SORITES

σωρίτης

An International Electronic Magazine of Analytical Philosophy
Indexed and Abstracted in ***THE PHILOSOPHER'S INDEX***
ISSN 1135-1349
Legal Deposit Registration: M 14867-1995

Regular-Mail Address:
Prof. Lorenzo Peña
CSIC [Spanish Institute for Advanced Studies]
Department of Theoretical Philosophy
Pinar 25
E-28006 Madrid
Spain
Fax +3491 564 52 52
Voice Tph +3491 411 70 60, ext 18

INTERNET ACCESS:
<**http://www.sorites.org/**>
<**http://www.ifs.csic.es/sorites/**>
Editorial e-mail inbox (esp. for submissions): <**sorites@sorites.org**>
Inquiries and subscription-requests: **<sorites@sorites.org>**

Issue #15 — December 2004

SORITES (ΣΩΡΙΤΗΣ)
ISSN 1135-1349
Issue #15 — December 2004

MAIN INTERNET ACCESS:
<**http://www.sorites.org/**>
<http://www.ifs.csic.es/sorites/>
<**sorites@sorites.org**> (Editorial e-mail inbox, esp. for submissions)
<**sorites@sorites.org**> (Inquiries and subscription-requests)

SORITES

ISSN 1135-1349

SORITES

ISSN 1135-1349

Issue #15 — December 2004

TABLE OF CONTENTS

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp. 3-6
Abstracts of the Papers

ABSTRACTS OF THE PAPERS

Quick thinking? Not so fast!

V. Alan White

Hud Hudson has argued that with a few assumptions one can prove that superluminal objects exist. I argue that even if the assumptions are true that his argument, if sound, leads to a proliferation of movers packing given spaces. I further argue that his argument as it stands cannot in fact entail that objects moving at *any* speed exist.

Homonymous Mistakes with Ontological Aspirations. The Persisting Problem with the Word 'Consciousness'

Rodrigo Becerra

In order to understand consciousness one would benefit from developing a more eclectic intellectual style. Consciousness is, as proposed by almost everyone except the stubborn reductionists, a truly mysterious concept. Its study and dissection merits a multidisciplinary approach. Waving this multidisciplinary flag has positively enlarged the discussion and neurologists, psychiatrists, mathematicians, and so on, have moved to the philosophy of mind arena, first with caution and now with a more powerful voice. Identifying what we mean by consciousness is a first step even when we want to deny its existence. The link between consciousness and some other mental activity (e.g., awareness, memory, executive functioning, etc.) is a logical next step and there is abundant literature doing this, but not all of them differentiate among associated yet different phenomena.

Memetics: An Evolutionary Theory of Cultural Transmission

Asunción Álvarez

In this essay, we introduce memetics, a new theory of cultural transmission based on Darwinian evolutionary theory. A brief account of the two man methodological trends within current memetics — «epidemiological» and «evolutionary» memetics — is given. Memetics differs from other evolutionary accounts of human behaviour in two main points: first, it posits replication as the mechanism for the transmission of «memes», or cultural units; and second, it claims that imitation is the only learning process by which the replication of memes takes place. These premises are discussed, as well as some of the main objections raised against them.

Ontic Vagueness in Microphysics

Silvio Seno Chibeni

This article aims to examine the import of science to the contemporary philosophical debate on ontic vagueness. It is shown, first, that our best theory on the structure of mater, quantum mechanics, clearly ascribes vague properties to objects. This point is explained by both a general theoretical analysis and by some simple examples. The advantage of these examples over that which has been hotly discussed in the literature (Lowe 1994) is underlined. Secondly, it is pointed out that stronger evidence for the existence of vague objects is available through a series of theoretical and experimental results in microphysics, imposing severe constraints on any theory purporting to restore sharpness in the properties of quantum objects.

Roman Suszko and Situational Identity

Charles Sayward

This paper gives a semantical account for the (i)ordinary propositional calculus, enriched with quantifiers binding variables standing for sentences, and with an identity-function with sentences as arguments; (ii)the ordinary theory of quantification applied to the special quantifiers; and (iii)ordinary laws of identity applied to the special function. The account includes some thoughts of Roman Suszko as well as some thoughts of Wittgenstein's *Tractatus*.

David Miller's Defence of Bartley's Pan Critical Rationalism

Armando Cíntora

W. W. Bartley argued that Popper's original theory of rationality (1945) opened itself to a tu quoque argument from the irrationalist and to avoid this Bartley proposed an alternative theory of rationality: pancritical rationalism (PCR). Bartley's characterization of PCR leads, however, to self-referential paradox.

David Miller (1994) outlaws self-reference (and in this way he avoids PCR's paradoxical nature) by distinguishing between positions and statements, Miller's distinction looks, however, suspiciously like an ad hoc manoeuvre or as a stipulation that has to be accepted dogmatically.

Furthermore, Miller's move is inadequate because it is a second world answer (i. e., it involves attitudes or thoughts) to a third world problem, that is, to logical paradox.

It is then argued that given the paradoxical nature of PCR, Popper's old justificationist critical rationalism with its minimum of dogmatism and irrationalism is malgré tout a better option.

On Quine's Arguments Concerning Analyticity

Shaun Baker

In a detailed examination of Quine's Two Dogmas of Empiricism, I argue that Quine fails to make the case that there are no analytical truths in ordinary language. Drawing on admissions he makes with regard to definitions and languages' relationship to pragmatic considerations, and an examination of his arguments concerning the interdefinability of the terms 'synonymous', and 'analytic', I argue that analytic truths exist as deducible consequences of the various uses to which language or sub-languages are put.

Against Compatibilism: Compulsion, Free Agency and Moral Responsibility

William Ferraiolo

Free agency and moral responsibility are incompatible with causal determinism because causal determinism, properly understood, entails that originating conditions beyond the agent's control ultimately compel all human choices and actions. If causal determinism is true, then causal antecedents and laws of nature nomologically necessitate all deliberation, choice and action. If conditions beyond the agent's control ultimately compel the agent's behaviors, then the agent is not free and is not morally responsible. Compatibilists claim that externally compelled acts are not free, but fail to recognize that causally determined acts are, ultimately, externally compelled.

Mad, Martian, but not Mad Martian Pain

Peter Alward

David Lewis attempts to accommodate the possibility of both mad pain and Martian pain by giving a functionalist account of pain *for a population*, and an identity theoretical account of pain *for individual members of a population*. I argue that Lewis's fails because no satisfactory account of the conditions under which a given individual is a member of a given population can be provided.

The Veil of Perception and Contextual Relativism

Dimitris Platchias

In this paper I point out main shortfalls of the three main families of theories of perception and I propose a sort of inferential realism. In addition, I argue that there cannot be a scientific variant of direct realism and illustrate this point with reference to P.F.Strawson's attempt to reconcile, not naïve realism and the scientific variant as he amounts to, but rather, direct and indirect realism. I draw the distinction between four cases of illusion, and I refer to one of these, namely to the case of *veridical illusion*, to show that Strawson's view, put in terms of the Fregean sense-reference distinction, fails. As regards indirect realism, I argue against the representationalist account and the Lockean picture of primary and secondary qualities. Phenomenalism is rejected in terms of the impossibility to identify an object throughout different contexts and I suggest that what is for *x* to be *that x* in different contexts can be given only by a realist analysis of a material object. Finally, I provide an account of

what it is for A to perceive *that x* with respect to different contexts and I conclude with what conditions should veridical perception meet and therefore propose the framework of a new theory of perception.

Johnston on Fission

Brian Garrett

In this discussion paper, I evaluate some arguments of Mark Johnston's which appear in his articles «Fission and the Facts» (1989) and «Reasons and Reductionism» (1992). My primary concern is with his description of fission cases, and his assessment of the implications of such cases for value theory. In particular, Johnston advances the following three claims:

(1) Rejecting the intrinsicness of identity is an arbitrary response to the paradox of fission;

(2) Fission cases involve indeterminate identity;

(3) Contra Parfit, fission cases have no implications for value theory in the actual world.

I argue that (1) and (2) are false, and that (3), if true, is not true for any reason that Johnston gives.

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp. 7-10
Quick thinking? Not so fast!

QUICK THINKING? NOT SO FAST!

V. Alan White

In 'Moving faster than light' Hud Hudson [2002] argues that by employing simple reasoning with a few explicit metaphysical assumptions, one can demonstrate that, contrary to accepted physics, there must be objects that move at superluminal velocities. Though there is without doubt some very quick thinking on Hudson's part that is more than a little reminiscent of Zeno's, I will show that Hudson's argument no more requires anything in the world go at dazzling speed than Zeno's arguments stood the world still.

Hudson's argument is by way of the metaphysical construction of a supposedly material object dubbed 'Quick'. Quick lives up to its name by superluminally traversing a spatial distance of just over two-billionths of a light-second (about two feet) in the following way. The distance constitutes the height of a 3-D object 'Cone'[1] (one dimension is an hour's temporal duration), which itself is comprised of a nondenumerable cross-section stack of 2-D space-slice discs that taken together make up the so-called 'Disc Set'. These discs are ordered in relation to the temporal extension of Cone so that they can be matched with appropriate instantaneous moments in a familiar asymmetric temporal manner. Hudson then posits an interval coincident with that of Cone named 'T' which is just over one-billionths of a second in duration. Then Hudson set-theoretically maps instantaneous moments of T — the 'T Set' — with members of the Disc Set in the aforementioned well-ordered manner so that the result mirrors time traversal throughout T corresponding with space movement across Cone. The result of this mapping generates members of another set — the 'Quick Set' — the fusion of which produces the aforementioned material object 'Quick'. Since Quick is a moving object (composed of spatiotemporal parts), and Quick traverses the length of Cone in just over one-billionth of a second, Quick moves at twice the speed of light [2002, pp. 203-204].

