclopedia of Philosophy* **http://plato.stanford.edu/archives/sum2002/entries/qm-manyworlds/**.

Wallace, D. (2002), 'Worlds in the Everett Interpretation', *Studies in the History and Philosophy of Modern Physics* **33**, pp. 637-661.

Wesley Cooper
University of Alberta
123drc@gmail.com

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 117-124
Kant and the Expression of Imperatives

KANT AND THE EXPRESSION OF IMPERATIVES

by Ronald Cordero

---

1. Introduction

Not far into the Second Section of the *Foundations of the Metaphysics of Morals*, Kant makes an interesting claim about the expression of imperatives. In the original German it is 'Alle Imperativen werden durch ein *Sollen* ausgedrückt …'[1], and in the popular Lewis White Beck translation it is rendered as «All imperatives are expressed by an 'ought' …»[2]. The assertion in the English version is extremely perplexing. Are not most imperatives expressed without an «ought»? In fact, are *any* imperatives *ever* expressed with «ought»?

From the point of view of moral theory, the question is important. Kant is going to claim that the moral law, which he believes to be both universal and necessary, is created by the categorical imperative that comes from reason. Can this crucially important imperative be expressed by an «ought» judgment? And if it can, how is it that such a judgment can give rise to apodictic law?

Moreover, such a claim about imperatives and «oughts» has significant consequences for the analysis of ethical utterances. To say that imperatives are expressed by «oughts» is to say that at least some «ought» judgments are expressions of imperatives. If this is in fact what Kant has in mind, then he would seem to be saying that using «ought» can be a way of issuing a command; and this has a distinctly noncognitivistic ring to it. On this interpretation, Kant would appear to be in sympathy with A. J. Ayer's famous position on «ought»:

> The sentence «It is your duty to tell the truth» may be regarded both as the expression of a certain sort of ethical feeling about truthfulness and as the expression of the command «Tell the truth.» The sentence «You ought to tell the truth» also involves the command «Tell the truth,» but here the tone of the command is less emphatic.[3]

On the same interpretation, Kant would also seem to be assenting, at least in part, to Hare's insistence that «ought» in moral contexts is used with a prescriptive rather than a descriptive force.[4] Are we then to suppose that Kant is proposing a noncognitivist or

---

1. Immanuel Kant, *Grundlegung zur Metaphysik der Sitten*, ed. Karl Vorländer, (Hamburg: Felix Meiner, 1952) 33. (Royal Prussian Academy edition page 413).

2. Immanuel Kant, *Foundations of the Metaphysics of Morals*, trans. Lewis White Beck (New York: Macmillan, 1990) 29.

3. Alfred Jules Ayer, *Language, Truth and Logic* (New York: Dover, 1952) 108.

4. R. M. Hare, *Moral Thinking* (Oxford: Clarendon, 1981). See, for example, 215.

prescriptivist analysis of at least certain «ought» judgments? Does he believe that an utterance like «One ought not to lie,» may be nothing more than the issuance of an imperative against lying — and thus not capable of being either true or false?

What is needed is a careful analysis first of just what is involved in the expression of imperatives and the creation of laws — and then of Kant's text. This is the only way to dispel the appearance of misstatement on his part, settle the question of his possible endorsement of a noncognitive interpretation of «ought» judgments, and clarify his claim that morality consists of law created by an imperative.

2. Expressing Imperatives and Making Rules

There are at least two important senses in which one can speak of an imperative being expressed. In one of these, to express an imperative is simply to issue it, to give it as a command; in the other, to express an imperative is to state or report it in indirect discourse. Expressing a certain kind of imperative in the first sense can create a rule. Expressing the same kind of imperative in the second sense can state or report a rule.

When an imperative is expressed in the first sense — when it is issued — the utterance expressing it is noncognitive, having only what J. L. Austin would call the illocutionary force of commanding or directing.[5] Suppose we wish to direct (to order or command) Smith to be here by five. We could, of course, issue the imperative in any of thousands of languages. In English, we could issue it by saying to Smith any of the following …

1) Be here by five.

2) You will be here by five.

3) You are to be here by five.

The first version simply uses the imperative mood. The second employs a future tense in a way that stipulates rather than predicts. And the third involves a construction that can plausibly be regarded as equivalent to «You are hereby directed to be here by five.» Each of these three utterances can be meant and taken as an issuance of the imperative, and if they are so taken, none of them can be regarded as cognitive. None could be either true or false. In using the imperative mood one is not making a statement. Nor is one doing so by stipulating required future behavior as in (2) or by explicitly directing it as in (3). If «You will be here by five,» were understood as a prediction, it would constitute a statement, not a directive. And if «You are to be here by five,» were taken to mean «You have been ordered to be here by five,» it would constitute a statement about a directive given in the past and not an issuance of the directive in the present. But taken respectively as a stipulation and a directive, the latter two utterances are just as noncognitive as the first.

It is important to understand at this point that imperatives expressed in the sense of being issued may or may not be standing imperatives. They may or may not, that is, concern more than a single action on a single occasion. Consider these examples:

1) Always be home by eleven.

2) Students will not run in the halls.

---

5. *How to Do Things With Words* (New York: Oxford University Press, 1962) 99-103.

3) Travelers are to carry identification at all times.

Parents who issue the imperative in (1) to their child are giving an order that will remain in effect until it is rescinded. The directive given by some school authority in (2) is a standing order about conduct by any student at any time in any hall in the school. And the order issued in (3) is meant to apply to an unlimited number of travelers on an unlimited number of occasions. All three examples represent the issuance of imperatives that are, as it were, universally quantified. And this, presumably, is precisely the kind of imperative which Kant takes the categorical imperative issued by reason to be. Reason says to us, as it were, «Always act in such a way that ….» The command reason issues is not just about a single occasion. It is a standing imperative, applying to all times and all places.

The second important sense in which one can speak of imperatives being expressed involves the giving of cognitive reports. Here to express an imperative is not to issue it, but rather to report it. Suppose the media want to report the directive we issued when we said to Smith, «Be here by five.» They could do so in a variety of ways:

1) They said, «Smith, be here by five.»

2) They said for Smith to be here by five.

3) They said Smith was to be here by five.

4) They said Smith should be here by five.

5) They directed Smith to be here by five.

Here the media would not be issuing or reissuing our imperative: they would simply be reporting what we did, and their statement could be accurate or inaccurate, true or false. The first version reports the order in what is called «direct discourse» (using a direct quote). The next three examples use «indirect discourse» to report the same order without a direct quote. Any of the five utterances would count as a statement, a cognitive account of what transpired that could be true or false. And any of the five could also be said to express our imperative by reporting it.

While examples (1) through (5) involve statements about a past event, it should be noted that present-tense statements can report commands just as readily. Suppose we are in the process of parking in a foreign country when a police officer approaches and says something in a language we do not understand. We may ask someone to translate and the translator may say, «The officer says you are to park somewhere else,» expressing the officer's directive in the present tense in a way that reports it cognitively but does not, as it were, reissue it.

As in the case of the issuance sense of the expression of imperatives, it is essential to note that imperatives expressed in the sense of being reported may also be standing, rather than «one-time-only,» imperatives:

1) The king says we are to pay taxes annually.

2) The king said, «All subjects will pay taxes annually.»

In asserting either (1) or (2) we would be reporting a standing command originating from the monarch and could certainly be said to be expressing — though not issuing — a standing imperative. We could, of course, be mistaken in either case. We might, for example, simply

have misunderstood the king's command.

Here it is important to note that reporting a *standing* command through a statement is tantamount to stating a rule — since rules are commonly established through the issuance of standing commands. To report a standing command, that is, can be to state a rule. If we direct our daughter to be home tonight by eleven, we are issuing a «one-time-only» command, not laying down a rule. But if we instruct her always to be home by eleven — if we issue the standing command «Always be home by eleven,» we *are* laying down a rule. The standing imperative establishes the rule. And a statement reporting our standing command would be a statement of the rule. Our daughter might, for example, express the rule in indirect discourse by saying, «My parents say I always have to be home by eleven.» Similarly, if the authorities direct all travelers to carry identification, their standing imperative gives rise to a rule, which can be stated by reporting the directive: «The authorities say all travelers are always to carry identification.» The cognitive expression of the standing order is a statement of the rule established by that order.

Similarly, on Kant's theory, the categorical imperative delivered by the reason of each rational being establishes the apodictic moral law. And that law can be stated through a cognitive report of the categorical imperative. The noncognitive standing categorical imperative, «Act always in such a way that …,» establishes the basic moral law, and that law can be stated through a cognitive report of the categorical imperative: «Reason directs us always to act in such a way that ….»

3. What Kant Says

Turning now to Kant's text, we must deal first with his contention that all imperatives are expressed by a «sollen» (an «ought» in Beck's translation). Should we not reject this claim outright as simply mistaken? After all, we know that imperatives, German and English alike, can be — and indeed characteristically are — expressed in the imperative mood. There is absolutely no need for a «sollen» or an «ought» to be used in the expression of any imperative. And this is the case whether we are talking about imperatives that are hypothetical, such as «If you want to have a good reputation, keep your promises,» or imperatives that are categorical, like «Keep your promises.» The imperative mood provides a perfectly good way of expressing imperatives, with no need for «sollen» or «ought.» As a matter of fact, in the *Foundations* Kant himself *repeatedly* expresses the fundamental moral imperative simply through the imperative mood, without «sollen»:

> Act only according to that maxim by which you can at the same time will that it should become a universal law.[6]
>
> Act so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only.[7]
>
> [T]he categorical imperative can be expressed also as follows: Act on those maxims which can at the same time have themselves as universal laws of nature as their object.[8]

---

6. Beck, 38. (RPA 421).

7. Beck, 46. (RPA 429).

8. Beck, 54. (RPA 437).

So why would Kant say that all imperatives are expressed by a «sollen»? He clearly cannot mean that they are *always* so expressed.

The most reasonable assumption to make here may be that what Kant means is simply that all imperatives are *capable* of being expressed with a «sollen» — not that they are always expressed in that way. Such an interpretation saves the passage from being obviously false, though it also leaves unanswered the question of what Kant means when he talks about imperatives being expressed in such a manner. Exactly how does he conceive this expressing? Is he thinking of imperatives being expressed in the sense of being issued or of imperatives being expressed in the sense of being reported?

As far as expression through issuance is concerned, there is, in fact, a good basis for Kant's claim with regard to «sollen.» It is a fact that in German a form of the verb «sollen» is commonly used for issuing commands.[9] And it is important to notice that this is done in a way that contrasts with a common way in which commands are issued in certain other European languages. Consider for example the following different versions of the commandment about adultery found in *Exodus 20*:

English: Thou shalt not commit adultery.

Spanish: No cometerás adulterio.

French: Tu ne commettras point d'adultère.

German: Du sollst nicht ehebrechen.

There is an obvious difference here between the way the commandment is issued in German and the way it is issued in the other three languages. In the English, Spanish, and French versions of the commandment, a future-tense form is used — as indeed was the case in the original Hebrew.[10] In German, on the other hand, «sollen» is employed and there is no future tense: the «sollst» in the German is a present-tense indicative form of the verb.

As already noted, the use of the future tense to issue commands is not uncommon in English. If we want to tell someone not to park on the grass, we can say, «You will not park on the grass.» And it is significant to observe that in such a case we could usually *not* issue the order unambiguously with «ought.» If we said, «You ought not to park on the grass,» we would run the risk of not being understood to be giving an order at all. We might be taken to be making a recommendation, offering advice, or a giving a suggestion instead.

German, however, does not use the future-tense for issuing commands.[11] When it does not simply use the imperative mood, it frequently uses «sollen» in the present tense; and Kant, with his well known pietist background, must have been familiar with moral commandments issued in this way. Indeed, in the *Preface* to the *Foundations* he cites the commandment

---

9. See, for example, George O. Curme, *A Grammar of the German Language* (New York: Frederick Ungar, 1964) 321.

10. For the point about Hebrew, I am indebted to my colleague Marshall Missner.

11. See Bill Dodd, Christine Eckhard-Black, John Klapper, and Ruth Whittle, *Modern German Grammar: A Practical Guide* (New York: Routledge, 1996) 339-41.

against lying as «Du sollst nicht lügen,»[12] and Beck translates it using the future tense as «Thou shalt not lie,»[13] not «Thou ought not to lie.» Again, in the *Second Section* of the *Foundations*, Kant formulates the commandment against false promises as «[D]u sollst nichts betrüglich versprechen …»[14] and Beck translates «Thou shalt not make a false promise …,»[15] not «Thou ought not to make a false promise.»

So we might well expect Kant sometimes to think of the fundamental moral imperative being issued in such a way, using «sollen» rather than the imperative. In fact, he often does so, though he uses the first person rather than the second person. This is perfectly understandable, of course, since Kant is thinking of the categorical imperative being issued to each individual by her or his own reason. Reason is telling us what to do. In the *First Section* of the *Foundations*, for example, Kant gives a preliminary formulation of the categorical imperative using «sollen» in this way:

> … d.i. ich soll niemals anders verfahren als so, *daß ich auch wollen könne, meine Maxime solle ein allgemeines Gesetz werden.*[16]
>
> That is, I ought never to act in such a way that I could not also will that my maxim should be a universal law.[17]

Here Beck translates «ich soll niemals» as «I ought never.» But by analogy to the preceding examples, the phrase could also be translated as «I shall never act in such a way that …»

Evidently, if Kant is thinking of expressing commands in the sense of issuing them when he speaks of imperatives being expressed by a «sollen,» it is far from certain that «ought» is the corresponding term to use in English. It would, in fact, be misleading to say that Kant holds that imperatives can be issued by an «ought.» It would be truer to his meaning to say that on Kant's position, imperatives can be formulated in terms of what one is to do or shall (stipulatively, not predictively) do. If, for example, I express (issue) the categorical imperative in German with «sollen,» the corresponding expression in English might best be realized with «shall» for the stipulative future tense or with «are to»: «Rational beings shall (are to) act only in ways in which they can consistently will everyone to act.» Saying, «Rational beings ought to act only in ways which they can consistently will everyone to act,» could mistakenly give the impression that the utterance was intended as counsel or advice, rather than as a command.