Hudson admits throughout this argument to a number of controversial metaphysical assumptions: that at least one *n*-D object may have (*n*-1)-D cross-sectional spatial parts, that any extended object has spatiotemporal parts, that fusions of the members of well-ordered spatiotemporal-parts-sets (such as the Quick Set) result in whole objects. He maintains,

---

[1] I note that this distance is only an intelligible quantity in the context of a background space/spacetime within which Cone exists. Thus there may be hidden assumptions about such a space/spacetime that further complicate Hudson's argument, but I will try to avoid these issues in my analysis.

however, that these are not incredible propositions, and thus that superluminal motion is not merely possible, but a fact (if the assumptions are true) [2002, pp. 203-204].[2]

However controversial these assumptions are, and even if they are true as Hudson allows, I will show that Hudson's argument, when fully explicated, entails consequences that are counterintuitive if not outright false. Further, under a (perhaps most) plausible interpretation of what fusions of spatiotemporal-parts-sets are, I will argue that this form of argument in general *cannot entail* that Quick and its ilk exist.

To see the full scope of Hudson's reasoning, review the following variation on it. Let Cone and the Disc Set remain as before. Say, however, that T* is an interval one *trillion* times longer than T — around 17 minutes duration, but still within Cone's time of existence. (Since Hudson places no constraint on the selection of a time interval within Cone's existence parameters, it would appear that T*, and any such interval less than or equal to Cone's temporal duration, can be used in the following way.) Now map T*'s instantaneous-moment set to the Disc Set as T's was before. The result is (what I must call) the 'Tortoise Set' — though the resultant fusion-object 'Tortoise' would lose a race to all able-bodied tortoises, I suspect. So Tortoise is much, much slower than Quick, taking nearly 17 minutes to traverse the same space — yet Tortoise is constructed in the same kind of way. Clearly then, within the two-billionths-light-second distance and one-hour time limits of Cone's existence, one can construct a very large number of movers (and even an infinite number, given that temporal intervals within Cone's hour of existence are subsets of the nondenumerable set of Cone's instantaneous moments), bounded on one extreme by asymptotically near-instantaneous movers (like Quick, but even quicker and quicker and …) and on the other by the slowest hour-crawler (like Tortoise, but taking the full hour to move the distance). Therefore, Hudson's revised point should be that *material objects that move at every speed possible within any given well-ordered spacetime volume can be proven to exist* (again, with his admitted assumptions).

Given the intuitively vast array of Cone-like spacetime volumes where such metaphysical construction might occur, Hudson's argument, when applied to the universe at large, indeed yields quite a prodigious progeny of moving objects (certainly enough to induce the infamous Lewisian stare).

Further, since distinct time intervals of any given Cone-like object can overlap — T and T* might well overlap by T's duration, for example — then resultant fusion-objects of time-and-space sets such as 'Quick set' and 'Tortoise Set' themselves are in such cases *coincidently* existing and moving things.[3] While it is quite understandable that real objects might overlap in this way — consider lightspeed cosmic rays that might happen to coincide with my movement with the Earth by flashing through me within a certain interval of terrestrial motion — such an overlap is typically interpreted as an accident of nature, *not a metaphysical*

---

[2] Afterward Hudson expands this argument by quantizing the space and time mappings into four dimensions with 'ThickQuick' [2002, pp. 204-205], though this expansion in no way dispenses with a reliance upon the essentials of his earlier 'Quick' argument (or its accompanying deficiencies).

[3] And apart from an additional assumption that Hudson's time-space mappings must not overlap spacetime intervals distinct from Cone-e.g., as would 'Coneahead', which contains Cone but includes an additional anterior spacetime of about .4 inches and another minute's duration (to preserve scale with Cone)-the overpopulation problem here worsens! Such an assumption, however, seems entirely arbitrary, and in fact contrary to intuitions about spacetime continuity.

requirement of existence. Thus every one of the magnificent offspring that result from fertile argument, fast and slow alike, always exist in the company of a staggering assemblage of fellow travelers, all of which exist merely in virtue of sharing their various times with the same space. While this is certainly not logically impossible, filling spaces with a plenum of such movers is *at best* an excess of ontological exuberance.

But there is an additional sense in which Hudson's argument is, as it were, unsafe at any speed. Clearly the above extension of the argument does not constitute a *reductio* or anything of the sort. Given assumptions and perspective, it only serves to populate the universe with objects *that move at all logically possible speeds* (if one additionally assumes that there exists at least one infinite spacetime Cone-like stretch, to allow for infinite sloth in that case). *But what objects?* Hudson says, as he does of Quick, that these are *material* objects [2002, p. 204]. But why should anyone accept this claim? After all, even given a fusion of the members of the 'Quick' Hudson's and and likewise sets, these are still *sets*, the members of which are the result of *logical* mapping of *arbitrarily selected* dense time-lapses to equivalently dense space-extensions (it is this Cantorean equipotence of all such space and time sets that gets Hudson' 'Tortoise's argument off and running, so to speak). As such, I would argue that fusions of the members of these sets constitute at best *logically possible* objects in spacetime-perdurant objects that must, *if they exist*, move at such speeds.[4] A recognizable example of such a possible object construction is that of Santa Claus delivering gifts to the world's children in one night. Sophisticated versions of this scenario estimate that this would require a Santa-sleigh-velocity of many thousands of miles per second. What is done here is to assign Santa's work to one night, map that time to the needed distance of travel, and voila! — a logically possible Santa. Presumably, Hudson would not seriously want to argue that a speedy temporal-parts Santa really exists. On the other hand, nothing in Hudson's form of argument *prevents* one from concluding that the Jolly Old Elf, conceived as such a moving object, exists (largely due to Hudson's failure to address identity questions about fusion objects — see note 4). That alone suggests a laxity within Hudson's argument that undermines the force of his conclusion that superluminally moving objects *must* exist, since they are on a metaphysical par with a speedy but presumably mythical Santa.

Ultimately, the ingenuity of — and the problem for — Hudson's argument is that it is Zeno turned inside out. Instead of producing paradoxes of motion over distance and time, Hudson uses (or, as I argue, *should* have used) the continua of distance and time to produce a surfeit of movers at all possible rates of motion relative to a space traversed. The fact that Hudson restricts his argument to proving that only superluminal movers exist indicates that either he is unwilling to acknowledge, or unaware of, the breathtaking existential commitment of its complete scope should he be right. If, however, as argued above, Hudson's fusion-object movers constitute only *possible objects*, then whether such objects are existentially instantiated

[4] Though I won't pursue it here, I should register my protest against Hudson's assumption that Quick and its relatives (including 'Tortoise') are properly objects at all, since they are fusions of nothing more than times correlated to spaces apart from any properties instantiated within these times and spaces that could satisfy trans-spatiotemporal identity conditions of what counts as the same object. At best I would call the result of fusion of the Quick, Tortoise, etc. sets abstract objects. Still, for simplicity's sake I will yield to Hudson to the extent of calling such 'bare' spatiotemporal-parts fusion-objects logically possible ones.

is in fact a further question beyond any fanciful entertaining that they do.[5] Hudson's Zenophilic argument is not sufficient warrant in itself to conclude that any movers, super- or subluminal, exist.

References

Hudson, Hud. 2002. «Moving faster than light». *Analysis* 62: 203-205.

V. Alan White

University of Wisconsin Manitowoc

awhite@uwc.edu | awhite@lsol.net

[5] In this way my criticism reflects something of the same criticism Aristotle leveled at Zeno-namely that the latter's arguments conflate questions of potentiality and actuality.

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp 11-23
Homonymous Mistakes with Ontological Aspirations. The Persisting Problem with the Word «Consciousness»

HOMONYMOUS MISTAKES WITH ONTOLOGICAL ASPIRATIONS. THE PERSISTING PROBLEM WITH THE WORD «CONSCIOUSNESS»

Rodrigo Becerra

In order to understand consciousness one would benefit from developing a more eclectic intellectual style. Consciousness is, as proposed by almost everyone except the stubborn reductionists, a truly mysterious concept. Its study and dissection merits a multidisciplinary approach. Waving this multidisciplinary flag has positively enlarged the discussion and neurologists, psychiatrists, mathematicians, and so on, have moved to the philosophy of mind arena, first with caution and now with a more powerful voice. This is, to be sure, a welcome phenomenon for several reasons. First, it reminds us of those old timers who used to be all in one (mathematician-philosopher-political commentators, etc.) and their natural preoccupation with *grand themes* rather than with specific disciplinary boundaries. It demands from philosophers a sharper rigour when dealing with specific scientific sub-areas. It demands from scientists a more global picture and intellectual projection of their specific findings. Collectively taken, these processes will advance the study of consciousness. However, the negative side of this exercise is that this collaboration may lead to an overinclusion which at times might mislead the direction and make us «feel as if» we are getting closer to solve the mystery, the hard problem, but in fact we might be sometimes getting away from it. Common sense also warns us against this intrusion by advice epitomized in sayings such as «can't see the wood from the trees». We are all familiar with the story of the drunk who is looking for something under the lamppost and someone asks him, «what are you looking for? My keys. Where did you drop them? Back in the alley. Why are you looking for them here then? Because there is more light here».

The principle that guides this essay is that the study of consciousness requires a multidisciplinary approach because consciousness is, in all probability, a multifactorial phenomenon. However, once this honeymoon period of disciplines is over, each discipline will need to generate *relevance* if we are to continue exploring the same target. The word consciousness is used with a variety of connotations and it means something different to many. However, once agreed on the use of one connotation, the explorative research should adhere to this, if not universal, at least operational connotation. The exhaustive analysis of this phenomenon might force us to rethink the definition. This would be a healthy outcome. However, a clear identification of the target is still necessary in the beginning. Many publications deal with related but not relevant «consciousness» terms and this is dangerous because it is adding related but irrelevant information that is counterproductive to the study of «consciousness».

A classic illustration is the paper published by *PSYCHE*: «The Decoupling of «Explicit» and «Implicit» Processing in Neuropsychological Disorders. Insights Into the Neural Basis of Consciousness?» by Faulkner and Foster (2001). This is an informative summary of explicit and explicit cognitive processes but its impact on the consciousness debate is minimal at best and very likely to be counterproductive because of its philosophically ill informed background. The merit of this paper is simple; it is a good catalogue of neuropsychological syndromes that involve neglect. It clearly shows that there is a discrepancy between what «neglect» patients report and their true preserved, albeit implicit, information in the impaired domain; this discrepancy applies to a series of neuropsychological disorders. A brief summary of this paper is pertinent here.

In neuropsychology, «neglect» refers to disorders characterised by a lack of awareness of (usually) half of the presented information. For example, patients suffering «unilateral neglect» show impaired processing of information presented on the side that is contralateral to the location of the brain insult. Studies however, suggest that some processing, occurring on the presumably neglected area does take place implicitly. Therefore, the patient is not overtly aware of this. The evidence presented by the authors comes from studies investigating blindsight, amnesia, object agnosia, prosopagnosia, hemi-neglect, and aphasia. Their analysis, they claim, will have repercussions on «a) possible clinical therapies in brain-injured patients as well as b) the architecture of cognition and c) the neural basis of consciousness in non brain-damaged individuals» (p.1). A crucial distinction offered by the authors is about their understanding of «consciousness». They state that, although they acknowledge existing distinctions of the word consciousness, they prefer to refer in their paper, to «explicit knowledge» as relating to those aspects of cognition which the individual has access to. Conversely, implicit processing refers to cognitive processes, which the individual does not have access to. Thus, explicit memory for example can be tested by asking the individual to report on the content of a specific memory, whereas implicit memory can be detected by behavioural changes due to the influence of previously «unaware» acquired information. I will briefly explain the phenomenon of blindsight to illustrate the authors' strategy.

Blindsight occurs when a brain insult causes a loss of vision, as reported by the patient, but there is preserved implicit processing based on information presented in the «blind» area. There is ample experimental evidence demonstrating that patients alter their behaviour in tasks influenced by implicitly gained information. This adjustment of behaviour is beyond what chance could bring about and it appears to be directly influenced by the relevant stimulus, with the patient remaining unaware. An important figure in the introduction of this concept is Weiskrantz and colleagues (1974) who described a case study in which the patient developed left hemianopia (blindness in the left visual field of both eyes) after surgical removal of a significant portion of the striate cortex in the right hemisphere. After the patient was asked to visually discriminate between stimuli presented in the blind field, the patient would report that he could not do so. However, when asked to guess, the answer he gave would be strikingly accurate. He was asked to point the location of stimuli, orientation of different lines (vertical versus horizontal), and shapes (cross versus circle). Other studies have found other types of visual discrimination equally accurate in hemianopia patients; for example, reaching and grasping, discerning meaning of words, varied shapes and so on (in Faulkner and Foster 2002).

These findings are analogous to those found in memory research whereby a brain insult can damage the ability to recall new information (anterograde amnesia) and the ability to

recall information prior to the insult (retrograde amnesia). There appears to be ample evidence suggesting the existence of preserved implicit memory. A traditional experimental method uses amnesic and normal subjects who would be presented with word stems for example and be required to complete the stem with the first word that comes to mind. Both controls and amnesic subjects were more likely to complete the stems with words from a list that had been previously studied of which amnesic subjects had not explicit recollection. In an experiment of this nature, the priming effect consists in the fact that amnesic subjects, although scoring poorly on a simple recall test (explicit mode), still showed an influential effect (implicit mode). Similar findings are reported from the visual agnosia area. In visual agnosia patients fail to identify familiar objects by sight even though there is no physical damage to the visual apparatus. However, they can identify objects if presented in a different sensory modality (e.g., touch, hearing). The same phenomenon has been established in Prosopagnosia (impairment of facial recognition) whereby patients have reported an inability to recognise familiar faces. However, they have significantly different (larger) skin conductance responses as compared to their reaction when exposed to unfamiliar faces. In the language domain, aphasic patients (expressive and/or receptive language impairment) have also demonstrated preserved implicit language comprehension and expression in spite of impairment. Further examples of implicit processing come from the neglect area, which is an attentional deficit of (generally) left visuo-spatial stimuli, normally as a consequence of right hemisphere damage. This hemineglect has been observed in behaviours such as dressing only the right part of the body, eating the food from the right side on the plate only, omitting words placed in the left visual field when requested to read a paragraph and so on. However, preserved processing on the neglected space has been reported. This has been usually investigated using preferential choices in the absence of explicit recognition of the stimulus presented in the neglected space. That is, on subsequent stimuli presentations (hemineglect) subjects would still choose a stimuli (e.g., an intact house) as opposed to a less preferred option (e.g., a house with the left part of it in flames) even when the portion representing adverse stimuli was in the neglected field.

Faulkner and Foster's paper was published in *Psyche,* (a very eclectic and multidisciplinary journal) which deals with the mystery of consciousness and its various contributors are supposed to approach the topic in a pluralistic fashion. However, it appears they are making the topic flexible enough to suit every single connotation of the word consciousness — Faulkner and Foster's paper is an example. The discussion on consciousness is plagued with these sorts of nominal accidents. One wonders if the question asked in the title (unanswered in the paper) refers to the type of question with which this journal is concerned. The authors of the paper deliberately equated consciousness with «awareness». From that moment on, everything is permitted and the relevance to the consciousness domain is avoided. Terms such as explicit, implicit, unawareness, (Freudian)unconsciousness, covert/overt, belong to a family of concepts that denote primarily cognitive «awareness». As briefly described above, all the neuropsychological disorders mentioned are analysed in the light of awareness or accessibility from the patient's point of view. This awareness represents a challenge of its own accord but it is not necessarily equated with the subjective experience of phenomenal character and the potential link between awareness and consciousness (phenomenal experience) is not even mentioned. The literature on awareness is vast and cognitive psychology has embarked on a prolific study. For the last 20 years it has covered varied areas such as implicit memory, with Graf and Schafter, (1985) being some of the pioneers in revitalising interest; skill acquisition processes (controlled) leading to «automatic» (that is, lacking awareness) processes, work which started importantly with Schneider and Schiffrin (1977); and the

learning domain which was pioneered by Reber (1967). The terminology used in these areas moves freely between, «implicit learning», «unconscious learning», and «unaware learning». The same applies to memory, and skill acquisition. These are just some of the domains cognitive psychology has ventured into exploring the notion in which the subject reports being «aware» (of something recalled, learnt, or practiced).[1]

More contemporary work in this area has developed and applied to abnormal manifestations of information processing. For example MacLeod's work[2] has focused on implicit selective information processing of depressive and anxious patients, and drawing from earlier work by Beck (1967) has extensively investigated the idiosyncratic selective cognition of this population. MacLeod focuses on the cognitive biases of individuals suffering from affective disorders and suggests that these cognitive processes are outside the realm of the individual's awareness. Thus, the thought content that this population engage in is systematically distorted by cognitive biases that take place «automatically», and outside the realm of unawareness. Perhaps the best known proposal about the unconscious is that offered by the psychoanalytical tradition initiated by Freud and his invention of the concept of «unconscious». Given that it is a particularly familiar proposal I will therefore not examine it here. Suffice to say that the characteristic feature of Freud's unconscious is exactly the contention that is a whole set of memories, desires, fixations, etc. that are not accessible to the individual, that is, that the individual is not aware of them (e.g., a childhood experience/trauma).

Faulkner and Foster's review of implicit processes in neuropsychological disorders resonates entirely with the domain investigated in fields like learning, memory, skills acquisition, emotional pathology, and the Freudian tradition. However, these areas do not *necessarily* refer to consciousness, that is subjective experiences, or phenomenal experience, (qualia) or other more precise descriptors that have been used to refer to consciousness. Their work fit nicely within the psychological tradition, which explains behaviour but does not examine «consciousness». It does, fit, as put by Chalmers (1996), within the characteristic concern of psychology, that is, mind as the internal basis of behaviour. It follows the tradition of explaining «how» a mental process causes behaviour, not «what» a mental state «feels like». I have purposefully used the word awareness when describing the content of their paper but the authors used interchangeably «unconscious», «unaware», and «implicit» terms. This reflects the lack of familiarity with the philosophical inquiries into consciousness, which is not a feature restricted to Faulkner and Foster. But one wonders if by publishing in a journal like *Psyche* this confusion might not be perpetuated. One wonders also if mistakes they generate such as the use of expressions like «conscious awareness» (used by others too) are innocent tautologies (in that they incessantly used consciousness as a synonym of awareness) or whether it reflects a deeper confusion.

---

[1] A good review is provided by Kirsner et al (Eds., 1998) in their «Implicit and Explicit Mental Processes». Another interesting recent review is «Out of mind. Varieties of unconscious processes» (Gelder, De Haan and Heywood, 2001).