And what of the expression of imperatives in the sense of reporting? Could Kant have had this in mind as well? The relevant syntactic fact here is that «sollen» is indeed commonly used in German to express commands in the sense of reporting them. It is, one might say, the

---

12. Vorländer edition, 5.

13. Op. cit., 5. (RPA 389).

14. Vorländer edition, 40.

15. Beck, 36. (RPA 419).

16. Vorländer edition, 20.

17. Beck, 18 (RPA 402).

verb of choice for reporting commands in indirect discourse. And in the passage originally referenced, Kant could definitely have had indirect discourse in mind. Indeed, the assumption that that is at least part of what he has in mind certainly helps make sense of the passage.

In German «sollen» is used in the subjunctive, not the indicative, mood for reporting commands in indirect discourse. In formal English the verb form used for the same purpose is «should»:[18] If last night the director said to us «Get those reports in by Friday,» we would report the imperative in indirect discourse in formal English by saying, «The director said we should get those reports in by Friday.» This would be a cognitive report of a command. In less formal English, of course, we have other options when it comes to expressing imperatives in indirect discourse. We might, for example, prefer a construction with the infinitive: «The director said for us to get those reports in by Friday.»

Further, because a standing command establishes a rule, cognitive reports which express standing commands through indirect discourse can count as statements of rules. To report the rule about getting home at a certain time, our children could say, «Mom and Dad say we should always be home by eleven,» or «Mom and Dad say for us always to be home by eleven.» So since Kant conceives of the basic moral law as something created by a standing command given by reason, he might be expected to express (state) it with «sollen» in the subjunctive in indirect discourse. And so he does — in what is referred to as the second formulation of the categorical imperative in the *Foundations*:

> Denn vernünftige Wesen stehen alle unter dem Gesetz, daß jedes derselben sich selbst und alle anderen *niemals bloß als Mittel*, sondern jederzeit *zugleich als Zweck an sich selbst* behandeln solle.[19]
>
> For all rational beings stand under the law that each of them should treat himself and all others never merely as means, but in every case at the same time as an end in himself.[20]

Here the «solle» at the end of the German passage is the subjunctive form of «sollen» in question — and Beck renders it with a «should» rather than an «ought.» One could also say: «under the law which says for each of them to treat himself and all others ….» On Kant's position, the basic moral law, established by the categorical imperative issued by reason, can be reported in terms of what reason says for us to do.

But what about «ought»? Can it be used to express imperatives in the sense of reporting them? And can statements about what ought to be done express standing imperatives in such a way as to constitute statements of rules or laws?

First consider non-standing imperatives. Can we, for example, report the directive «Get those reports in by Friday,» by saying «The director says we ought to get those reports in by Friday,»? Here again — as in the case of issuing imperatives — to do so would be to run a significant risk of being misunderstood. Someone might think that the director only recommended — and did not order — that we get the reports in by Friday. Formal English calls for a «should,» not an «ought,» for the expression of commands in indirect discourse; but even «should» may not avoid ambiguity in these cases. If we say, «The director says we

---

18. T. Herbert Etzler and Harvey I. Dunkle, *A German Review Grammar* (New York: Odyssey, 1965) 34.

19. Vorländer edition, 57.

20. Beck, 50. (RPA 433).

should get those reports in by Friday,» someone might think the director merely advised us — rather than told us — to get them in by Friday. This may explain why everyday English tends to prefer «The director says for us to get them in,» to «The director says we should get them in.»

As much can be said for the expression of standing imperatives. If the authorities say that travelers are always to carry identification, it could be misleading to say the authorities say travellers ought always to carry identification. The authorities are not recommending something: they are requiring it. And if what the law says is «All citizens shall pay taxes annually,» it would be misleading to describe the law as saying all citizens ought to pay taxes. What the law says is that they have to do so.

For Kant, the categorical imperative issued by reason does much more than merely issue a recommendation or give advice. Because of this, it would be highly misleading to report it in a way that could be understood to present it as merely recommending or advising. But to assert that the categorical imperative says we ought to act only in certain ways would be to do precisely this — and so would be to misrepresent the imperative completely. According to Kant, the categorical imperative establishes a universal and necessary moral law. And such a law cannot be adequately or unambiguously stated by talking about what ought to be done.

4. Conclusion

When Kant speaks of the expression of imperatives, whether he has in mind the issuance of imperatives by means of a noncognitive utterance or the reporting of imperatives by means of a cognitive utterance, he is *not* saying that imperatives can be expressed by «ought» judgments. He does not think that reason phrases its categorical imperative in terms of what ought to be done. Nor is he saying that the basic law of morality can be stated in terms of what ought to be done. The suspicion raised earlier about the possibility of Kantian noncognitivism regarding «ought» judgments in the Second Section of the *Foundations* can thus be seen to be unwarranted.

Ronald Cordero
University of Wisconsin at Oshkosh
cordero@uwosh.edu

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 125-134
Knowing that *p* rather than *q*

KNOWING THAT *P* RATHER THAN *Q*

by Bjørn Jespersen

Introduction

In a number of papers Jonathan Schaffer has outlined a programme of an epistemological position called *contrastivism* and ably defended it. The idea informing contrastivism is that there is no such thing as knowing something *simpliciter*. In the final analysis, knowledge is always knowledge standing out against a foil of eliminated relevant alternatives. Schaffer's phrase for knowing that something is true while its relevant alternatives are not is 'knowing that *p* rather than *q*'. In Schaffer's words, «The 'rather than'-clause is a mechanism of contrastivity. It explicitly articulates *q*.» (2005, p.11.) For instance, it is wrong to say that you know that it is snowing, but correct to say that you know that it is snowing rather than raining, in case the proposition that it is raining is one of the eliminated epistemic possibilities.

In this paper I wish to investigate the cornerstone of contrastivism, the notion of knowing that *p* rather than *q*. I shall not discuss, for instance, whether knowledge always presupposes the elimination of epistemic possibilities. Instead I shall address the following two questions:

- What is the notion of *contrast* that is central to knowing that *p* rather than *q*?
- What sort of object is the *object of knowledge* in the case of knowing that *p* rather than *q*?

Neither question has so far received sufficient attention so as to provide an exhaustive answer. Or put more pointedly, just as Schaffer objects that, «No contextualist has ever offered anything near a precise account of relevance» (2004, p.88), so it may be objected that no contrastivist has ever offered anything near a precise account of knowing that *p* rather than *q*. Below I broach (without advocating it) an account of knowing that *p* rather than *q* in order to get a firmer grip on the notion. The alternative I broach construes knowing that *p* rather than *q* as knowing the conjunction of *p* and the negation of *q*. As may have already transpired, I have misgivings about this alternative. My primary reservation about it is that it is debatable whether a mere conjunction succeeds in accommodating the *contrast* between *p* and its alternatives that contrastivism revolves around. Whatever its merits and flaws, however, the conjunction account is incompatible with a key constraint that Schaffer imposes on knowing that *p* rather than *q* and, therefore, cannot help advance Schaffer's contrastivist programme. (This is not to say, though, that the account might not turn out to suit an alternative epistemological theory based around the notion of knowing that *p* rather than *q*.)

First, I set out the bare bones of contrastivism, in particular its so-called ternicity constraint. Then I look into the notion of contrastivity, arguing why I find it problematic to

reconcile with the ternicity constraint. Finally, I argue that neither Schaffer's propositions nor my conjunction $p \wedge \neg q$ is appropriate as an object of knowledge due to hyperintensionality concerns.

1. Ternicity

Two strands come together in contrastivism. One is the contextualist tenet that the truth-conditions of sentences containing reference to knowledge, «… know …», are context-dependent because, according to contextualism, 'know' is an indexical term. The other is the notion of relevant alternative. Schaffer rejects that 'know' is an indexical and intends instead to obtain context-sensitivity by means of *q*: any of a range of relevant alternatives to *p* may be assigned to *q*. Relevant alternatives, in turn, provide the backdrop against which *p* stands out as a known truth: *p* is *true* independently of *q*, but cannot be *known* without *q*.[1]

Contrastivism brings the two strands together by construing knowledge as a ternary relation between a knower *s*, a proposition *p* that is known by *s*, and a so-called contrast proposition *q*, which is a relevant alternative to *p* that has been eliminated. Where there are multiple relevant alternatives, the contrast propositions are the conjunctions $q_1 \wedge \ldots \wedge q_n$. The contrastivist conception of knowledge is summed up in the ternicity constraint (2004, p. 77).[2]

Ternicity: 'Know' denotes a three-place relation *Kspq*.

The contrastivist semantics of 'know' gives rise to the following schema of the truth-condition of instances of «… know …».

Kspq iff

(i) *p* is true

(ii) *s* has conclusive evidence that *p* rather than *q*

(iii) *s* is certain that *p* rather than *q* on the basis of (ii).

Apart from the slippery notions of conclusive evidence and certainty, the truth-condition as it stands lends itself to at least two very diverse interpretations known from epistemological contextualism (see DeRose, 1992, pp. 918ff; Brendel and Jäger, 2004, §3). According to subject contextualism, *s*'s context, or vantage point, determines the truth-condition of *Ksp*, such that *s* must know that *p* is true, *s* must have conclusive evidence for the truth of *p*, and *s* must be certain of the truth of *p* in virtue of that evidence. According to attributer contextualism, the attributer's context, or vantage point, determines the truth-condition, since it is the attributer who selects the range of alternatives to *p*. (Of course, the roles of subject and attributer may coincide in the same individual.) In both cases it falls to *s* to eliminate the relevant alternatives to *p* in order to get to know that *p* relative to the given range of alternatives.

Schaffer's contrastivism qualifies as attributer contrastivism, which is in turn a species

---

[1]. For a recent objection to the contextualist semantics of 'to know', see Douven (2004). Roughly, Douven's argument is this. If 'to know' is an indexical term, then it has context-sensitive comparatives and superlatives, as expressed by «Ann knows more than Berthold» and «Ann knows most of everybody». But it hasn't, so it isn't.

[2]. Schaffer adds that the *K* relation ought to be expanded to include a fourth variable *t* ranging over instants of time (2004, p.95, n.8), though leaving *t* out for convenience.

of epistemological externalism. By selecting a particular range of contrast propositions $q_1 \wedge \ldots \wedge q_n$ the attributer imposes stricter or laxer constraints on what qualifies as conclusive evidence for *p* rather than its contrastive contenders. Due to the flexibility as for the choice of relevant alternatives to *p*, it depends on the choice of values to 'saturate' *q* (2004, p. 90) which particular truth-condition a given instance of the '*Kspq*' schema is.

The '*Kspq*' schema is just that — a schema. If we are going to turn the formula into a sentence, ∀-binding the *q* variables would seem the obvious choice, since knowing that *p* rather than *q* means having eliminated all of the *q*'s. The variable '*s*' may be substituted by the constant '*a*'. The variable '*p*' should be replaced by a constant '*P*', since we obviously wish to express, not that *a* knows just any old proposition *p* rather than its alternatives, but that *a* knows some particular proposition *P* rather than its alternatives. When the schema is transformed into a closed expression, the result is

*Ternicity** : $\forall q$ (*KaPq*).

We can use Ternicity* to express that *a* knows that *P* rather than *all* of its relevant alternatives (where '$\forall q$' is shorthand for '$q_1 \wedge \ldots \wedge q_n$'). We also need universal quantification over *q* when 'keeping score of the overall progress of inquiry' across contexts (2004, p. 84): all the relevant alternatives to *P* have been eliminated, all the relevant alternatives to *P'* have been eliminated; etc. We would need '$\forall q$ (…*q*…)', '$\forall q$' (…*q*'…)', etc., to express this.

2. Contrastivity

The preposition 'rather than' is, according to the *Oxford Advanced Learner's Dictionary*, ambiguous between (a) 'in preference to' and (b) 'instead of'. An example of (a) would be, «Ann prefers tea to coffee», with the implication that with no tea around Ann would be fine with coffee. An example of (b) would be, «Ann drinks tea instead of coffee», which I find equivalent to (though emphatically not synonymous with) the conjunction, «Ann drinks tea, and Ann does not drink coffee». This conjunction can be telescoped into, «Ann drinks tea and not coffee». Whether 'instead of' or 'and not', the idea is *one at the expense of the other*.

It is plain that the phrasing of contrastivism in English prose employs (b) only. Without an explanatory caption, however, the phrase 'to know that *p* rather than *q*' might just as well mean that *s* knows that *p* is true, but does not know that, or whether, *q* is true; of *p*, *q* it is *p* that *s* knows. What needs to be made explicit is that *q* has been *eliminated* instead of being simply *neglected*. This might suggest that a rendering in natural language of this particular contrastivist tenet is not readily available. (Perhaps «*s* knows that *p as opposed to q*» might be an alternative formulation.) Thus, I take it to be a *stipulation* on Schaffer's behalf that the 'rather than'-clause is a mechanism of contrastivity in the exacting sense of involving eliminated relevant alternatives. In the remainder of this paper I adopt the contrastivist notion of contrast as expressed by 'rather than'.[3]

[3]. Schaffer argues that there is a linguistic analogy between 'to prefer' and 'to know', in that both verbs are, in the final analysis, ternary (2005, §3). According to Schaffer's analogy, just as the sentence «Ann prefers tea» is elliptical for «Ann prefers tea to *F*» (one substitution instance of which being, «Ann prefers tea to coffee»), the sentence «Ann knows that *p*» is elliptical for «Ann knows that *p* rather than *q*». Obviously, if Ann prefers tea to coffee and coffee to root beer then, by the transitivity of the preference relation, Ann prefers tea to root beer. There is a ranking with tea at the top, coffee in the middle and root beer at the bottom. But Schaffer's analogy between 'to prefer' and 'to know' is easily overstretched. If Ann knows that *p* rather than *q* we cannot introduce an *r* such that Ann knows

A first step toward implementing the contrast between *p*, *q* is to require that the *p*, *q* that occur in *Kspq* be *mutually exclusive*, or incompatible (cf. Schaffer, 2005, fn. 5; 2005, p. 13). That is, '*RA*' abbreviating 'relevant alternatives':

$p \rightarrow \neg q$, for any $q \in RA(p)$

or equivalently

$q \rightarrow \neg p$, for any $q \in RA(p)$.

However, I can think of three reasons why this way of cashing out the contrastivist's ternicity constraint would be problematic. First, said requirement remains exterior to '*Kspq*', showing it to be a shallow logical formalisation. Second, the effect is that since *s*, *p*, *q* are all in the scope of *K*, the knower gets saddled with the knowledge that *p*, *q* are incompatible. This seems to me to exceed the boundaries of contrastivism, which, as I understand the position, does not require that the knower know about this relation between *p* and any of its relevant alternatives. If the contrastivist wishes to maintain that the propositions are incompatible without *s* knowing, the contrastivist needs to show how while preserving the basic ternary form '*Kspq*'. (Some form of the distinction between knowledge *de dicto* and knowledge *de re* might perhaps come in handy.) Third, and most importantly, as I understand the ternary *Kspq*, *K* needs to be a relation between a knower and two propositions that are, at the very least, ordered in a sequence. The first element of the sequence would be the selected proposition, the second element the eliminated proposition. What *s* knows is that *p* rather than *q*, not that *q* rather than *p*. Arranging *p*, *q* as an ordered pair makes it explicit that the order in which they occur matters. The result is

$K\langle s, \langle p, q\rangle\rangle$.

But the pairing of *p*, *q* makes *K* binary, since the arguments of *K* are now an individual and an ordered two-tuple. This is in flagrant breach of Ternicity. Since *p*, *q* need to be ordered as $\langle p, q\rangle$, it would appear that the outcome of the quest for contrast between *p*, *q* inherent in knowing that *p* rather than *q* is a binary *K* rather than a ternary.

I am not sure how to square Ternicity with the required contrast between selected and eliminated propositions. Instead, I wish to suggest an alternative conception of contrastive knowledge. Consider the above example, «Ann drinks tea instead of coffee», which was unwrapped as, «Ann drinks tea, and Ann does not drink coffee». It will receive this formalisation in propositional logic.

$p \wedge \neg q$.

If *s* knows that Ann drinks tea rather than coffee, a first stab at an analysis would be

$Ks\ (p \wedge \neg q)$.

That is, *s* knows that the conjunction of *p* and the negation of *q* is true.[4] To arrive at this

---

that *q* rather than *r*, and therefore knows that *p* rather than *r*. Strained grammar aside, there is no such thing as knowing that *p* rather than *q* rather than *r*, with *p* at the top, *q* in the middle and *r* at the bottom.

4. Schaffer claims (personal communication) that an analysis in terms of $Ks\ (p \wedge \neg q)$ results in what Keith DeRose calls an 'abominable conjunction' in his (1995, pp. 27ff). In general, the conjuncts of an 'abominable conjunction' are that some skeptical hypothesis is not known to be false and that some run-of-the-mill proposition is known. The following analysis translates the assumptions (1) and (2) into (1') and (2').

attribution of knowledge to *s*, the attributer would have to carry out an instance of the following inferential schema.

[1] $p \rightarrow \neg q$ Assumption

[2] p Assumption

[3] $\neg q$ 1, 2, MPP

[4] $p \wedge \neg q$ 2, 3, $\wedge$I

.

.

.

[*m*] *Ks* $(p \wedge \neg q)$…, *Ks*I

*Ks*I is an introduction rule for *Ks* validating the prefixing of '*Ks*' to line [4]. The details of the introduction rule need not detain us here, but (ii) and (iii) from Schaffer's statement of the truth-condition of *Kspq* would have to be in the mix. My counterpart of Ternicity* then is

$$\forall q \in RA(P)\ (Ks\ (P \wedge \neg q)).$$

There are various ways of reading this formalism; for instance, «The conjunction of *P* and the negation of all of its relevant alternatives $q_1 \wedge \ldots \wedge q_n$ is known by *s*», or «For any relevant alternative *q* to *P*, *s* knows that *P* is true and no *q* is». However, my concern is not so much the most accurate interpretation of the formalism as what it would mean to know that the conjunction of *p* and $\neg q$ is true (for arbitrary *p*, *q*). For instance, does it mean to know of either of the conjuncts that it is true? If so, the following distribution qualifies as valid.

$$Ks\ (p \wedge \neg q) \models (Ksp \wedge Ks\neg q).$$

Knowing one thing, namely that the conjunction $p \wedge \neg q$ is true, implies knowing two things,

---

1. Moore knows that he has hands rather than stumps
2. Moore does not know that he has hands rather than vat images of hands

1'.	$Km(H \wedge \neg S)$	Assumption
2'.	$\neg Km(H \wedge \neg V)$	Assumption
3.	$KmH \wedge Km\neg S$	1', Distribution
4.	$\neg KmH \vee \neg Km\neg V$	2', Distribution, DeMorgan
5.	KmH	3, $\wedge$E
6.	$\neg Km\neg V$	4, 5, Disjunctive syllogism
7.	$KmH \wedge \neg Km\neg V$	5, 6, $\wedge$I.

The premises (1'), (2') are contrastive intuitions to the effect that (1) and (2) are true. The conclusion is the 'abominable conclusion' that Moore does not know that he is not a (bodiless, hence handless) brain in a vat, but does know that he has hands. The rationale behind the transition from (2') to (4) would be that if you fail to know a given conjunction to be true it is because you fail to know at least one of the conjuncts to be true. However, the argument is in my opinion *invalid*, by deploying Distribution of *K* over $\wedge$ twice to obtain two pairs of conjuncts each of which with a *K* at the head. *K* Distribution is, in my view, a clearly invalid rule in any properly restrictive epistemic logic. So no such 'abominable conclusion' is validly inferable from the proposal presented above. (I am indebted to the Editor for discussion of this 'abominable conjunction'.)

namely that p is true and that $\neg q$ is true.[5] However, it is problematic that the consequent severs the relation between p and $\neg$q (or q) that is supposed to provide the contrast between the selected proposition and its rejected alternative(s). «$Ksp \wedge Ks\neg q$» just says that *s* knows two things, with no indication of their being related.

Does the antecedent also sever the relation between p and $\neg q$ (or q)? Why, no; $\wedge$ is, after all, a (binary) relation, its arguments being p and $\neg q$ in the present case. But it is to no avail to explain the notion of knowing that p rather than q in terms of knowing $p \wedge \neg q$ as long as it is not clear what it means to know a conjunction to be true or whether knowing that p rather than q can be adequately translated into knowing the conjunction of p and $\neg q$ in the first place. In fact, what especially speaks against the conjunction interpretation of contrastivity is that the mere conjunction of two propositions lacks the biff of contrast between its conjuncts. The conjunction, if true, just records the fact that p is true and q is false.[6]

Another contentious issue is the fact that elimination translates into negation. This translation probably captures the core idea of elimination, but records only the outcome, or exterior, of an act of elimination. The procedural, or interior, aspect of carrying out acts of elimination drops out of the story. This may well be unsatisfactory for a full statement of contrastivism, since processes of eliminating *p*'s rivals conceptually precede *s*'s knowledge of *p*. A full statement of contrastivism would then maintain that *s* sknows that *p* rather than *q*, *because s* (or *s*'s evidence) has already eliminated *q*.

For these reasons I am hesitant to put '$\forall q \in RA(P)$ (Ks ($P \wedge \neg q$))' forward as an analysis of the notion of knowing something rather than something else.[7]

It should be obvious why Schaffer could not possibly accept the conjunction analysis. Ks ($p \wedge \neg q$) violates Ternicity by being binary, and his q is my $\neg q$. What is more, should the

---

5. Evnine embraces this outcome, arguing that «being in a state of believing a conjunction simply is being in a state of believing its conjuncts. There is no state of believing *p and q*, distinct from the state of believing *p* and believing *q*.» (1999, p.215. I replaced '*A*', '*B*' by '*p*', '*q*'.) Unfortunately, Evnine does not reveal how to individuate such 'states' (which I suppose to be the same for believing and knowing), so it remains unclear how strong or weak his claim is. Even so I would have to be convinced that this identity claim is true: Ks ($p \wedge q$) = $Ksp \wedge Ksq$. I am not fond of $\wedge$ flitting in and out of the scope of *K*. For one thing, it is inferable from $Ksp \wedge Ksq$, via $\wedge$E, that Ksp (or Ksq, if you wish), but not inferable from Ks ($p \wedge q$) without first factoring out the two conjuncts and placing them under *K* scope. And it is not logically necessary that knowers be capable of extracting individual conjuncts from a conjunction they know. (Indeed, in pragmatic terms it would be downright irrational for resource-bounded knowers to do so with each and every conjunction known to them.) So it is doubtful whether Ks ($p \wedge q$) is even equivalent to $Ksp \wedge Ksq$.

6. This problem will be familiar from the translation of 'but' into '$\wedge$' when used to conjoin propositions that contrast. For instance, if «Ann is poor but honest» (a telescoped phrasing of «Ann is poor, but Ann is honest») goes into «$Fa \wedge Ga$», the contrast between Ann being poor and Ann being honest conveyed by 'but' is lost in translation. A consequence of the contrast is that «Ann is poor» and «Ann is honest» do not commute in «… but …». See also (Blakemore, 2000).

7. Consider the following as an alternative interpretation which also articulates *q*: *s* knows *p* rather than *q* iff *p* and (*s* knows whether *p* or *q*). The idea underlying *knowing whether p* is to know which disjunct (possibly both) of $p \vee q$ is true. In rough notation (x ranging over propositions):

s knows whether *p* iff $Ks\ \iota x\ (x \wedge ((x = p \vee x = q) \vee x = p \wedge q))$.

ι is a function from sets of propositions to propositions that returns the only member of a singleton, and is otherwise undefined. Its value is the respective unique proposition that is known when it is known whether *p* or whether *q* or whether $p \wedge q$. This alternative is not sufficiently strong, however, since when *p* rather than *q* is known, *q* cannot be true. Thus, we need to add that not both disjuncts are true. Still, as with the alternative in terms of Ks ($p \wedge \neg q$), I fail to see how the *K* of $p \wedge$ (Ks (whether *p* or *q*)) would meet the Ternicity constraint.

implication $Ks\ (p \wedge \neg q) \models (Ksp \wedge Ks\neg q)$ be deemed true, the right-hand conjunct $Ks\neg q$ of the consequent is too strong. It says that *s* actually *knows* something (namely that a certain negation is true) whereas Schaffer carefully settles for the less exacting notions of the subject having 'conclusive evidence' for and being 'certain' about *p* rather than *q* when stating his truth-condition for *Kspq*. This is a natural move, since knowing that $\neg q$ would have to be spelt out as knowing that $\neg q$ rather than some *r*. And $\neg r$ would in turn have to be known rather than something else. And so on, with an infinite regress looming on the horizon that would turn contrastivism into an unmanageable theory of human knowledge.

3. Objects of knowledge

Even if a contrastivist should be prepared to flout Ternicity, there is an additional reason to be wary of

$$\text{«}\forall q \in RA(p)\ (Ks\ (p \wedge \neg q))\text{»}$$

(for arbitrary *p*) as a workable analysis of knowledge *á la* contrastivism.

The reason is that, whether or not this analysis should be found satisfactory, the following problem needs to be addressed: is $(p \wedge \neg q)$ appropriate as an object of knowledge? I think not. I am going to argue that neither Schaffer's propositions nor my conjunction $(p \wedge \neg q)$ is appropriate as an object of knowledge.