[2] MacLeod, 1992; MacLeod, Mathews and Tata, 1986; MacLeod and Rutherford, 1992, MacLeod and Hagan, 1992, MacLeod and Mathews 1991; MacLeod and McLaughlin, 1994; MacLeod and Lawrence Cohen, 1993 and MacLeod, Rutherford, Campbell, Ebsworthy and Holker, 2002.

A similar criticism can be applied to Schiff and Plum's attempt to shed light on normal consciousness (Schiff and Plum, 2000). An underlying assumption of their article is that the understanding of the pathology of human consciousness is a first step in understanding mechanisms underlying human consciousness. The paper aims to «(1) explore the neurology of impaired consciousness and detail a brief taxonomy of global disorders of consciousness, (2) place these neurological diseases in the context of the underlying anatomy and physiology of arousal and 'gating' systems, (3) examine the role of the gating systems in fluctuations of cognitive function and recovery from states of impaired consciousness, and (4) consider the possible contributions of these clinically-rooted approaches to further understanding of human consciousness» (p.1). The authors' expectations are clear; understanding what human consciousness is via studying disorders of human consciousness is a promising project. The definition of consciousness used by Schiff and Plum is in a way more confusing than that used by Faulkner and Foster in that from the outset, this paper appears to center on *wakefulness* more than anything else. They offer, as their basic definition of consciousness, the one given by James (1894 in Schiff and Plum, 2000), which appears to draw attention to «awareness» (of the individual and his/her environment).

> At its least, normal human consciousness consists of a serially time-ordered, organized, restricted and reflective awareness of self and the environment. Moreover, it is an experience of graded complexity and quantity (p.1)

The detailed neurological account offered by the authors is beyond the scope of this essay. Of considerable however interest is their understanding of consciousness. What follows is a brief description of the first part of their essay, which is of more interest to this paper in that it specifies the «type» of consciousness to which they refer. The taxonomy favoured by the authors involves a classification of global disorders of consciousness that include: stupor and coma, the vegetative state, akinetic mutism, absence and partial complex seizures, delirium, dementia and hyperkinetic mutism.

Coma is an unarousable state characterised by unresponsiveness to internal or external stimuli and a loss of all neuropsychological aspects of normal functioning. In its observable behaviour, it resembles a deep, sleep-like unconsciousness. Coma contrasts with stupor in that stupor refers to impairment of arousal but some responses (although inconsistent) to the environment can still be detected in stupor. The persistent vegetative state (PVS) differs from coma in that in PVS cyclic arousal recovers but there is still no evidence that the patient is aware of his/her environment. The overall cerebral metabolism in PVS, as revealed by positron emission tomography (PET), is reduced by 50% (similar to those patients undergoing deep surgical anesthesia). Some behavioural and physiological «activity» has been reported in PVS patients. For example, one patient of one of the article's co-authors expressed occasional single (understandable) words. The patient's PET investigation revealed isolated activity of the left cerebral structure, which was operating at a very low metabolic rate. However, it was still prominently activated as compared to the rest of the patient's brain. Another disorder included in their analysis is akinetic mutism, which refers to a state in which patients appear vigilant and attentive but remain motionless and have a profound impairment of neuropsychological functioning. Neuropathological analysis reveals the involvement of the frontal lobes, either directly or indirectly. An opposite condition is hyperkinetic mutism characterised by unrestrained but coordinated motor activity with no apparent awareness of self or environment. This rare condition appears to involve bilateral destruction of temporal parietal occipital junctions and wider lesions compromising occipital-parietal regions. The authors also postulate seizure disorders as part of consciousness disorders. In both absence

seizures and complex seizures, patients develop momentary vegetative-like states with attentional and intentional failures and loss of working memory and intra-ictal perceptual dissociation. Other less pervasive conditions (in that some functioning is preserved) are Delirium and Dementia. Schiff and Plum state that delirium's main characteristic is a temporal disorientation whereas dementia's main feature is at first a gradually increasing memory dysfunction reaching a stage which might be difficult to differentiate from a vegetative state.

The term 'consciousness' used in their paper is clearly related to a family of terms, which resonate with «sustained wakefulness», «awakening», «arousability» and so on. The fact that we are «awake» is uninteresting and interesting at the same time. It is intellectually interesting in the sense that to reach a level of knowledge that allows us to know the neurological substrates implicated in arousability and eventually in sustained wakefulness will solve one of the mysteries haunting the study of biological entities. This knowledge is fascinating; much needed and will, in all likelihood, have consequences in many other areas of the functioning of the human brain. However, it is uninteresting when one wants to examine a particular aspect of the human mind, that is, consciousness, and to focus on its basic nature. The state of being awake is shared with many other species and appears to be related to consciousness in that «most» consciousness appears to take place when one is awake (though this is highly debatable as well, for example, in pain sensations while one is asleep). But they are still two different phenomena. Breathing, for example, is in a sense even more necessary for consciousness for if we were not breathing consciousness would not be possible. However to include studies on breathing (a function which is also shared with all other living species) in a debate about «consciousness» would appear intuitively overinclusive and intellectually interesting but strictly speaking irrelevant.

Because the word consciousness can be used to describe wakefulness as well as a phenomenal experience, it does not seem to be powerful enough to be included in the same debate (or it shouldn't anyway). To do this, creates the sense that we are getting closer when in fact it is adding unnecessary distractions. The study on wakefulness (as the study on breathing) does not appear to shed much light on the subjective phenomenal experience, which one would think, is the main target interest for the «consciousness community». And if it does, the link is consistently missing. Different neurological disorders include different levels of cyclical arousal and different levels of awareness of the self and the environment. For example, coma epitomises unresponsiveness to the environment whereas PVS appears more complex in that some responsiveness has been reported. Similarly, absence seizures involve a momentary loss of attentional and mnestic faculties. All these disorders create a massive dissociation within the individual and with his/her environment, momentarily or permanently, which makes the researcher's access to the patient's neuropsychological functioning impossible. Therefore, we can only assume, and rightly so in all probability, that their cognitive abilities are impaired. Needless to say, their subjective experience of colour or pain, etc. is equally inaccessible and the researcher assumes therefore equally impaired. Is this assumption warranted? Is it the same in Delirium as in PVS or coma? Do we have enough information for this conclusion? The authors conduct neuropathological analyses via neuroimaging techniques, which reveal the organic lesion site or low metabolic activity, thus giving a good account of the putative neurological substrates subserving wakefulness, and there is where they have to stop. It appears as if the study of consciousness as such (as in the philosophical interest) requires a different level of analysis.

This present paper does not favour discarding studies of implicit/explicit processes or the neurology of impaired wakefulness. On the contrary the examination of these domains would

greatly enhance the study of consciousness. However, first we need to agree on what we are studying. For example if we decide that consciousness refers to a subjective experience which is characterised by intentionality and volition (this is an example only) then a re-examination of implicit /explicit processing in the light of volition and/or intentionality would generate interesting and *relevant* information. The same analysis can be re-considered with neurological disorders affecting these constituent aspects of consciousness. For example, are there neurological disorders that affect intentionality or volition? Perhaps some do and some do not. Here the taxonomy proposed by Schiff and Plum would need to be readjusted according to which aspect of consciousness is being employed.

The entanglement into which philosophy has driven itself when exploring consciousness has been unnecessarily complicated by lack of agreement on terminology. It would be pretentious and ambitious, for example, to re-baptise Freudian unconscious as a theory of awareness (the implicit). However, something like this is needed. There is a clear distinction between the Freudian unconscious and the state of being awake.

Phenomenal consciousness, access consciousness, content of consciousness, double consciousness, consciousness proper, real consciousness, transitive consciousness, intransitive consciousness, self-consciousness, creature consciousness, core consciousness, peripheral consciousness, primary consciousness, secondary consciousness; Please! To understand and follow the technical study of these various aspects of consciousness has become an exercise on its own. Do these «different types of consciousness» allude to different ontological entities? Are they overlapping, partly overlapping, or the same? What a mess. Paraphrasing Searle's observations regarding the criticisms of his basic philosophical «default» positions, it is as if in order to keep business running, some philosophers have created an artificial language which eventually becomes an independent subject of study with little to do with «consciousness». These terms seem to have departed from the notion of consciousness so violently that there seems to be something called consciousness and something totally different from «the study of consciousness». A preliminary step to any analytical definition is to establish identity criteria and this is the step that is missing and keeps messing up the landscape so badly. Consciousness is intrinsically complex and does not need extra, unnecessary, yet avoidable complications.

There seems to be a general agreement on the fact that there is such a mess in the literature. Many authors acknowledge this in their introductory paragraphs as they proceed to explain their own theories on consciousness. However, no serious attempts are made to solve this and we end up with yet more «surnames» for consciousness. A notable paper on this is that of Block (1995) who carefully points out the existence of this nominal problem. I have chosen to present a brief summary of Block's paper here as it illustrates the main propositions of the present essay. However, Block's solution tends to obscure the discussion rather than solving it. His criticism of the «messy» state of affairs in the study of consciousness is sharp and precise, but when attempting to solve it, he generates a new nomenclature, namely phenomenal consciousness and access consciousness. I will use his proposition to illustrate the fact that he could have used the word «awareness» to distinguish access to information, instead of piling more names on already cluttered consciousness terminology. The tenet of Block's proposition is that there are different notions of consciousness and this has led to numerous confusions about its analysis. He uses the phenomenon Blindsight to illustrate the existence of a fallacy that characterises this field. Blindsight has led many authors to postulate that one of the functions of consciousness is to harness or to enable information represented

in the brain to guide action. This fallacy, Block claims, is revealed when one uses a more pertinent distinction in the consciousness area, that is, Phenomenal consciousness (P-C) and Access consciousness (A-C).

Phenomenal consciousness

Block's argument uses Schacter's model to suggest that P- consciousness is

a) experience, that is, what makes a state phenomenally conscious is that there is something it is like to be in that state (as proposed by Nagel, 1974, in Block, 1995). And

b) P-consciousness has a function, which is to act as a gateway between the Knowledge modules and the Executive system module, which is in charge of reasoning, reporting and guiding action. The phenomenal consciousness module integrates the outputs of the specialized modules and transmits the integrated content to mechanisms of reasoning and control of action and reporting. (Notice that eliminating the phenomenal consciousness module would raise the possibility of the existence of zombies). The basic concepts suggested by this model is that a) C is single system. That is, phenomenal consciousness is *different* to cognitive processes. Note that this conceptualisation would immediately exclude Faulkner and Foster's catalogue of neuropsychological syndromes of neglect; and b) C has a function, which is helping/facilitating access to the executive system. This is in sharp contrast to other views of C, which *identify* C with information processing or postulate that C is *correlated* with that information processing function. Disambiguating between cognitive processes and consciousness has been elegantly explained by Chalmers (1996).

Using the phenomenon Blindsight (see previous section), Block postulates that a module specialising in visuo-spatial information has information about verticality of stimulus. A lesion has damaged a pathway between this specialized module and the phenomenal conscious system creating the «blind spot». The consequence is that the patient has no phenomenally consciousness experience of the line. And hence his executive system has no information about the whether the line is horizontal or vertical. The specialized module has a direct connection to the response system therefore when the subject is given a forced binary choice, the specialized module can somehow directly affect the response. It has been postulated that if the subject is given longer exposure (studies using alexics) their guesses are inaccurate. The explanation would be that the executive system takes over, and it prevents the peripheral system from controlling the response.

Many theories dealing with consciousness fail to address the question about P consciousness. For example, Crick and Koch (1990 in Block, 1995) claim to explain the mystery of C but their physiological examinations and propositions probably deal with the binding problem more than P consciousness (35-75 Hertz hypothesis). Many language/thought cognitive and neuropsychological experiments are addressing cognitive phenomena with no theoretical perspective of what P consciousness is. This is a powerful insight on Block's part and the basic argument of his observation of conflations. Block argues for example that Baars's theory «global worspace» is presented as a theory of P-consciousness but it is clearly a theory of A-consciousness. The same thing applies to Mandler's theory. Similarly, Schallice states that his theory of consciousness is about P-C, but it is an information processing theory of A-C. Edelman's proposition is about A-C and self-consciousness not P-C. Kosslyn and Koening's theory suffers the same problem, although it claims to be a theory of P-C, their theory is about A-C and monitoring consciousness. Kihlstrom postulates that the phenomena

of implicit perception are about P-C. However, they claim that self-consciousness is what is lacking in implicit perception. They state that events come into consciousness when there is contact between the representation of the event (fact node) and the representation of oneself (self node). Blocks counterargument is that there are creatures (e.g., babies or other species) that have P-consciousness but have no «self-node». Andrade identifies C with the «executive systems» (systems that coordinate lower-level information processing). Here again they are conflating P and A-C. Jacoby and colleagues, examine experiences in which there is perception without «subjective experience» (subliminal perception). Therefore, their work focuses on A-C but they assimilate P and A-C. Schacter postulates that his work focuses on P-C but he frequently reverts to A-C when examining neuropsychological defects (e.g., anosognosia). Dennett postulates that consciousness is a cultural construction. For Block, A-C is as close as we get to the official view of Dennett's C, but closer comparisons are difficult as Dennett's view is that C is not a real ontologically independent phenomenon. Searle normally uses the notion of degrees of consciousness, but when he does so he ends up admitting somewhat Block's distinction of P and A-C.

There are quite a few observations about Block's strategy, but to review them here would be beyond the scope of this essay. They are well examined by several commentators in the replies to his target article. A salient criticism is his lack of evidence when hypothesising the existence of Access Consciousness (see for example Baars, 1995). My main observation, as I wish to stick to the nominal arena, is that there is nothing stopping Block from using «awareness» instead of «access consciousness». Why use «consciousness» to signify something different? There is something, a phenomenon, or set of phenomena that is encapsulated by the word «awareness». There is nothing controversial about this; however, the controversy arises when one is invited to accept not only his highlighting of the confusion but also the interaction between his putative P and A-C. The relationship of awareness and consciousness is another matter and it deserves much examination. However, since we know that they refer to something different let us start by using different terms to distinguish them. Block's initial concern about the conflation of P and A-C is unreservedly supported in this essay. The problem comes with his alternative proposition. To attribute some sort of consciousness to blindsighters is potentially confusing, as attributing some sort of consciousness to zombies would be equally misleading.

Chalmers's approach when defining consciousness appears to be less contaminated (1996). He speaks of the error of identifying the psychological with the phenomenal. He acknowledges the confusing state of affairs regarding its definition but favours remaining closer to the phenomenal aspect of consciousness. His characterization of consciousness involves the essential subjective experience, and he uses a set of terms that resonate with consciousness which include: experience, qualia, phenomenology, phenomenal, subjective experience, «what is like». The analytical approach to consciousness, and the other family of terms, will not be embarked upon here.

With this background in mind, one could use the terminology presented thus far and build a clear distinction of family of terms:

Wakefulness	**Consciousness**	**Awareness**
Sustained wakefulness	Qualia	Implicit
Awakening	Phenomenology	(Freudian) unconscious
Arousability	Phenomenal	Covert
	Subjective experience	Automatic
	«What is it like»	

Wakefulness, Consciousness and Awareness as different terms

Notice that this classification responds to a necessity to «identify» phenomena in the consciousness area. The «analysis» of these phenomena is something different and was not attempted in this essay. How does wakefulness relate to consciousness? How do implicit information processes relate to consciousness? and so on, are meritorious analyses and needed for further clarification. However, to use these families as one entity is to start off on the wrong foot, and a considerable source of confusion. This sort of distinction would allow to clearly conceptualise issues such as the fact one can be not awake and have consciousness (e.g., sensation of pain while sleeping), one can be awake and process information and not aware, therefore devoid of the experience (e.g., blindsight), and so on.

Where does this come from? At a content level analysis this is not a confusion, it is a deliberate attempt to, on one hand accept the existence of consciousness as experience (thus rejecting accusations of radical eliminitivism) while keeping the methods and ideas within the known framework of empirical science (Varela, 1995). This comfort zone makes cognitive scientists reluctant to value and explore areas more difficult to operationalise (e.g., memory versus the feeling of red). Some blame Descartes and perhaps rightly so. Behaviourism is another offender. If Behaviourism admitted to any mental life, it was deemed irrelevant in understanding behaviour. The less dogmatic behaviorists referred to mind exclusively in psychological processes; they would not even contemplate the phenomenal aspect of mind. In addition, the advent of cognitive science and their entire focus on the psychological aspect of mind as ultimately responsible for behaviour is probably a significant influence in this prevailing attitude. At a more generalized level, and perhaps significantly more influential because of this, the *common* use of the word consciousness (with its numerous derivative connotations) has been unduly reproduced in the expert literature by some philosophers. For example, a very basic common dictionary (there are a few) would give two definitions: **1)** being conscious: *We have no consciousness during sleep. The blow caused him to lose consciousness. He did not recover/regain consciousness until two hours after the accident.* **2)** all the ideas, thoughts, feelings, wishes, intentions, recollections, of a person or persons: *the moral consciousness of a political party.* Definition one refers to the state of being awake and the second to consciousness as an umbrella term, which contains lots of «mental» activities such as memory, thoughts and feelings. In a better dictionary (e.g., Oxford) the word consciousness contains eight entries. A few are obsolete, or no longer in use, and date as far back as 1639 (it is paradoxical that after three and a half centuries we are still unsure about its definition). This dictionary also refers to consciousness in two basic dimensions; one denotative of wakefulness and another denotative of «mental activity». One could hardly blame an unspecialised dictionary for not disambiguating between subjective experience and awareness (or grouping every mental activity under one word).

Unfortunately, in many cases the expert literature has inherited this ambiguity, making progress a little slower than it should. For example, John Searle's (1990) definition of consciousness:

> By consciousness I simply mean those subjective states of awareness or sentience that begin when one wakes in the morning and continue throughout the period that one is awake until one falls into a dreamless sleep, into a coma, or dies or is otherwise, as they say, unconscious

Notice that Searle includes wakefulness, subjective states and awareness in the definition of consciousness. Issues arising from this definition will necessarily be misleading in that some will focus on awareness or wakefulness when trying to establish the nature of consciousness (e.g., Faulkner and Foster, 2001; Schiff and Plum 2000, respectively). Or simply, they will advance the psychological aspect of the awareness tradition. In this respect the paper by Faulkner and Foster for example appears to be a nice continuation of an already established tradition in the implicit processing of emotional pathology (e.g., MacLeod, 1991) applied to the neuropsychological domain. However, nothing is said about the impact of this not-so new field on «consciousness». I am fully aware of Searle's defense of the irreducibility of consciousness, but his overinclusive definition does, perhaps tacitly, justify an unrestricted invitation to cognitive scientists to «explore» consciousness. The problem with this open invitation is that the subjective experience tends to remain unexamined.