First of all, what are *p*, *q*? Schaffer introduces *p*, *q* as propositional variables, construing propositions in the vein of possible-world semantics, to wit, as sets of possible worlds, or functions from the logical space of possible worlds to truth-values. The problem with this sort of propositions is that they are arguably both 'too little' and 'too much'. It is a thrice told tale that, and why, they are too little. They fail to accommodate a principle of individuation finer than logical equivalence. Cresswell in (1975) coins the phrase 'hyperintensional' to characterize any two intensions that are logically equivalent yet distinct; Bealer speaks of a 'conception 2 intension' as what is *intended* when entertaining an attitude (1982, p. 166). For instance, one thing is to know that Seoul is south of Pyongyang; quite another that Pyongyang is north of Seoul. Yet possible-world propositions cannot distinguish between inverse relations. Nor can they distinguish between any two necessarily true propositions, or between any two necessarily false ('impossible') propositions, since there is only one necessary proposition (the one true at all worlds) and only one impossible proposition (the one false at all worlds). Also, due to the classical definitions of the truth-functions, $Ks\ (p \rightarrow q)$, for example, would come out identical to $Ks\ (\neg p \vee q)$. Yet, intuitively, *s* may be innocent of the equivalence between $p \rightarrow q$ and $\neg p \vee q$, and so might know the former without knowing the latter. Schaffer notes that it is problematic using possible-world propositions as objects of knowledge, but proposes no remedy (2007, n. 28). The moral, I submit, is that it is not rewarding to construct either philosophical epistemology or epistemic logic around such propositions, because they are too crude to figure as objects of knowledge. The set-theoretic intensions of possible-worlds semantics are intensionality on the cheap.[8]

---

[8]. It is interesting to note that Fred Dretske — contrastivist *avant la lettre* — implicitly presupposes hyperintensional objects of knowledge (cf. 1970, pp. 1022-23). He says, «We have subtle ways of changing […] contrasts and, hence, changing what a person is said to know *without changing the sentence that we use to express what he knows*.» If the neutral sentence is «Lefty killed Otto», the different things a person can be said to know can be spelt out by means of what is in effect topic/focus articulation: «It was Lefty who killed Otto», «It was Otto whom Lefty killed», and «What Lefty did to Otto was kill him». All three sentences are associated with 'the fact that Lefty killed

But there is also a sense in which they are too much by being too fine-grained. Though extensionally individuated, they are still intensional entities and as such inappropriate as arguments for the truth-functions. For instance, in

$p \rightarrow q$

the truth-function $\rightarrow$ does not operate on intensions. Instead it operates on the *extensions* of *p*, *q*, trading two truth-values for a third truth-value in accordance with its truth-table. In the truth-table the material implication $p \rightarrow q$ is nothing other than a truth-value, which can figure as argument for other truth-functions, as in $\neg(p \rightarrow q) \vee r$.

On the other hand, in

$p \Rightarrow q$

the arguments of the intensional relation of entailment, $\Rightarrow$, are propositions, or truth-values-in-intension, i.e., *p*, *q* themselves. (Entailment takes two propositions, or a set of propositions and a proposition, and yields a truth-value.) Yet standard notation fails manifestly to flag the difference between intensions and their extensions.[9] This sin of omission catches up with us as soon as we prefix '*Ks*' to, e.g., '$p \wedge \neg q$' to form

«*Ks* $(p \wedge \neg q)$».

For if $p \wedge \neg q$ is a truth-value then *s* knows that a certain truth-value is true. But truth-values cannot be known to be true, as little as a kettle of fish can. The standard escape is semantic contextualism: in such-and-such contexts '*p*', '*q*', «$p \wedge q$», etc., denote truth-values, while in such-and-such other contexts they denote propositions. Yet this makes the notation '*p*', '*q*', «$p \wedge q$» ambiguous, and the notion of ambiguous logical notation runs counter to the very idea of introducing logical notation in the first place.

All the same, let it be granted that, for instance, in «*Ks* $(p \wedge q)$» the complement '$p \wedge q$' denotes the *intension* of $p \wedge q$ so that *s* knows that $p \wedge q$ is a true proposition. This still will not do, though, exactly because of the crude individuation of *p*, *q*. The moral is that it is not going to be enough to sort out the contrast between *p*, *q* and then simply prefix some epistemic operator to whatever is the result. There is a leap from extensional and possible-world intensional contexts to hyperintensional contexts that cannot be neglected either conceptually or notationally.

Conclusion

Schaffer's contrastivist programme lacks as yet a theory of appropriate objects of knowledge. Whatever theory may be advanced to fill the lacuna, it seems nonnegotiable that the objects would need to be hyperpropositions. Another nonnegotiable requirement would

---

Otto', but Dretske maintains that «in knowing that Lefty *killed* Otto […] we do not *necessarily* […] know that *Lefty* killed Otto […].» It is this differentiation that calls for hyperintensional objects of knowledge.

9. I suspect the historical culprit for this sort of notation must be the conception of modalities due to possible-world semantics, which treats '$\Box$', '$\Diamond$' as being syntactically on a par with '$\neg$'; both «$\neg p$» and «$\Box p$» are well-formed formulae. This makes for handy notation, but it remains implicit that the argument of $\neg$ is a truth-value of *p* and the argument of $\Box$, *p* itself, i.e., the entire function. If '*K*' is introduced as a notational variant of '$\Box$' we get formulae like «*Kp*», and we are allowed to generate strings like, «$\neg p \wedge K\neg p$», where the extension/intension ambiguity of the notation is manifest. Moreover, if *K* is a hyperintensional operator, and $\Box$ an intensional operator, then we are in for three-way ambiguity as in, «$(\Box p \rightarrow p) \wedge Kp$».

seem to be that the internal structure of the contrastivist hyperproposition *p*_rather_than_*q* must accommodate a *contrast* between the selected *p* and the eliminated *q*. The fundamental problem, however, remains how to reconcile Ternicity with the fact that *p*_rather_than_*q* is one object rather than two.[10]

References

Bealer, G., 1982, *Quality and Concept*, Oxford: Clarendon Press.

Blakemore, D., 2000, 'Indicators and procedures: *nevertheless* and *but*', *Journal of Linguistics*, vol. 36, pp. 463-86.

Brendel, E. and Jäger, C., (2004), 'Contextualist approaches to epistemology: problems and prospects', *Erkenntnis*, vol. 61, pp. 143-72.

Cresswell, M.J., 1975, 'Hyperintensional logic', *Studia Logica*, vol. 34, pp. 23-38.

DeRose, K., 1992, 'Contextualism and knowledge attributions', *Philosophy and Phenomenological Research*, vol. 59, pp. 913-29.

DeRose, K., 1995, 'Solving the sceptical problem', *Philosophical Review*, vol. 104, pp. 1-52.

Douven, I., 2004, 'The context-insensitivity of 'knowing more' and 'knowing better', *Canadian Journal of Philosophy*, vol. 34, pp. 313-26.

Dretske, F., 1970, 'Epistemic operators', *Journal of Philosophy*, vol. 67, pp. 1007-23.

Evnine, S.J., 1999, 'Believing conjunctions', *Synthese*, vol. 118, pp. 201-28.

Schaffer, Jonathan, 2004, 'From contextualism to contrastivism', *Philosophical Studies*, vol. 119, pp. 73-103.

Schaffer, Jonathan, 2005, 'Contrastive knowledge', in: Gendler and Hawthorne (eds.), *Oxford Studies in Epistemology*, vol. 1.

Schaffer, Jonathan, (ms.), 'Contrastive closure and answers', paper read at Free University of Amsterdam, October 2004. [Published as Schaffer (2007)].

[10]. The material presented herein is based on my reply to Jonathan Schaffer's talk 'Contrastive closure and answers', which was delivered at Vrije Universitiet, Amsterdam, in October 2004 and published as Schaffer (2007). Thanks to Jonathan Schaffer for follow-up discussion via email.

Schaffer, Jonathan, 2007, 'Closure, contrasts, and answers', *Philosophical Studies*, vol. 133, pp. 233-55.

Bjørn Jespersen
Section of Philosophy
Delft University of Technology
Delft, The Netherlands
B.T.F.Jespersen@tudelft.nl

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 135-140
The Two Envelope Paradox and Using Variables Within the Expectation Formula

THE TWO ENVELOPE PARADOX AND USING VARIABLES WITHIN THE EXPECTATION FORMULA

by Eric Schwitzgebel & Josh Dever

The paradox

You are presented with a choice between two envelopes. You know one envelope contains twice as much money as the other, but you don't know which contains more. You arbitrarily choose one envelope — call it Envelope A — but don't open it. Call the amount of money in that envelope X. Since your choice was arbitrary, the other envelope (Envelope B) is 50% likely to be the envelope with more and 50% likely to be the envelope with less. But, strangely, that very fact might make Envelope B seem attractive: Wouldn't switching to Envelope B give you a 50% chance of doubling your money and a 50% chance of halving it? Since double or nothing is a fair bet, double or half is more than fair. Applying the standard expectation formula, you might calculate the expected value of switching to Envelope B as (.50)½X [50% chance it has less] + (.50)2X [50% chance it has more] = (1.25)X. So, it seems, you ought to switch to Envelope B: Your expected return — your return on average, over the long run, if you did this many times — would seem to be 25% more. But obviously that's absurd: A symmetrical calculation could persuade you to switch back to Envelope A. Hence the paradox.

Where have we gone wrong? What's the flaw in the reasoning? Despite many interesting discussions of *alternative* ways to reason through the Two Envelope paradox, no one has given a fully adequate answer to *this* question — no one has fully exposed the nature of the misstep.[1] The problem, surely, has something to do with how variables are deployed in the fallacious argument. Proper diagnosis of the fallacy, then, should help clarify more generally what counts as proper or improper use of variables within the expectation formula.

Other discussions of the Two Envelope paradox have tended either to focus on the «open envelope version» of the paradox, in which one gets to see the contents of the chosen envelope before deciding whether to switch (and we agree with the general consensus here that whether to switch depends on what you see, and only weird probability distributions generate the result that you should switch no matter what you see); or they have satisfied themselves with vague remarks the mathematical grounding of which is unclear; or they have advocated constraints on the use of variables in the expectation formula that are, we think,

---

1. Discussions include Nalebuff 1989; Marinoff 1993; Jackson, Menzies, and Oppy 1994; Broome 1995; Chihara 1995; Jeffrey 1995; Rawling 1997; Clark and Shackel 2000; Horgan 2000; Chalmers 2002; Meacham and Weisberg 2003; Priest and Restall 2003; Dietrich and List 2004; Langtry 2004.

considerably more restrictive than necessary.

An analogously absurd case. Our solution to the paradox essentially analogizes the reasoning above to the following reasoning, where the source of the problem is more obvious: You are presented with an envelope containing either $1, $2, $10, or $20 with equal probability. You are given the choice between two wagers. On the first, you receive twice the amount of money in the envelope, if the amount in the envelope is $1 or $2, or just the amount of money in the envelope if the amount in the envelope is $10 or $20. The second wager is the reverse: You receive twice the amount of money in the envelope if the envelope contains $10 or $20 and just the amount of money in the envelope if it contains $1 or $2. Assigning X to the amount in the envelope, you reason that on either bet there is a 50% chance you will receive X and a 50% chance you will receive 2X (for an expectation of 3/2 X), so you are indifferent between the two bets.

Wager 1	Wager 2
$1 * 2	$1
$2 * 2	$2
$10	$10 * 2
$20	$20 * 2

Clearly, however, the second wager is preferable. It's much better to have the chance to double $10 or $20 than to have the chance to double $1 or $2. The proposed fallacious calculation is fallacious because it does not take that into account. (The actual expectation, which can be calculated on a case-by-case basis, is $9 for the first wager and $15.75 for the second.) In Wager 1, the expected value of X in the «2X» part of the formula is much lower than the expected value of X in the «X» part of the formula; in Wager 2, the reverse is the case. A decision-theoretic calculation in which a random variable does not maintain the same expected value in each of its occurrences has no guarantee of producing proper results.

The Solution. Analogously, in the Two Envelope Paradox, the expected value of X in the «2X» part of the formula (where Envelope A is the envelope with less) is less than the expected value in the «½X» part of the formula (where Envelope A is the envelope with more).[2] You would expect less in Envelope A if you knew that it was the envelope with less than you would if you knew it was the envelope with more. Allowing X to have different expectations in different parts of the formula in this way is like comparing apples and oranges. The «X» in the «2X» just isn't the same as the «X» in the «½X» part.

The proper course of action in the Two Envelope Paradox can be non-paradoxically calculated by setting X to the amount in the envelope with less and calculating the expected value of Envelope B as (.50)X + (.50)2X = 3/2 X — and the expected value Envelope A likewise as (.50)X + (.50)2X = 3/2 X. In these calculations the expectation of X in the first term of each equation is identical to its expectation in the second term: The expected amount

---

[2]. We assume, contra Langtry (2004), that X here is a random variable with an expected value. If that seems troublesome to you in the case as described, imagine the following variation: For every possible value of X from one cent to $20 trillion, you have some (small!) determinate degree of confidence that that value is the value in X. For higher and lower values, your subjective probability is zero. Perhaps, indeed, the person offering you the wager, in whom you have absolute faith, provides you with a full list of those probabilities. Suppose, also, that you also have a ballpark sense of how much richer you'll be if you take the contents of the envelope: probably just a few dollars. We see no reason, in such a case — or indeed in the more sparsely presented case — not to suppose, for decision-theoretic purposes, that X can be interpreted as a random variable with a determinate or approximate expected value.

of money in the envelope with less does not change depending on whether Envelope A is the envelope with more or Envelope B is. The availability of such a non-paradoxical calculation is old news, of course; the novelty here is the identification of the crucial difference between the paradoxical and non-paradoxical calculation.

In general, we propose as a constraint on the use of variables within the expectation formula that their expected value be the same at each occurrence in the formula. More formally: For all events A_i in the partition of the outcome space, $E(X/A_i) = E(X)$. Abiding by this constraint guarantees the legitimacy of calculations using X as a variable, if all the equations involved are linear (as we will explain more fully below).[3]

Stronger constraints are too restrictive

Jackson and Oppy (1994) and others following them have proposed a stronger constraint: that to use a formula like E(Y) = (.50)½X + (.50)2X it must be the case that for all values of X there's a 50% chance that the value of Y is half the value of X and a 50% chance that the value of Y is twice the value of X.