In summary, common dictionary definitions of consciousness are extremely ambiguous and this has been carried through to more specialized definitions (e.g., Searle). These types of definitions are overinclusive and refer to at least three different types of phenomena. Solving this type of mistake would lead to content progress not just nominal clarifications. For example, studies of neuropathology affecting consciousness would be different to studies of neuropathology studying wakefulness. Similarly, studies connecting neuropsychological neglect and awareness would look different to studies linking neuropsychological neglect and consciousness. Perhaps based on this distinction, refocusing on alternative methodologies with a phenomenological orientation is warranted, (e.g., Neurophenomenology, Varela, 1995; the three principles proposed by Chalmers, 1996, i.e., coherence, organizational invariance and the double aspect theory of information). However, because there are fundamental differences between Varela and Chalmers, analysis of these differences would deviate us from the less controversial nominal issues. Thus, identifying what we mean by consciousness is a first step even when we want to deny its existence (e.g., Dennett, 1990). Furthermore, it appears clumsy to acknowledge the ambiguity and at the same time introduce more confusion by not differentiating enough among different types of phenomena or differentiating it by introducing more descriptors of consciousness (e.g., Block, 1995). The link between consciousness and some other mental activity (e.g., awareness, memory, executive functioning, etc.) is a logical next step and there is abundant literature doing this, but not all of them differentiate among associated yet different phenomena.

References

Baars, B. (1995). Evidence that phenomenal consciousness is the same as access consciousness. *Behavioral and Brain Sciences* 18 (2): 249.

Beck, A. (1967). Depression. New York: Hober Medical.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18 (2): 227-287.

Chalmers, D. (1996). *The conscious mind. In search of a fundamental theory*. New York: Oxford University Press.

Dennett, D. C. (1991). *Consciousness Explained*. Little Brown. Boston MA.

Faulkner and Foster, J. (2001) The Decoupling of «Explicit» and «Implicit» Processing in Neuropsychological Disorders. Insights Into the Neural Basis of Consciousness PSYCH. An interdisciplinary journal of research on consciousness. 8 (02). 1-28.

Gelder, B., De Haan, E. and Heywood, C. (2001). *Out of mind. Varieties of unconscious processes.* Great Britain: Oxford University Press

Graf, P. and Schafter, D. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition,11*, 501-518.

Hornby, A. (1987). *Oxford Advanced Learner's Dictionary of Current English*. Great Britain: Oxford University Press

Kirsner, K., Speelman, C., Maybery, M., O'Brien-Malone, A., Anderson, M. and MacLeod, C. (Eds.) (1998). *Implicit and Explicit Mental Processes*. USA: Lawrence Erlbaum Associates, Inc.

MacLeod, C (1991). Clinical anxiety and the selective encoding of threatening information. *International review of psychiatry. 3*,279-292

MacLeod, C., Mathews, A. and Tata P. (1986) Attentional Bias in Emotional Disorders. *Journal of Abnormal Psychology*. 95 (1), 15-20.

MacLeod, C. and Rutherford, E. (1992). Anxiety and the selective processing of emotional information: Mediating roles of awareness, trait and state variables, and personal relevance of stimulus materials. *Behaviour Research Therapy*. 30 (5), 479-491.

MacLeod, C. and Hagan, R. (1992). Individual differences in the selective processing of threatening information, and emotional responses to a stressful life event. *Behaviour Research and Therapy*. 30 (2), 151-161.

MacLeod, C. and Mathews, A. (1991). Biased cognitive operations in anxiety: accessibility of information or assignment of processing priorities. *Behaviour Research and Therapy*. 29 (6), 599-610.

MacLeod, C. and McLaughlin, K. (1994). Implicit and explicit memory bias in anxiety: A conceptual replication. *Behaviour Research and Therapy*. 33 (1), 1-14.

MacLeod, C. and Lawrence Cohen, I. (1993). Anxiety and the interpretation of ambiguity: A text Comprehension Study. *Journal of Abnormal Psychology*. 102, 2, 238-247.

MacLeod, C., Rutherford, E., Campbell, L., Ebsworthy, G. and Holker, L. (2002). Selective attention and emotional vulnerability: Assessing the causal basis of their association through the experimental manipulation of attentional bias. *Journal of Abnormal Psychology*. 111, 1, 107-123.

MacLeod, C. and Mathews, R. (1988). Anxiety and the allocation of attention to threat. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 38*, 659-670.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*

Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour. 6*, 855-863.

Schiff, D. and Plum, F. (2000) The Neurology of Impaired Consciousness: Global Disorders and Implied Models. *E-seminars. Association for the scientific study of consciousness.*

Searle, J. (1990) Who is computing with the brain? *Behavioral and Brain Sciences*. 13:4: 632-642.

Varela, F. (1996). Neurophenomenology: A Methodological Remedy for the Hard Problem. *Journal of Consciousness Studies*, «Special Issues on the Hard Problems», J.Shear (Ed.) June 1996.

Weiskrantz, L., Warrington, E.K., Sanders, M.D., & Marshall J. (1974). Visual capacity of the hemianopic field following a restricted occipital ablation. *Brain 97*, 709-28.

Rodrigo Becerra

Murdoch University

Australia, 6150

rbecerra@central.murdoch.edu.au

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp. 24-28
Memetics: An Evolutionary Theory of Cultural Transmission

MEMETICS: AN EVOLUTIONARY THEORY OF CULTURAL TRANSMISSION

Asunción Álvarez

---

1. Introduction

In his 1976 book *The Selfish Gene*, evolutionary biologist Richard Dawkins hypothesized that living beings, including humans, are mere «vehicles» for the transmission of the genetic information they bear. Genes, said Dawkins, are «replicators», information units which produce copies of themselves in order to be transmitted from generation to generation; and evolution can be understood as directed by those replicators in order to preserve their continuity.

But in the last chapter of his book, «Memes: the New Replicators», Dawkins took a step further. Dissatisfied with the usual Darwinian explanations of human behaviour in genetic terms, he postulated the existence of a unit of cultural transmission, analogous to the gene, which he termed ***meme***. Like genes, memes would be replicators, and the mechanism by which they produced copies of themselves would be imitation:

> Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation [1].

Ever since, the study of memes — ***memetics*** — has strived to attain a scientific status. Scott Atran has described the steps to be taken in order to achieve this goal: the initial stage would be specifying whether and how the gene / meme analogy stands under verifiable scrutiny. Were the analogy to hold, it would then lead to «significant and surprising discoveries about specific causal structures». Were the analogy not to hold, the whole endeavour would have to be discarded as unscientific, although Atran acknowledges the likelihood that

> such an original and enticing idea as the meme will endure with significantly altered content, or as an expedient trope that orients attention, like the etiological notion of «germ» [2]

The question is thus whether the notion of meme is to have a merely heuristic value, or whether on the contrary it may make possible the development of a scientific programme, be it in its original Dawkinsian form or in some modified version.

2. The Meme-as-Germ and the Meme-as-Gene

Literature on memetics tends to divide into two main branches, according to the analogy on which it is based. The most popular interpretation of memetics, sometimes known as the **«meme-as-germ»** interpretation, sees memes as similar to disease agents — as reflected in the

titles of well-known popularisations and / or vulgarisations of memetics: *Virus of the Mind* [3], *Thought Contagion* [4], etc. This point of view lays great emphasis on the memetic answer to the key question in evolutionary theory, which Daniel Dennett has formulated in Latin as «*Cui bono*?», «Who profits?» Whereas traditional social science has always assumed that cultural behaviour must ultimately benefit the individuals displaying it, according to memeticists it is not brains, individuals or societies that profit from evolutionary cultural dynamics, but memes themselves. The meme-as-germ approach thus makes great play of Dawkins's description of memes as leaping flea-wise from head to head, extending to cultural units his pithy description of genes as immortality-seeking parasites which use and then discard their hosts.

The meme-as-germ approach to the study of cultural transmission takes **epidemiology** as its model science. Memes would thus be the cultural equivalents of flu bacilli, transmitted through the communicational equivalents of sneezes. As a result of this epidemiological analogy B — until recently, the predominant strain in memetics — memeticists have tended to focus on how memes are transmitted from individual to individual. However, despite its popularity, the viral analogy tends towards distortion, endowing memes with a certain unwarranted malign character (which can be easily avoided by simply conceiving of memes in their more traditional guise as concepts or representations).

The second main approach to memetics, also known as the **«meme-as-gene»** interpretation, exploits in more depth Dawkins's original analogy between genetic and cultural transmission. This school of thought takes **evolutionary genetics** as its model science. It thus adopts a «meme's eye view» of cultural processes, similar to the «gene's eye view» which has characterized the development of evolutionary theory in the last years. It is this trend which is currently prevailing in memetics, and its aim is to precisely detail the ways in which cultural transmission is evolutionary. In order to do so, it has pushed the analogy with the gene to its logical extremes, seeking cultural equivalents for the main evolutionary genetic concepts, such as genotype, phenotype, transcription, code, etc. The main concept taken by evolutionary memetics from genetics is however that of **replication**: just as DNA strands replicate by producing identical copies of themselves B with an inevitable rate of mutation, which allows for evolution — so memes would replicate themselves in order to be transmitted from bearer to bearer. Standard evolutionary memetics also argues that the means by which memes replicate themselves is the form of learning known as **imitation**. In the next section, we shall examine these two concepts more closely.