While applying this constraint would indeed allow one to avoid the paradox, it also rules out other cases where the formula seems intuitively appropriate. Suppose for example that you're about to mug Mary. Around the corner comes someone else — either Terri or Geri, with 50% likelihood of each. You know that Terri usually carries about half as much money as Mary and Geri usually carries about twice as much. It's perfectly appropriate (moral remonstrances aside) to calculate the expected value of letting Mary go in favor of mugging the oncoming party as (.50)½X + (.50)2X. To calculate in this way, it is not necessary that for all possible dollar amounts in Mary's purse, there be a 50% likelihood that the person coming around the corner has half as much and a 50% likelihood she has twice as much. Perhaps when Mary has $84.57 (which can't even be halved), Terri always has $101.23. Maybe Geri sometimes has the same amount as Mary, sometimes four times as much, and never exactly twice as much — as long as *on average* she has twice as much, the calculation works, accurately reflecting the long-run expectations. What matters is *not* that the relationships among the *each particular possible value of* X and Y exactly mirror the relationships in the overall formula, but rather that the *overall expected values* of X and Y exhibit the right relationship.

In principle, of course, the expectations *could* be calculated case-by-case for different possible values carried by Mary, and some purists we've encountered insist that calculating case by case is the only «technically correct» approach — that one simply cannot legitimately combine random variables in the way suggested. The problem with this purism, of course, is that case-by-case calculation may often in practice be difficult or impossible. Thus, it's of potentially great value to the decision theorist to know when case-by-case calculation is genuinely necessary, and when it may be circumvented by short-cut techniques without affecting the outcome of the decision — which is, of course, exactly the question the Two Envelope paradox raises so forcefully.

---

[3]. Chihara (1995) and Horgan (2000) have proposed constraints that bear some similarity to ours, but which in our view are vague and difficult to interpret, and which fail to note the importance of linearity. Priest and Restall (2003) also offer a similar constraint, which is a bit clearer but too strong: that the actual value of X be the same in both events. That this constraint is needlessly strong can be seen from the example involving the coin flip and cards below, and from the proof.

Needless to say, we see little value in still stronger constraints, such as (per Jeffrey 1995) that one can discharge such X-for-Y substitutions only when X is a true constant. Such excess caution needlessly robs us of the convenience of simple calculations.

Generalizing

Abiding only by our constraint allows also us to generalize to other cases, less intuitive, that stronger constraints forbid us. Consider this case: You have a choice between two wagers. In the first wager, a fair coin is tossed. If it lands heads, you are to draw one of three cards, marked 0, 2, and 4, winning half the amount on the card. (i.e., \$0, \$1, or \$2). If it lands tails, you are to draw one of two cards, marked 1 and 3, winning two more than the amount on the card (i.e., \$3 or \$5). The second wager begins with a similar coin flip and drawing. However, given heads you win 70% of the amount on the card, plus 1 (\$1, \$2.40, or \$3.80). Given tails you win simply 70% of the amount on the card (\$0.70 or \$2.10).

Wager A:	Heads:	0 → \$0	Tails:	1 → \$3
		2 → \$1		3 → \$5
		4 → \$2		
Wager B:	Heads:	0 → \$1	Tails:	1 → \$0.70
		2 → \$2.40		3 → \$2.10
		4 → \$3.80		

We can let X be the amount on the card: The expectation of X is the same given heads or tails — 2 in both cases. The first wager is thus worth (.50)½X + (.50)(X+2), which simplifies to (.75)X + 1. The second wager is worth (.50)[(.7)X+1)] + (.50)(.7)X, which simplifies to (.70)X + .5. We can thus see that the first wager is preferable without calculating case-by-case — which is obviously a great advantage as the number of cards in the two decks increases! Stronger constraints forbid such calculations.

The proof

As long as one abides by the constraint we propose — *that the conditional expectation of the variable be the same in each term or condition of the equation* (i.e., in each event in the partition) — and by one additional constraint, *that the functions be linear* (this second constraint, though necessary, is perhaps not obvious), one will come to the same results in one's calculations as one would working by the more arduous case-by-case method, calculating the expectations for each particular value. Why? If the expectation of Y (the ultimate outcome you're interested in) is a linear function $g_i = m_i x + b_i$ of the expectation of X (the variable in question) in various conditions A_i (possibly a different linear function in different A_i), then

$$E(Y) = \Sigma_i[m_i E(X/A_i) + b_i]P(A_i).$$

If X has the same expected value in the different conditions A_i, then $E(X/A_i) = E(X)$, and consequently

$$E(Y) = \Sigma_i[m_i E(X) + b_i]P(A_i).$$

In other words, one can calculate the expectation of Y by summing the different g_i functions on the expectation of X (which needn't actually be calculated) times the probability of the A_i — the kind of stuff we were doing above, the kind of maneuver we'd like to make, that it often makes intuitive sense to make, but that the Two Envelope paradox may bring us to doubt the

validity of. Getting rid of the $E(X/A_i)$ in favor of $E(X)$ is crucial here: It means one can treat X as the same in every condition, which is key to simplifying the equation into an interpretable result (e.g., simplifying $(.50)X + (.50)2X$ into $3/2\ X$). The linearity is crucial to distributing the g_i functions outside the scope of the expectation of X in the first step.[4]

Conclusion

We don't claim to be presenting a novel or profound mathematical result. But we do hope these remarks will prove useful to the reader who feels the pull of puzzlement, as we do, about what has gone wrong in the reasoning of the Two Envelope Paradox but sees no straightforward solution that doesn't — as do all published solutions we've seen — forbid other sorts of calculations that it seems perfectly reasonable to make.[5]

References

Broome, J. (1995). The two envelope paradox. *Analysis* 55, 6-11.

Chalmers, D.J. (2002). The St. Petersburg two-envelope paradox. *Analysis* 62, 155-157.

Chihara, C.S. (1995). The mystery of Julius: A paradox in decision theory. *Philosophical Studies* 80, 1-16.

Clark, M., and Shackel, N. (2000). The two envelope paradox. *Mind* 109, 415-442.

Dietrich, F., and List, C. (2004). The two-envelope paradox: An axiomatic approach. *Mind* 114, 239-248.

Horgan, T. (2000). The two envelope paradox, nonstandard expected utility, and the intensionality of probability. *Noûs* 34, 578-603.

Jackson, F., Menzies, P., and Oppy, G. (1994). The two envelope paradox. *Analysis* 54, 43-45.

Jeffrey, R.C. (1995). *Probabilistic thinking*. Available at http://www.princeton.edu/~bayesway/ProbThink/TableOfContents.html.

Langtry, B. (2004). The classical and maximin versions of the two envelope paradox. *Australasian Journal of Logic* 2, 30-42.

Marinoff, L. (1993). Three pseudo-paradoxes in 'quantum' decision theory: Apparent effects of observation on probability and utility. *Theory and decision*, 35, 55-73.

Meacham, C.J.G., and Weisberg, J. (2003). Clark and Shackel on the two-envelope paradox. *Mind* 112, 685-689.

---

4. Consider the following case, where the violation of linearity is to blame: A coin is flipped. If it lands heads, $10 is put in an envelope. If it lands tails, either $0 or $20 is put in the envelope. You are given a choice of the following two wagers: (1.) The amount of money in the envelope, if the coin landed heads, or the amount squared if it landed tails, or (2.) the amount of money in the envelope, if the coin landed tails, or the amount squared if it landed heads. The expectation of X is the same, given heads or tails, but linearity is violated, and characterizing your expectation as $.5X + .5X^2$ in the two cases leads to an erroneous recommendation of indifference. No such trouble if instead of squaring, one doubles and adds one.

5. The outlines of this position were developed in 1993, with input from numerous people in the Berkeley Philosophy Department, most memorably Charles Chihara, Edward Cushman, and Sean Kelly. We have also profited from more recent discussions with Terry Horgan, Brian Skyrms, Peter Vanderschraaf, and others.

Nalebuff, B. (1989). Puzzle: The other person's envelope is always greener. *Journal of Economic Perspectives* 3, 171-181.

Priest, G., and Restall, G. (2003). Envelopes and indifference. Available at http://consequently.org/papers/envelopes.pdf.

Rawling, P. (1997). Perspectives on a pair of envelopes. *Theory and Decision* 43, 253-277.

Eric Schwitzgebel
Department of Philosophy
University of California at Riverside
Riverside, CA
eschwitz@ucr.edu

Josh Dever
Department of Philosophy
University of Texas at Austin
Austin, TX
dever@mail.utexas.edu

SORITES (Σωρίτης), ISSN 1135-1349
http://www.sorites.org
Issue # 20 — March 2008. Pp. 141-156
Hypothesis Testing Analysis

HYPOTHESIS TESTING ANALYSIS

by Mikael Eriksson

---

1. Introduction

Looking directly into some technicalities of the philosophical empirical tradition, the essence of hypothesis confirmation and hypothesis falsification is viewed such that hypotheses are based in, and perhaps are stronger than, the evidence that experience in turn provides. Confirmation singles out those hypotheses that base in, and perhaps go beyond, the evidence. Falsification, on the other hand, singles out those hypotheses being inconsistently based in evidence. The very same traditions recognise the well-known consequences that confirmation and falsification involve induction and correlation problems. Confirmation tends to include more information than the evidence provides, while falsification is not strong enough to generally tell what is false in the falsified hypothesis. As if this would not be problem enough, it is also generally viewed that confirmation and falsification clash with each other. These problems seem to put confirmation and falsification in a pretty awkward place of fulfilling the conditions of being a set of logical constants validating the natural phenomenon of empirical knowledge gaining activities.

For starters, look at the notion of falsification. Popper talked about a hypothesis h being better corroborated than a hypothesis h' on some evidence e, in the sense that h is capable of being tested for falsification by more observational statements than h'. (Popper,1972) Corroboration includes this falsification, which in turn has the consequence that some given observation, testing a hypothesis, falsifies it without being able to show which part of the hypothesis is false. The accustomed logical understanding of this is as follows. The predicate «e falsifies h» boils down to the formula «h implicates e». A well-known logical theorem says that in denying a formula consequent, it does not follow which part of the antecedent is false. Therefore, when e falsifies h and h is a complex formula, it is not clear which part of h is false. However, looking at the natural idea of falsification, it clearly indicates the idea of proper evidence and hypothesis correlation, demanding more of «e falsifies h» than the implication analysis shown. Following this, it seems to me that the suggested Popper logical understanding of falsification needs revision. Rather, how does h and e correlate?

Now, what has the empiricists to offer? The vast empiricist tradition searches for methods of gaining empirical knowledge along the notion of induction, taking off in the Hume tradition. (Hume,1777) In the early 20^{th} century the empirical ideas took further steps, by understanding confirmation as testing, which by that time was viewed as an inductive and non-demonstrative inference from evidence to hypotheses. Later on, empiricists developed the basic confirmation idea understanding into demonstrative probability, still emulating induction. There, Carnap stipulated that the formula «evidence confirms a hypothesis» means a

hypothesis becoming accepted scientific knowledge and nearly (increasingly) certain in the light of one (several) bodies of evidence — absolute (relevance) confirmation. (Carnap,1962) This induction idea was governed by the assumptions that hypotheses are based in evidence, and might be stronger than that evidence. For those cases when the proper hypotheses do not go beyond the evidence, the hypotheses and evidence are indeed identical. I claim that this seemingly fair view is false and leads to paradoxes as section 5.1 will show.

With this outline I have tried to show the problems involved in both empirical accounts — each aiming at the trophy of validating the natural phenomenon of empirical knowledge gaining. The rising question is whether there might be another way to define this empirical set of constants. We know that the falsification account claims formulas on the form «hypothesis h being better corroborated than hypothesis h' on evidence e». Empiricists, on the other hand, claim formulas on the form «evidence e confirms hypothesis h». Putting it this way, it seems to me that both accounts put their efforts on analysing the relation between e and h, while I think the emphasis first need to be on e and h in separate. Developing a system along this view give rise to a new terminology which I hope will bring some light to both the Popper notion of falsification modus tollens and the Carnap notion of confirmation probability.

2. Natural language analysis

2.1 Idea

In pursuing empirical knowledge reasoning, I search for valid argument schemata characterising the scientific everyday language. A usual sentence is «I am interested in this phenomenon or this hypothesis». Out of such ordinary scientific language, it could be possible to single out argument schemata. One way of unveiling these suggested schemata is to put the following question. What makes scientists to accept and be nearly certain of a hypothesis? It seems clear that scientists, after making their conjectures, use evidence to test their hypotheses, aiming at confirming or falsifying the hypotheses. However, the clearness is not preserved when it comes to understand the involved phenomenon of testing. Aiming at this seemingly opacity, I believe that this notion should be understood in terms of correlation of hypothesis and evidence as well as understanding the meaning of hypothesis and evidence. Consider this ancient example. In my view, when Galilei claimed that there is a halo around the moon (not the face of God turned to the observer as claimed by contemporaries), he used his perception to collect evidence. He also made a hypothesis claiming the idea of «halo around the moon». Finally, he used his human ability to relate his hypothesis with his evidence. In a language view, I refer to these as evidence formula, hypothesis formula, and procedure formula. I put these three formulas together in the following way. The procedure formula tests the hypothesis-formula «halo around the moon» with the evidence-formula «halo around the moon».

With this view of testing established, I go further by analysing what these three formulas are. By evidence-formula I mean a logical formula describing the evidence claim (for instance the perception of some entity). By hypothesis-formula I mean a logical formula describing the hypothesis claim (for instance an idea claimed by someone). I mean that the evidence and hypothesis formulas are distinct in character but similar to each other, having the same simple and logical form. Finally, the procedure formula I mentioned is a logical construction of these evidence and hypothesis formulas. This logical construction will depend on the logical form of the evidence and hypothesis formulas, as section 3 shows. Summing this up, I view and analyse testing as including procedures of hypothesis and evidence

formulas.

This analysis of testing makes the first face of empirical knowledge-gaining (besides conjecturing). Note that this face of testing involves the principle of similar evidence and hypothesis formulas, which later will be shown to explains the rationalist's evidence and hypothesis correlation problem. I believe there is a second face of empirical knowledge-gaining, which I claim to be hypothesis inclusion. Now, this means that my viewed two-faced formulation of empirical knowledge-gaining divides a scientist hypothesis claim into two parts. The first part is a sub hypothesis, which is tested by evidence. The second part is the rest of the hypothesis, which belongs to the intuition of the scientist, but is not present in the evidence. This way of understanding empirical knowledge-gaining explains the empiricist's induction inference problems, as I will show.

Summing the idea up, I view confirmation as constructive empirical knowledge-gaining and I analyse it in terms of testing and inclusion. This can be expressed by the conjunction of the two formulas «this evidence tests this hypothesis-part» and «the full hypothesis compares with that hypothesis-part». Along this view, I fit in falsification as reductive empirical knowledge-gaining and I analyse it in terms of reductive testing, as below will show. In this way, both confirmation and falsification is defined in terms of testing, seemingly opening a common ground for both the Carnap and Popper terminologies.

2.2 Analysis

2.2.1 Distinctiveness

The above show that I understand testing as including the condition of evidence and hypothesis formulas being distinct from each other. Going into this in detail, I will say that by formulas I mean logical formulas, where the evidence-formula simple parts refer to simple evidence-claims, and the hypothesis-formula simple parts refer to simple hypothesis-claims. By the term distinct, I mean that the simple parts of evidence-formulas are distinct in kind from the simple parts of hypothesis-formulas.

This view of distinctiveness bases in my intuition of what hypothesis and evidence are. To me, evidence-claims and hypotheses-claims are human activities about the natural world. A natural entity differs from both the human perception of it and the human understanding of it. Evidence-claims (related to science) are claims focusing upon natural entities from a perceptual point of view. Analogously, hypothesis-claims (related to science) are claims focusing upon natural entities from an idea point of view. These different points of views of the natural entity include that the evidence have at least some quality different from the hypothesis. This makes evidence distinct from hypotheses.

Let's see how this distinctiveness applies to the Galilei example. When Galilei made a claim that there is a halo around the moon (not the face of God turned to the observer), he focused upon the natural situation of the moon from his perceptual point of view. His perception included biochemical properties. The very same properties were not included in his hypothesis-claim that there is a halo around the moon. This makes his evidence claim distinct from his hypothesis claim. Relating to Carnap, he also points out on page 12 that neurological factors determine inductive reasoning. (Carnap,1971) Still, he claims that «e implicates h», if e and h refer to state-descriptions having inclusion relation. That is, h is included in e. Understanding the Carnap mentioned neurological factors as I do, makes his claim contradictory, as e and h has some neurological properties distinct from each other. However,

it would be possible to rewrite the Carnap claim in the following manner. My claim of hypothesis and evidence distinctiveness is consistent with world-states including the two kinds of evidence and hypotheses referring to the same situation in the world-state. The evidence, having its properties, is part of a two-faced world-state alongside with the hypothesis, having some distinct properties. However, contrary to Carnap the evidence and hypothesis relation, due to the distinctiveness, is rather conjunctive than implicative. To formalise this evidence E and hypothesis H distinctiveness, I use the following terminology. (A $\models$ B means B is true in A.)

$$E,H \models F,G \quad \text{iff} \quad E \models F \text{ , } H \models G$$

$$E,H \models F \wedge G \quad \text{iff} \quad E \models F \text{ and } H \models G$$

$$E,H \models \neg F \quad \text{iff} \quad \text{not } E,H \models F$$

$$E,H \models F \rightarrow G \quad \text{iff} \quad E,H \models \neg(F \wedge \neg G)$$

2.2.2 Similarity

After analysing the distinctiveness of e and h I go on investigating the e and h relation. Rather, I understand testing as including the condition of an evidence-formula F and a hypothesis-formula G being similar in simple form and the same in logical form. By similar simple form I mean that the logically simple evidence claim F of a natural entity and the logically simple hypothesis claim G of that natural entity is similar. Same logical form means that F has the same logical structure as G. I set a terminology for this, as follows.

The indexed formulas F_i and G_i mean that F has similar simple form and same logical form as G.

2.2.3 Matching, procedure and equivalence

I also understand the notion of testing as including the matching condition. Matching is the combination of the condition of distinct evidence and hypothesis formulas with the condition of similar evidence and hypothesis formulas. Matching M is the conjunction of the distinct evidence and hypothesis formulas, having similar simple and same complex logical form.

$$M(F_i,G_i) \quad \text{iff} \quad F_i \wedge G_i$$

To me, the notion of testing also includes a procedure showing how the test is about to be done — its procedure. In language, the procedure is defined as a logical construction of the involved matching predicates. That is, the logical construction depend upon which arguments the matching predicates have. For instance, the procedure of testing a conjunction formula is to first test one of the conjuncts and then to test the second conjunct.

Procedure of testing $G_i \wedge G_j$ is the procedure of testing G_i and the procedure of testing G_j

Procedure of testing $G_i \vee G_j$ is either the procedure of testing G_i or the procedure of testing G_j

My final understanding of testing is that it involves an expression condition, inspired by the elegant natural formulation «evidence tests hypothesis». This natural expression form denotes any test procedure, indifferent of its logical construction. In formal language, this behaves much like a Frege predicate showing free variables within parentheses. Here, the predicate variables are instead exchanged with formulas. I believe there is a brilliant natural language focusing feature involved here which I extract for this article purpose as follows. The test-predicate T, having the simple argument F_i and G_i, defines as the simple test procedure $M(F_i,G_i)$.

$$T(F_i,G_i) \quad \text{iff} \quad M(F_i,G_i)$$

Test procedures differ with the hypotheses being tested because the logical constructions of the hypotheses differ. This means that the logical properties of the test predicate T differ with the logical form of the arguments, as section 3.3 will show. This has the unexpected consequence that a system built on these premises include cases where two test predicates have logically equivalent formulas as arguments without the test predicates having the same logical properties. This phenomenon has special consequences for the well-known raven paradox, as section 6 will show.

3. Formal suggestion

My basic aim in this article is to view the idea of empirical knowledge gaining in terms of confirmation and falsification, but give both those notions a base in the notion of testing. I will now try to fully formalise the notion of testing into a system and then derive the notion of confirmation and falsification from that system.

3.1 Simple test predicate rule

For starters, using natural deduction style, the section 2.2 formally includes below deduction.

	$E \vdash F_i$ and $H \vdash G_i$
$I\wedge$	$E,H \vdash F_i \wedge G_i$

The formation rule of matching is as follows.

FM	$E,H \vdash F_i \wedge G_i$
	$E,H \vdash M(F_i,G_i)$

Let the indexed formulas F_i and G_i be «the evidence claim F has similar simple form and same logical form as the hypothesis claim G». Let F_i be a simple formula in system E and let G_i be a simple formula in system H. Let M be matching. Let $M(F_i,G_i)$ be «matching M of F_i and G_i denotes conjunction formula $F_i \wedge G_i$ focusing upon F_i and G_i». Read $M(F_i,G_i)$ as F_i matches G_i.

IM	$E,H \vdash F_i \wedge G_i$
	$E,H \vdash M(F_i,G_i)$

EM	$E,H \vdash M(F_i,G_i)$
	$E,H \vdash F_i \wedge G_i$

The section 2.2 also includes a way of expressing matching, formally expressed as the test predicate $T(G_i)$. The test predicate T takes hypothesis formula G_i, the outstanding part of the formula $M(F_i,G_i)$, as argument.

FT	$E,H \vdash M(F_i,G_i)$
	$E,H \vdash T(F_i,G_i)$

F_i and G_i are simple formulas. $M(F_i,G_i)$ is a formula in the systems E,H. Let T be testing. Let $T(F_i,G_i)$ be «testing T of F_i and G_i denotes $M(F_i,G_i)$ focusing upon F_i and G_i». Read $T(F_i,G_i)$ as F_i tests G_i.

IT'	$E,H \vdash M(F_i,G_i)$
	$E,H \vdash T\,`(G_i)$

Read $T\,`(G_i)$ as test of G_i

I will now reformulate the rule IT ‘ to below rule for more natural reading of testing.

IT	$E,H \vdash M(F_i,G_i)$
	$E,H \vdash T(F_i,G_i)$

Read $T(F_i,G_i)$ as F_i tests G_i

ET	$E,H \vdash T(F_i,G_i)$
	$E,H \vdash M(F_i,G_i)$

3.2 Probability rule

Carnap viewed confirmation as being defined as probability, which can be seen in his quantitative concept «evidence supports hypothesis to some degree». (Carnap,1962) In this article I surely claim that probability is part of the quantitative empirical knowledge gaining intuition, but it does not coincide with the qualitative notion of confirmation. Along this line, I suggest the probability notion «the probability of evidence testing hypotheses», distinguishing testing from probability. Here, the quantitative empirical knowledge intuition passes over to the notion of probability, leaving testing a purely qualitative empirical knowledge intuition. In this way, I can nest probability formulas inside test formulas and vice versa as the following two examples will show. In the first example, I test T that at least one individual being a swan G_i and white G_j; $T(F_i \wedge F_j , G_i \wedge G_j)$. (Section 3.3 shows this T conjunction case.) I also test precisely one thousand individuals being white; $T(F_k,G_k)$. This means two tested hypotheses. Now, the probability P of there being tested white individuals that also are tested as swans, is at least one of one thousand; $P(T(F_i \wedge F_j , G_i \wedge G_j) | T(F_k,G_k))$. (Below IP applied to 3.3 explains this formulation.) In this way, I talk about tested hypotheses, and probability applied to these. In the second example, I test the following probability hypothesis. At least one of one thousand whites is a swan. That is, $P(G_i \wedge G_j | G_k)$. To test this I need an evidence formula similar to the hypothesis. I use the evidence arguments of above T predicates, and apply probability to these. That is, $P(F_i \wedge F_j | F_k)$. Now, the evidence formula tests the similar hypothesis formula — $T(P(F_i \wedge F_j | F_k) , P(G_i \wedge G_j | G_k))$. This shows the relationships between the two notions of testing and probability.

IP	$S \vdash F$
	$S \vdash P(F)$

, where $P(F_i \vee F_j) = P(F_i) + P(F_j) - P(F_i \wedge F_j)$ and $P(F_i \wedge F_j) = P(F_i | F_j) * P(F_j)$ and the conditional $P(F_i | F_j) = P(F_i \wedge F_j) / P(F_j)$, as well as its converse that if $P(F_i | F_j)$ then $P(F_j | F_i) = P(F_j \wedge F_i) / P(F_j \wedge F_i) + P(\neg F_j \wedge F_i)$.

3.3 Complex test predicates

The section 2.2 makes clear that the test predicates emulate the test part of the natural phenomenon of confirmation. The section 3.1 IT rule singles out the test predicates having simple arguments. Below will define the remaining test predicates, having logically complex evidence and hypotheses.

T∧ $\_\_E,H \vdash T(F_i \wedge F_j\ ,\ G_i \wedge G_j)$

$\_\_\_\_\_\Leftrightarrow$

	$E \vdash F_i \wedge F_j\ ,\ H \vdash G_i \wedge G_j$
I∧	$E,H \vdash (F_i \wedge F_j) \wedge (G_i \wedge G_j)$
IM	$E,H \vdash M(F_i \wedge F_j\ ,\ G_i \wedge G_j)$

T→ $E,H \vdash T(F_i \rightarrow F_j\ ,\ G_i \rightarrow G_j)$

$\_\_\_\_\_\Leftrightarrow$

	$E \vdash F_i\ ,\ E \vdash F_i \rightarrow F_j$	$H \vdash G_i\ ,\ H \vdash G_i \rightarrow G_j$
I∧	$E,H \vdash (F_i \rightarrow F_j) \wedge (G_i \rightarrow G_j)$	
IM	$E,H \vdash M(F_i \rightarrow F_j\ ,\ G_i \rightarrow G_j)$	

T∨ $\_\_\_\_E,H \vdash T(F_i \vee F_j\ ,\ G_i \vee G_j)$

$\_\_\_\_\_\Leftrightarrow$

	Either_$E \vdash F_i$ or $E \vdash F_j$, either $H \vdash G_i$ or $H \vdash G_j$
I∨	$E,H \vdash (F_i \vee F_j)\ ,\ (G_i \vee G_j)$
I∧	$E,H \vdash (F_i \vee F_j) \wedge (G_i \vee G_j)$
IM	$E,H \vdash M(F_i \vee F_j\ ,\ G_i \vee G_j)$

T¬ $\_\_\_\_\_E,H \vdash T(\neg F_i\ ,\ \neg G_i)$

$\_\_\_\_\_\Leftrightarrow$

	$E \vdash F_i\ ,\ H \vdash G_i$
	$:\ ,\ :$
	$\bot\ ,\ \bot$
I¬	$E \vdash \neg F_i\ ,\ H \vdash \neg G_i$
I∧	$E,H \vdash \neg F_i \wedge \neg G_i$
IM	$E,H \vdash M(\neg F_i\ ,\ \neg G_i)$

T∀ $\_\_\_\_\_E,H \vdash T(\ \forall v F_i(v)\ ,\ \forall u G_i(u)\ )$

$\_\_\_\_\_\Leftrightarrow$

	$E \vdash F_i(c)\ ,\ H \vdash G_i(d)$
I∀	$E \vdash \forall v F_i(v)\ ,\ H \vdash \forall u G_i(u)$
I∧	$E,H \vdash \forall v F_i(v) \wedge \forall u G_i(u)$
IM	$E,H \vdash M(\ \forall v F_i(v)\ ,\ \forall u G_i(u)\ )$

, for arbitrary c in E and arbitrary $\vdash$ in H.