3. Replication and Imitation

The idea of applying Darwinian evolutionary theory to the study of culture is of course nothing new. The aim of sociobiology and its latest offspring, evolutionary psychology, is to close the gap between the natural and the social sciences, bringing about a general — and, according to their followers, long overdue — «Darwinization» of the study of the human being. These schools of thought consider all human behaviour as the consequence of the interaction of evolved physiological and psychological variables with the natural environment. Thus, human culture would ultimately be biologically determined by the evolutionary history of the species. This stance leads sociobiologists and evolutionary psychologists to regard certain widespread behaviours, «backfiring» from the evolutionary point of view (such as the use of contraceptives), as the maladaptive result of the clash between the Stone Age instincts hardwired into our brains and our current technological environment.

The extreme reductionist claims of sociobiologists and evolutionary psychologists have been somewhat tempered by the cultural selectionism approach, which admits the existence of a *dual* system of inheritance in the human species: genetic transmission and cultural transmission. (This approach is thus sometimes also termed gene-cultural coevolution). In this respect, memetics can be considered a subcategory within cultural selectionism, differing from mainstream coevolutionism in its insistence on replication as the mechanism of cultural inheritance (cultural selectionism merely posits an unspecified inheritance mechanism).

The idea that cultural units replicate as genes do assumes that discrete, definable units *can* be distinguished within culture. Which is, of course, a rather strong claim to make. The memetic stance has been criticized, mainly by anthropologists, on the grounds that culture constitutes a *continuum*, and any units within it will necessarily be arbitrary constructs of the observer.

But even if cultural units can be properly distinguished, the memeticists' insistence on replication as the only mechanism for cultural inheritance has also been criticized on the grounds that replication is the exception rather than the rule in processes of cultural transmission — the rule being almost always *transformation*. Such obvious examples as linguistic change or the often bewilderingly varying versions of rumours and urban legends indicate that mutation is indeed the default case in processes of cultural diffusion. As Dan Sperber has put it [5], when actual replication does takes place, it can be seen as the limit case of zero transformation.

A further argument is that the evolutionary character of cultural transmission is not compromised by the definition of its inheritance mechanism as other than replication. Indeed, as Joseph Henrich and Robert Boyd have shown [6] — and as cultural selectionists have always emphasized — evolutionary processes can take place without replication. Replication is only a particular instance of inheritance, peculiar to DNA.

Thus, it seems increasingly obvious that the only reason for the maintenance of replication as the mode of transmission is the adherence to the original gene / meme analogy (or, as Francisco Gil-White jokingly terms it, the «fetishism of the gene analogy»). Even Dawkins himself softened his initial views in *The Extended Phenotype*:

> The copying process is probably much less precise [in the case of memes] than in the case of genes: there may be a certain «mutational» element in every copying event […] Memes may partially blend with each other in a way that genes do not. New «mutations» may be «directed» rather than random with respect to evolutionary trends. [7]

Sperber has suggested that the «directedness» of cultural transmission may be due to cognitive mechanisms characteristic of the human species. Thus memes would not be replicators, but units subject to transformation in the course of constructive processes of an inferential nature.

Another difference between cultural selectionism and memetics is that cultural selectionism considers that any means of non-genetic transmission can take part in cultural transmission, and thus admits a wide array of possibilities: imprinting, classical conditioning, operative conditioning, observation, imitation, direct instruction. By contrast, standard memetics is more restrictive in that it defines imitation as the only learning process which makes cultural diffusion possible. According to one of the leading memeticists, Susan Blackmore [8], the reason for this emphasis on imitation is that it is the only means of transmission which makes possible the accumulation of modifications over generations — what

is known as the «ratchet effect», whereby cumulative changes eventually become irreversible. This would make imitation the distinguishing feature of the human species. In her book *The Meme Machine* [9], Blackmore also put forth the hypothesis that the «memetic drive» which had originally led to the development of the capacity for imitation in mankind would also account for such odd phenomena peculiar to the human species as hypertrophied brains, speech, or the tendency to engage in non-altruistic acts, even in large groups of non-kin.

The concept of the «memetic drive» is, according to Blackmore, unique to the mimetic perspective and what sets it apart from alternative evolutionary approaches to culture. However, its restriction to imitation has been questioned by philosopher David Hull [10], who argues that the consequent restriction to the human species seriously reduces the span B and interest B of memetics. An exclusively human memetics, argues Hull, cannot explain such general evolutionary trends as the increase of intelligence within some animal families.

4. Conclusions

As Robert Aunger has pointed out [11], a potential memetic science faces three main questions:

(1) Whether culture can be seen as composed of independently transmitted information units;

(2) whether the process whereby these units are transmitted is necessarily one of replication; and

(3) whether a Darwinian or selectionist approach is the most adequate form for a science of culture to take.

The answer given by standard memetics to all these three questions is «yes». However, the objections mentioned in this essay are but a sample of the complications which such an unqualified answer raises. In the future, memeticists must meet these complications, revising and redefining their current conceptions of replication and imitation, in order to provide a consistent theory, if memetics is to attain a truly scientific status.

References

[1] Dawkins, Richard. *The Selfish Gene* (revised ed.), pg. 192. Oxford: OUP, 1989.

[2] Atran, Scott. «The Trouble with Memes: Inference versus Imitation in Cultural Creation», in *Human Nature* 12(4): 351-381, 2001.

[3] Brodie, Richard. *Virus of the Mind: the New Science of the Meme*. Seattle: Integral Press, 1996.

[4] Lynch, Aaron. *Thought Contagion: How Belief Spreads through Society*. New York: Basic Books, 1996.

[5] Sperber, Dan. *Explaining Culture. A Naturalistic Approach*. Oxford: Blackwell, 1996.

[6] Henrich, Joseph & Boyd, Robert. «On modeling cognition and culture. Why cultural evolution does not require replication of representations», in *Journal and Cognition and Culture* (forthcoming).

[7] Dawkins, Richard. *The Extended Phenotype*, pg. 112. Oxford: Freeman, 1982.

[8] Blackmore, Susan. «The memes' eye view», in Aunger, Robert, *Darwinizing Culture*, Oxford: OUP, 2000.

[9] Blackmore, Susan. *The Meme Machine*. Oxford: OUP, 1999.

[10] Hull, David. «Taking memetics seriously», in Aunger, Robert, *Darwinizing Culture*, Oxford: OUP, 2000.

[11] Aunger, Robert. «Introduction», in Aunger, Robert, *Darwinizing Culture*, Oxford: OUP, 2000.

Asunción Álvarez

Universidad Complutense de Madrid

asun_alv@yahoo.com

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp. 29-41
Ontic Vagueness in Microphysics

ONTIC VAGUENESS IN MICROPHYSICS

Silvio Seno Chibeni

> The farther physical science progresses the less can it dispense with philosophical criticism. But at the same time philosophers are increasingly obliged to become intimately acquainted with the sphere of research, to which they undertake to prescribe the governing laws of knowledge.
>
> E. Schrödinger (1957, 51)

1. Introduction

It is difficult, if not impossible to characterize vagueness without prejudging the issue in favour of one or another of the main interpretations of vagueness. Perhaps the central element in the notion is the existence of a *fuzzy boundary*.[1] Thus, defenders of the linguistic interpretation say that a term is vague when its meaning is not precise, whereas proponents of the epistemic interpretation hold that vagueness results from lack of precise knowledge. Those, on the other hand, who defend ontic or metaphysical vagueness usually take a vague object as an object whose physical properties are blurred or indeterminate. Another way of expressing this point is to say that a vague object is an object whose properties are not — as a matter of fact — all precisely specifiable or definable. It is *not* that we are uncertain whether the property applies to the object (this would be epistemic vagueness), but that there is objectively «no determinate fact of the matter whether that object exemplifies that property.»[2]

Much of the voluminous literature on vagueness is devoted to the question of whether there is, or there can be, vagueness in the world itself, as contrasted with its representation in thought or language. The current disinclination of students to answer these questions positively appears to derive from two main sources. There is, first, the weight of the classical analyses of the issue. As is well known, Frege regarded vagueness as a defect of ordinary language. Indeed, this is one of the reasons why he and virtually all the early analytic philosophers concentrated their attention to artificial languages.[3] Also, in his seminal 1923 article on vagueness, Bertrand Russell maintained that «Vagueness and precision alike are characteristics

---

[1] Vagueness is also commonly characterized in terms of *borderline cases*. Here we shall focus on fuzzy boundaries, because the existence of borderline cases follows from the existence of fuzzy boundaries, whereas the converse does not seem to hold (see Keefe and Smith 1997, 15-16).

[2] Merricks (2001, 145). This paper offers particularly clear characterizations of the three main views of vagueness. For a more specific attempt to define ontic vagueness, see Sainsbury (1989).

[3] For an account of Frege's views on vagueness, see Williamson (1994, sect. 2.2).

which can only belong to a representation, of which language is an example. They have to do with the relation between a representation and that which it represents».[4] This was meant by Russell to apply also to thoughts, which he regarded as a kind of private representation. But the attribution of vagueness to the represented objects is denounced by him as an instance of the «fallacy of verbalism — the fallacy that consists in mistaking the properties of words for the properties of things».[5] Finally, Michael Dummett claimed, in a much-quoted phrase, that «the notion that things might actually *be* vague, as well as being vaguely described, is not properly intelligible».[6]

The second main source of antipathy to ontic vagueness is Gareth Evans's one-page 1978 article. On the face of it, the paper offers a formal proof that «the idea that the world might contain certain objects about which it is a *fact* that they have fuzzy boundaries», being therefore vague, is not «coherent». Not unexpectedly, the exact meaning and import of Evans's cryptic proof became the subject of hot controversy in the literature, which continues unabated to our days.[7]

In this article we shall not re-examine the classical arguments against ontic vagueness, nor discuss the details of Evans's proof. Our aim is to contribute to the debate through a philosophical analysis of some central elements of our best scientific understanding of the nature of the material world. More specifically, we shall explain in some detail why our basic theory of matter, quantum mechanics (QM), describes objects as being irreducibly vague. We shall also indicate that there are strong theoretical and experimental reasons for taking this aspect of QM as having come to stay.