T$\exists$ $\quad$ $\dfrac{E,H \vdash T(\exists vF_i(v) , \exists uG_i(u))}{\Leftrightarrow}$

	$E \vdash F_i(c)$, $H \vdash G_i(d)$
I$\exists$	$E \vdash \exists vF_i(v)$, $H \vdash \exists uG_i(u)$
I$\wedge$	$E,H \vdash \exists vF_i(v) \wedge \exists uG_i(u)$
IM	$E,H \vdash M(\exists vF_i(v) , \exists uG_i(u))$

, for some c in E and some d in H.

The T disjunction case is exclusive disjunction. In the T quantification case, the variables are relative their E and H domains. The conjunction case for understanding above test predicates is as follows. Suppose the hypothesis conjunction and find an evidence conjunction matching the hypothesis. Introduce T to denote the performed logical steps and read it as «evidence tests hypothesis».

The system developed above has a special effect on the well-known substitution principle. In that principle, any formula including $A \rightarrow B$ means the same as the formula instead including $\neg B \rightarrow \neg A$. Now, check this in the system above. Suppose that the formula including $A \rightarrow B$ is above T implication case. The T implication case definition shows that $T(A \rightarrow B)$ includes A in separate. It is easy to see by the same T implication definition that $T(\neg B \rightarrow \neg A)$ includes $\neg B$ in separate. Now, take the three formulas A, $A \rightarrow B$, and $\neg B$ included in above two T predicates. These formulas show that a contradiction follows from claiming that $\neg B \rightarrow \neg A$ substitutes $A \rightarrow B$ in T. Therefore, the substitution principle does not hold for extensional logic with T predicates added.

3.4 Test predicate logical properties

I will now show some logical consequences of the 3.1 and 3.3 test predicates.

Simple F_i and G_i $\quad T(F_i,G_i) \leftrightarrow (F_i \wedge G_i)$

Conjunction $\quad T(F_i \wedge F_j , G_i \wedge G_j) \leftrightarrow (T(F_i,G_i) \wedge T(F_j,G_j))$

Implication $\quad T(F_i \rightarrow F_j , G_i \rightarrow G_j) \leftrightarrow (T(F_i,G_i) \wedge G_i \rightarrow G_j)$

Disjunction $\quad T(F_i \vee F_j , G_i \vee G_j) \leftrightarrow$ (either $T(F_i,G_i)$ or $T(F_j,G_j)$)

Contradiction $\quad (T(\neg F_i,\neg G_i) \rightarrow T(\bot)) \rightarrow T(F_i,G_i)$

Negation $\quad (T(F_i,G_i) \rightarrow T(\bot)) \rightarrow T(\neg F_i,\neg G_i)$

$\quad (T(F_i,G_i) \wedge T(\neg F_i,\neg G_i)) \rightarrow T(\bot)$

All quantification $\quad T(\forall vF(v) , \forall uG(u)) \leftrightarrow T(F(c) , G(d))$, for arbitrary c and d.

Existence quantification $\quad T(F_i(c) , G_i(d)) \rightarrow T(\exists vF_i(v) , \exists uG_i(u))$, for some c and d.

$\quad (((T(F_i(c) , G_i(d)) \rightarrow I) \rightarrow (T(\exists vF_i(v) , \exists uG_i(u)) \rightarrow I)) \wedge$

$\quad (T(F_i(c) , G_i(d)) \rightarrow T(\exists vF_i(v) , \exists uG_i(u)))) \rightarrow I$

Probability $\quad (P(T(F_i,G_i)) \wedge (T(F_i,G_i) \leftrightarrow T(F_j,G_j))) \rightarrow P(T(F_j,G_j))$

$\quad (T(P(A_i)) \wedge P(A_i) = P(B_i)) \rightarrow T(P(B_i))$

With this formalism hanging over our heads, here is an intuitive example of understanding above formulas. $T(F_i \wedge F_j , G_i \wedge G_j)$ means that the evidence formula tests the hypothesis formula, following the conjunction test procedure (logical steps). The natural language reading of above conjunction formula is «testing a conjunction formula means that each conjunct tests in separate». I will not prove all these theorems due to the length of this paper. I will only prove the conjunction case.

Suppose	$E,H \vdash T(F_i \wedge F_j , G_i \wedge G_j)$	(i)
By 3.3 T∧ def. below (ii)-(iii)	$E \vdash F_i \wedge F_j \ , \ H \vdash G_i \wedge G_j$	(ii)
I∧	$E,H \vdash (F_i \wedge F_j) \wedge (G_i \wedge G_j)$	
IM	$E,H \vdash M(F_i \wedge F_j , G_i \wedge G_j)$	(iii)
EM	$E,H \vdash (F_i \wedge F_j) \wedge (G_i \wedge G_j)$	
E∧	$E \vdash F_i \wedge F_j \ , \ H \vdash G_i \wedge G_j$	
E∧	$E \vdash F_i , E \vdash F_j \ , \ H \vdash G_i , H \vdash G_j$	
Permutation	$E \vdash F_i , H \vdash G_i \ ,\_ \ E \vdash F_j , H \vdash G_j$	(iv)
I∧	$E,H \vdash F_i \wedge G_i \ , \ E,H \vdash G_j \wedge G_j$	
IM	$E,H \vdash M(F_i,G_i) \ , \ E,H \vdash M(F_j,G_j)$	(v)
By T def. above (iv)-(v)	$E,H \vdash T(F_i,G_i) \ , \ E,H \vdash T(F_j,G_j)$	
I∧	$E,H \vdash T(F_i,G_i) \wedge T(F_j,G_j)$	(vi)
I→_(i),(vi)	$E,H \vdash T(F_i \wedge F_j , G_i \wedge G_j) \rightarrow (T(F_i,G_i) \wedge T(F_j,G_j))$	

Suppose	$E,H \vdash T(F_i,G_i) \wedge T(F_j,G_j)$	(i)
E∧	$E,H \vdash T(F_i,G_i) \ , \ E,H \vdash T(F_j,G_j)$	
By T∧ def. (ii)-(iii)	$E \vdash F_i , H \vdash G_i \ , \ E \vdash F_j , H \vdash G_j$	(ii)
I∧	$E,H \vdash F_i \wedge G_i \ , \ E,H \vdash F_j \wedge G_j$	
IM	$E,H \vdash M(F_i,G_j) \ , \ E,H \vdash M(F_j,G_j)$	(iii)
EM	$E,H \vdash F_i \wedge G_i \ , \ E,H \vdash F_j \wedge G_j$	
E∧	$E \vdash F_i , H \vdash G_i \ , \ E \vdash F_j , H \vdash G_j$	
Permutation	$E \vdash F_i , E \vdash F_j \ , \ H \vdash G_i , H \vdash G_j$	(iv)
I∧	$E,H \vdash F_i \wedge F_j \ , \ E,H \vdash G_i \wedge G_j$	
IM	$E,H \vdash M(F_i \wedge F_j , G_i \wedge G_j)$	(v)
By T∧ def. (iv)-(v)	$E,H \vdash T(F_i \wedge F_j , G_i \wedge G_j)$	(vi)
I→_(i),(vi)	$E,H \vdash (T(F_i,G_i) \wedge T(F_j,G_j)) \rightarrow T(F_i \wedge F_j , G_i \wedge G_j)$	

3.5 Complex test predicate rules

Based on 3.3 and 3.4 it is now possible to form natural deduction rules for the complex T predicate. I show this for the conjunction case.

FT∧	$E,H \vdash T(F_i,G_i) \wedge T(F_j,G_j)$
	$E,H \vdash T(F_i \wedge F_j , G_i \wedge G_j)$

F_i is a formula in E and G_i is a formula in H. $T(F_i,G_i)$ is a formula in the systems E,H. Let T be testing. Read $T(F_i \wedge F_j , G_i \wedge G_j)$ as $F_i \wedge F_j$ tests $G_i \wedge G_j$.

IT∧	$E,H \vdash T(F_i,G_i) \wedge T(F_j,G_j)$
	$E,H \vdash T(F_i \wedge F_j , G_i \wedge G_j)$

ET∧	$E,H \vdash T(F_i \wedge F_j , G_i \wedge G_j)$
	$E,H \vdash T(F_i,G_i) \wedge T(F_j,G_j)$

3.6 Confirmation

Until now I have definied the notion of testing in a formal system. The remaining task

is to derive the notion of confirmation and falsification in this system. I start with the notion of confirmation. In my view, confirmation includes more than the above test notion. Scientists often claim hypotheses that after some consideration are shown to partly base in evidence and partly base in their own intuition. Therefore, I view testing as the confirmation base step and complement it with an induction step that will deal with comparing the tested hypothesis part with the full hypothesis. In natural language this will sound like «the evidence F_i tests T the hypothesis-part G_i« and «the full hypothesis G_j compares with G_i». This conjunction is then expressed by «F_i confirms G_i» using the natural language feature of focusing mentioned in section 2.2.3. Formally, the confirmation C base step is testing T the hypothesis part G_i using the evidence part F_i. The confirmation C induction step is a measure function P' of G_i and G_j.

IC	$E,H \vdash T(F_i,G_i) \wedge P'(G_i\|G_j)$
	$E,H \vdash C(F_i,G_i)$

, where P' is a measure tool used to measure the difference between G_i and G_j.

Read $C(F_i,G_i)$ as F_i confirms G_i

3.7 Falsification

Traditionally, confirmation is an empiricist term about hypothesis-support, while falsification is a rival rationalist term about hypothesis-denial. However, in my analysis both terms have testing T in common. In consequence, it seems that it is possible to define both these concepts without rival views emerging. Rather, they go side by side forming a three-unity with testing as base.

I view the natural notion of falsification as reductive testing. That is, falsification is a way of focus upon the evidence and hypothesis parts of the test procedure, where the evidence denies the hypothesis. Formally, this is formulated as follows as section 3.4 test-predicate negation-introduction.

IC	$E,H \vdash (T(F_i,G_i) \rightarrow T(\perp))\_ \rightarrow T(\neg F_i,\neg G_i)$
	$E,H \vdash F(\neg F_i,G_i)$

, where F focuses upon $\neg F_i$ and G_i in the first row formula

Read $F(\neg F_i,G_i)$ as $\neg F_i$ falsifies G_i

The first row antecedent includes that the test argument F_i implicates contradiction $\perp$. Considering the T index-condition, the same applies to G_i. By negation-introduction and test-introduction, the consequent $T(\neg F_i,\neg G_i)$ follows. (Accordingly, E and H revise to exclude F_i and G_i.) In the falsification-introduction, the second row focuses upon the evidence and hypothesis parts in the first row. That is, the rule IF starts with a test procedure going wrong and ends up with focusing upon the evidence and hypothesis parts of that procedure. Falsification application is not used for hypothesis choice, but for evidence and hypothesis revision. Here is an example. If a claimed individual has certain properties, and some is falsified. Then, that individual is falsified.

4. Semantics

The language shown above has a reference as follows. Language L includes symbols for constants c, variables v, functions f, assignment function a, satisfication s, and logical

operators $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\neg$, $\forall$, $\exists$. U is the set of terms, that is constants C, variables V, and functions. S is semantics. M is model set.

4.1 Notation convention

Read the formalism $f : A \rightarrow B\ [X]$

as function f from set A to set B such that condition(s) X

4.2 Assignment function

$a : V \rightarrow C\ [a(v_i) = c_i]$

4.3 Interpretation function

$i : C \rightarrow S\ [(a),(b),(c)]$

(a) $i(c_j) = s_j$

(b) $i(\ f(c_1,c_2,\ldots,c_n)\ ) = i(\ f(i(c_1),i(c_2),\ldots,i(c_n))\ )$

(c) $i(\ f(v_1,v_2,\ldots,v_n)\ ) = i(\ f(a(v_1),a(v_2),\ldots,a(v_n))\ )$

4.4 Satisfication

$M \models c_j = c_k \Leftrightarrow i(c_j) = i(c_k)$

$M \models v_j = v_k \Leftrightarrow a(v_j) = a(v_k)$

$M \models f(t_1,t_2,\ldots,t_n) \Leftrightarrow i(\ f(t_1,t_2,\ldots,t_n)\ )$

4.5 Logical operators

Logical definition over formula A. A is a Frege predicate, a test predicate, or a logical construction of these.

$M \models A_i \wedge A_j$ iff $M \models A_i$ and $M \models A_j$

$M \models \neg A$ iff not $M \models A$

$M \models A_i \vee A_j$ iff either $M \models A_i$ or $M \models A_j$

$M \models A_i \rightarrow A_j \Leftrightarrow M \models \neg(A_i \wedge \neg A_j)$

$M \models \forall vA(v)$ iff for any v if $v \in U$ then $M \models A(v)[a(v)=c]$

$M \models \exists vA(v) \Leftrightarrow M \models \neg\forall v\neg A(v)$

$M \models T(A_i,A_j) \Leftrightarrow M \models A$, where T focuses upon A_i and A_j in A.

5. Application

The basic aim in this article is to contribute to understand the notion of empirical knowledge gaining. I have suggested that the traditional notions of confirmation and falsification can be somewhat modified and derived in an ordinary logical system complemented with the test rule. It is now time to show the strength of this system by defining stronger notions.

5.1 Evidence-devices

Scientists use instruments to get information out of the natural world by quantifying properties. I emulate these instruments as sets of evidence formulas, and name the set evidence-devices E.

Let E be the set of evidence-formulas

Let $E_j \in \wp(E)$, where $\wp$ is power-operator and E_j is evidence-device.

So $E \vdash F_i(v) \Rightarrow \exists E_j(\ E_j \vdash F_i(v)\ )$

Evidence-devices are language descriptions of the instruments used by scientist. Each such device is a language expression of what the instrument has registered. If there should be any use with the hypothesis testing theory in this article, a central task would be to paraphrase the scientific everyday instruments as evidence-devices. This is a vast task.

5.2 Hypothesis-devices

Scientists claim ideas in order to suggest natural-world views. I emulate these ideas as sets of hypothesis formulas, and name the set hypothesis-devices H.

Let H be the set of hypothesis-formulas

Let $H_j \in \wp(H)$, where $\wp$ is power-operator and H_j is hypothesis-device.

So $H \vdash G_i(u) \Rightarrow \exists H_j(H_j \vdash G_i(u))$

Hypothesis-devices would usually be associated with humans being smart enough to come up with something scientifically interesting. However, it would also be possible for machines to be hypothesis-devices.

5.3 Test-devices.

Test-devices $T(E_j,H_j)$ are evidence devices testing hypothesis devices. The test-devices show the proper test procedure for any hypothesis and evidence possible. This shows the strength of the theory, which would be able to guide any kind of empirical knowledge gaining activity. I call the following theorem *hypothesis-testing completeness*.

Def. $E_j,H_j \vdash T(E_j,H_j) \Leftrightarrow E_j,H_j\ d\_\forall F_i,\exists G_i\ T(F_i,G_i) \wedge \forall G_i,\exists F_i\ T(F_i,G_i)$

The test-devices claim that every hypothesis-device can be tested by some evidence-device and every evidence-device can test some hypothesis-device. The proof is as follows. By the definition $T(E_j,H_j)$ is the case. Let H_j be a simple formula. Then by the definition, there is a simple formula E_j testing H_j. Let E_j be a simple formula. Then by the definition, there is a simple formula H_j tested by E_j. Let H_j be a conjunction formula. Then by the definition, there is a conjunction formula E_j testing H_j. Let E_j be a conjunction formula. Then by the definition, there is a conjunction formula H_j tested by E_j. Let H_j be an implication formula. Then by the definition, there is an implication formula E_j testing H_j. Let E_j be an implication formula. Then by the definition, there is an implication formula H_j tested by E_j. Let H_j be an exclusive disjunction formula. Then by the definition, there is an exclusive disjunction formula E_j testing H_j. Let E_j be an exclusive disjunction formula. Then by the definition, there is an exclusive disjunction formula H_j tested by E_j. Let H_j be a negation formula. Then by the definition, there is a negation formula E_j testing H_j. Let E_j be a negation formula. Then by the definition, there is a negation formula H_j tested by E_j. Let H_j be an all quantification formula. Then by the definition, there is an all quantification formula E_j testing H_j. Let E_j be an all quantification formula. Then by the definition, there is an all quantification formula H_j tested by E_j. Let H_j be an existence quantification formula. Then by the definition, there is an existence quantification formula E_j testing H_j. Let E_j be an existence quantification formula. Then by the definition, there is an existence quantification formula H_j tested by E_j. So for any H_j there is an E_j testing H_j and for any E_j there is an H_j tested by E_j.

5.4 Empirical knowledge-gaining programs

Scientists build up theories by putting ideas and information about the world together, perhaps letting the ideas going beyond the actual information at hand. I understand this phenomenon as confirmation, in terms of section 3.6. Applying the above notion of devices to this definition will result in the following formula. Example. Let H_k be Newton mechanics. Then there is a hypothesis part H_j, of H_k, being tested by its proper evidence E_j.

$$E,H \vdash C(E_j,H_j) \Leftrightarrow E,H \vdash T(E_j,H_j) \wedge P'(H_j|H_k)$$

Another knowledge-gaining part is when scientists revise their theories according to hypothesis contradiction or falsification. Section 3.7 shows this. In the device manner, it is possible to formulate.

$$E,H \vdash F(\neg E_j,H_j)$$

Besides confirmation and falsification, scientists use probability. Section 3.2 shows how this works with the notion of testing, and therefore how it works with confirmation and falsification.

$$E,H \vdash P(\ T(E_j,H_j) \mid T(E_k,H_k)\ )$$

Finally, I define the strongest notion in this theory, called empirical knowledge-gaining program. T(Q,M,A) focuses upon the reasoning characteristics of at least one of above three knowledge-gaining parts. This focusing uses the same natural language principle claimed in above 2.2.3. Q is question, M is method, and A is answer. Example. Let T(Q,M,A) be $T(E_j,H_j)$. Q is the supposition of $T(E_j,H_j)$, M is the deduction, and A is the conclusion. I think this theorem shows the procedure of confirming any imaginable hypothesis.

6. Some brief ontological-consequential notes

6.1 Distinctiveness

Empirical knowledge-gaining programs T(Q,M,A)_include the condition of evidence being distinct in kind from hypotheses. This condition challenges the traditional confirmation view that «hypotheses are based upon evidence and go beyond the evidence». In this view, «based upon» means referring to, and «going beyond» means that the hypothesis includes the evidence. In my view, «based upon» means that evidence tests a distinct hypothesis part, and «going beyond» means that the full hypothesis includes this hypothesis part. To me it seems that natural entities differ from those creatures observing those. Intelligent creatures make notes both by perception and understanding and forms two distinct evidence and hypothesis entities.

6.2 Similarity

T(Q,M,A) also includes the condition of evidence being similar to hypotheses. I use indexes like F_i and G_i to denote that F and G have similar simple and the same logical form. To me it seems that there is no point in confirming or falsifying a hypothesis-claim using an evidence-claim not similar to it. The hypothetical-deductional principle, (Popper,1972) includes the problem of identifying the hypothesis part being false out of a given falsification of the hypothesis. This problem partly includes the problem of correlating evidence with hypotheses. The T(Q,M,A) similarity condition directs evidence to its proper hypothesis, avoiding this part of the hypothetical-deductional problem. However, the correspondence of evidence and hypotheses is not definite. Someone might suggest that evidence F is similar to hypothesis G, someone else might instead suggest that F' is similar to G. One such way of defining similarity for evidence and hypotheses results in one set of test-predicates. That is, one empirical knowledge-gaining program. Using another similarity defined set of test-predicates ends up with some other empirical knowledge-gaining program. These programs cannot compare properly, due to the different conventions or paradigms of similarity. Therefore, such empirical knowledge gaining programs are incommensurable. This is an application of the notion of paradigm. (Kuhn,1970) The consequence for the empirical knowledge gaining programs is that programs are true in their paradigms. A supported hypothesis is true if and

only if it is confirmed in the section 3.6 formal sense, where the involved simple predicates are true in the supported paradigm. There are also other consequences of the notion of similarity. Relative verisimilitude is a relative way of approximating to truth presupposing observational nesting, using the idea that increasing theory truth-contents entails increasing observational success, predictive power. (Newton-Smith,1990) In my analysis, the natural relation of evidence and hypotheses makes an empirical request out of this thesis. This follows by its presupposition of truth-values relative a way of defining test predicates, having its set of suggested evidence and hypothesis similarities. Going on; suggestions that scientific methods are no better than other knowledge gaining methods, (Feyerabend,1975) could be interpreted such that the formulas F and G cannot fulfil the index condition suggested in F_i and G_i.

6.3 Procedure

Evidence being distinct from, and similar to, hypotheses defines evidence matching hypotheses. Logical constructions of matching define the test procedures, as section 3.3 shows. Now, consider the raven paradox. (Hempel,1985) The premise of the paradox is «confirmation of all ravens are black» and paraphrases in my analysis as the formula $T(\forall x(F_i(x) \rightarrow F_j(x)) , \forall y G_i((y) \rightarrow G_j(y)))$, where $F_i(x)$ is the evidence that any individual is raven and $F_j(x)$ is the evidence that any individual is black. $G_i(x)$ is the hypothesis that any individual is raven and $G_j(x)$ is the hypothesis that any individual is black. The conclusion of the paradox is «confirmation of all non-black are non-ravens» and is analogously paraphrased as above T formula, but with the two arguments in logical counter position. (My analysis of confirmation also includes probability, but this is not needed to comment the Hempel paradox. The second part P' of confirmation is not used in the paradox either.) The sections 3.3 and 3.6 show that the two C formulas define as two different test procedures. 3.3 shows that in this case it is not possible to substitute the arguments in T with logically equivalent arguments, due to the non equivalent logical constructions of matching predicates defining the T predicates. Therefore, the raven paradox is explained and avoided in this theory. Going on; the Carnap concept of confirmation includes the relevance concept, (Carnap,1962) involving terminology as positive, negative, and irrelevant confirmation. The T(Q,M,A) makes this concept pointless.

Empirical knowledge-gaining programs T(Q,M,A)_view the test part of confirmation as focusing upon evidence and hypotheses being parts of test procedures. This view is the result of stressing the meaning and relation of evidence and hypotheses. (This might be somewhat related to the Keynes idea of explicating the confirmation evidence part.) One consequence of this ontological view is that it puts new challenging questions to notions like induction, confirmation and falsification. For instance, I see it as Carnap that confirmation involves a classificatory, a comparative, and a quantitative part. With the first part I mean «evidence confirms hypothesis» T(E,H). With the second and third part I mean «this evidence confirms this hypothesis better than that evidence confirms that hypothesis» $T(E,H) \wedge T(E',H') \wedge P'(H|H')$, and «evidence confirms hypothesis to some degree» $P(T(E_j,H_j) \mid T(E_k,H_k))$. However in my account, the classificatory part does neither define the comparative nor the quantitative part. Another consequence example is induction, where there might be more to it than logical and psychological issues. A third effect is to question whether the rationalistic falsification problem can be solved. To me it seems that the traditional clash between empiricists and rationalists do not hold in the aspect of testing, as the empiricist hypothesis-support view and the rational hypothesis-denial rival view seem to converge in the

same principle of testing T.

Finally, in my view empirical knowledge-gaining does not aim at theory choice, but at the cultural use of tested hypotheses. I view T(Q,M,A) as a possible template of the empirical knowledge-gaining method of hypothesis testing, complementing the traditional knowledge-gaining method of theory proof.

7. Conclusion

Empirical knowledge gaining is traditionally viewed as being about confirmation and falsification. Those natural notions, in turn, come with suggested analyses, known to include both insights and paradoxes. I try to show that there is a third terminology about testing available. The essence of this testing terminology is to stress the ontology of confirmation and falsification and to clarify the meaning and relation of evidence and hypotheses. With this settled, the testing terminology tries to wedge itself into the traditional terminologies of confirmation and falsification. With this done, the testing terminology shows the ability to define strong theorems like empirical knowledge gaining programs. These programs are non-inductive and correlative and aim to template some logical aspect of insightful everyday scientific empirical knowledge-gaining in perhaps a less non-paradoxical way.

---

8. References

Carnap, R., *Logical Foundations of Probability* 2nd ed., 1962, University of Chicago Press.

Carnap, R., *Studies in Inductive Logic and Probability* vol. I, 1971, University of Chicago Press.

Feyerabend, P.K., *Against Method*, 1975, London: New Left Books.

Frege, G., *Begriffsschrift*, 1879, Halle: Verlag Louis Nebert.

Goodman, N., *Fact, Fiction, and Forecast*, 1955, Cambridge Mass.

Hempel, C.G., *Epist., Methodology, and Phil. of Science*, 1985, Repr. from Erkenntniss, Vol. 22, Nos. 1, 2 and 3.

Hume, D., *Treatise on Human Nature* ed. by L.A. Selby-Bigge, 1958, Clarendon Press.

Hume, D., *Enquiry Concerning Human Understanding* ed. by L.A. Selby-Bigge, 1927, Oxford.

Keynes, J.M., *A treatise on probability* 2nd ed., 1929, London and New York.

Kuhn, T.S., *The Structure of Scientific Revolutions* 2nd ed., 1970, University of Chicago Press.

Newton-Smith, W.H., *The Rationality of Science*, 1990, Routledge.

Popper, K., *Objective Knowledge*, 1972, London: Oxford University Press.

Mikael Eriksson
Karolinska Institutet
Stockholm
mikael.eriksson@ki.se

THE SORITES Charter

· Version 2.2 ·

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-divulgation, 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>, <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 of 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/> 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) is made available in several formats, but its only official version is the WordPerfect 5.1 document released with filename:

sorite20.wp

Although each version, whether official or not — as initially released on March 31, 2008, 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.