2. The import of science to the debate on ontic vagueness

Vagueness has traditionally been regarded as a philosophical issue, and this is just right, since it has traditionally been associated with language and thought, and these undoubtedly are philosophical provinces. Nonetheless, philosophy has also traditionally been interested in reality, except within certain philosophical schools. Given that vagueness was brought to fore in contemporary philosophy when one of these schools dominated the philosophical scene, it

---

[4] P. 62, as reprinted in Keefe and Smith (1997b). In addition, Russell maintained, controversially, that all words in natural languages, even logical terms, are to some extent vague.

[5] *Ibid*., p. 62. For a recent criticism of Russell's position, see Colyvan (2001). The present article can be taken as providing support to Colyvan's general criticism, as it presents a concrete, fully developed scientific case for the existence of vague objects.

[6] (1975, 260), as reprinted in Dummett (1978). It is fair to remark, however, that Dummett later recanted from this strong position; see his (1981, 440).

[7] For a sample of the most important attempts at clarification, see e.g. Lewis (1988), Burgess (1989) and (1990), Parsons and Woodruff (1995), Over (1989), Johnsen (1989), Keefe and Smith (1997a, 49 ff), Williamson (1994, sect. 9.2), Pelletier (1989). A different line of attack on ontic vagueness has been proposed by Sorensen (1998); for a criticism, see Markosian (2002).

is no surprise that it was, and still is, largely or exclusively taken as a linguistic or mental phenomenon.

But abhorrence to metaphysics has faded away, and the time is ripe for rehabilitating reality as a genuine philosophical subject. In the early modern period, however, philosophy gave birth to science, which has taken upon itself the task of investigating the material world. From that time on, we cannot, thus, afford to ignore what science says about the nature of the material objects. However, most of the philosophers engaged in discussing vagueness — even ontic vagueness — do not appear to recognize this point fully.

Recent analyses of ontic vagueness have focused almost exclusively on macroscopic objects, such as clouds, mountains and cats. Furthermore, when the constitution of these objects is discussed (as formed by atoms, for instance), the analysis is implicitly guided almost exclusively by theories of classical physics. Since these theories leave no place for vagueness — in the sense that, according to them, all the properties of the *elementary* constituents of matter are in principle specifiable with complete precision — this has the effect of biasing the whole discussion against ontic vagueness from the very beginning.

Thus, claims of vagueness in the material objects have been easily dismissed as merely «superficial» vagueness. On the usual (but debatable: see Chibeni 2004) assumption that the properties of the macroscopic objects supervene on the properties of their microscopic constituents, any vagueness in the former could in principle be eliminated by their theoretical reduction to the latter.

Attention to this important distinction between superficial and non-superficial, or fundamental, vagueness has been drawn by Keefe and Smith (1997a, 56-57) and Burgess (1990). Whereas the former authors do not take any position on the dispute, Burgess appears to regard superficial vagueness as genuine ontic vagueness, irrespective of what happens at the basic level. Although disagreeing with Burgess on this point, we strongly support his view that the issue of whether «the world is microscopically divisible into sharp objects … is best treated as an empirical claim» (p. 285).

Now, we obviously get different answers to this question, depending on which theory we choose. We believe that our guide here should be the *best currently available* physical theory. The fact that this theory will, like any other, be fallible does not imply that the choice is immaterial. Even if our interest is restricted to the question of whether there *can* be vague objects — as opposed to whether there actually are such objects in the world —, the theoretical choice is important. It is just silly to rely — for whatever purpose — upon a theory which is *known* to have met with refuting evidence. Curiously, this point has been largely ignored by the students of ontic vagueness. The first noticeable exception was, to our knowledge, provided by Lowe 1994.

In this article Lowe argued that a certain quantum mechanical system involving a pair of electrons constitutes a genuine instance of ontic vague *identity*. Lowe's example was, thus, directly addressed to Evan's proof.[8] Lowe's paper has generated some interesting discussion

[8] Besides offering the quantum counterexample to the proof, Lowe endeavoured, as many did before him, to locate its «flaw» by direct analysis.

in the literature.[9] Although fully agreeing with Lowe's line of inquiry, we think that he was unfortunate in the choice of his example, since it involves the thorny issue of the identity of quantum objects. Also, the whole controversy over Evans's proof piggybacks on his analysis, making the issue rather too complex. We shall not enter into this discussion here. In the following section we explain, through a general theoretical analysis, how ontic vagueness arises in QM. In section 5 we illustrate the point by offering a sample of straightforward examples of ontically vague quantum objects which do not involve the identity relation. And in section 4 we point out that certain theoretical and experimental results, made available in the second half of the twentieth century, impose forbiddingly severe constraints on any microphysical theory purporting to avoid ontic vagueness.

3. A theoretical analysis of quantum mechanics vis-à-vis the issue of ontic vagueness

Both in classical theories and in QM, the properties of objects fall into two classes: static, or state-independent properties, such as rest mass, charge and spin, and dynamic, or state-dependent properties, such as position, momentum, energy, spin components, etc. The former are always precisely definable, in both kinds of theories. Dynamic properties, on the other hand, are typically *not* sharply definable in QM, in contrast with what happens in classical theories. This fundamental difference arises from the peculiar way quantum mechanics characterizes the *states* of physical objects.

Whereas in classical mechanics the state of a particle is represented by a set of six numbers — the three components of its position and of its momentum — in quantum mechanics the pure states of an object are complex-valued functions — usually referred to as *wavefunctions* — or, more generally, vectors in a Hilbert space. In both classical and quantum mechanics the purpose of defining states is to allow the prediction of the physical properties belonging to the object. In the former theory, the specification of the state allows, in principle, the prediction of *all* the dynamical quantities of the object, such as its kinetic energy, angular momentum, etc. Quantum mechanical states, however, do *not* afford a complete value assignment to all the quantities which can legitimately be measured on and therefore, apparently, attributed to the object. It should be stressed that this holds even for the *pure* quantum states, i.e. the states embodying *maximal* information about the object. This unique situation in the history of physics is illustrated in section 5 by three simple examples.

The fact that *no* quantum mechanical state gives precise values to all the dynamical properties of quantum objects immediately leads to the suspicion that the theory is *incomplete* as a description of physical reality. This apparent incompleteness of QM is at the root of most of the intriguing features of this theory, and separated the founding fathers into two opposite camps. Led by Bohr and Heisenberg, most of them *denied* that there is anything missing in the quantum mechanical theoretical description («*position 1*»), whereas Einstein and Schrödinger insisted that the theory is, indeed, incomplete («*position 2*»). The two most powerful arguments to sustain the latter view appeared in 1935: Einstein, Podolsky and Rosen's argument concerning certain pairs of correlated quantum objects (EPR 1935) and Schrödinger's argument concerning the measurement process, known as the «cat paradox» (Schrödinger 1980).

---

[9] See e.g. French and Krause (1995, 1996 and 2003), Noonan (1995), Hawley (1998), Odrowaz-Sypniewska (2001), Lowe (1997 and 1999).

This is not the place to examine the controversy over the completeness of QM. We just want to explore its connections with the issue of ontic vagueness. We begin by noticing that the incompleteness view (position 2) clearly suggests an epistemic interpretation for quantum vagueness. According to this view, the lack of sharp values of physical magnitudes in QM is to be regarded a theoretical aspect only, to be eliminated through the addition of more information on the object, in the scope of a more complete theory.

The interpretation of the opposite view (position 1) is more complex. There are three general options open to the proponents of the completeness of QM:

1a) *Anti-realism:* the concept of a physically describable reality is abandoned. This stand was often taken by Bohr and his followers. The problem of ontic vagueness is thereby bypassed; the theory is meant as referring to phenomena only, not to real objects lying behind them. The lack of a complete value assignment in QM is interpreted as a trait of quantum theoretical language only.

1b) *Heisenberg's disturbance doctrine:* the objects are conceived as possessing sharp attributes only, but they are mostly «unknowable *in principle*». Due to the existence of the so-called «quantum of action», the act of observation would introduce an unavoidable and uncontrollable disturbance in the state of the objects, so that the precise values of many of their properties are always beyond our reach. The positivist doctrine (fashionable in the 1930's) would then discharge QM from the task of describing these properties. The theory should, thus, be considered complete, at least with respect to what can be *known* about reality. In this case there is no real ontic vagueness, just epistemic vagueness. The difference with respect to position 2 (incompleteness) is that now the missing information is claimed to be experimentally unobtainable.

1c) *Reality itself is fuzzy:* the attributes lacking theoretical values in QM are objectively blurred. The classical ontologies of sharp objects are replaced by a notion of reality with fuzzy objects, exactly to match what is found in the quantum formalism.

Failure to distinguish clearly these positions has often led to deep confusions in the historical debate concerning QM. In his classic 1927 article on the indeterminacy relations, for instance, Heisenberg first deduced his relations — which were to become the *locus* for discussing the failure of QM to provide a complete property assignment — from the mathematical properties of wavefunctions, and then tried to confirm them physically by the famous gamma-ray microscope thought experiment. Now, whereas the initial deduction presupposes that reality is conceived as a literal counterpart of the wavefunction (a possible way to instantiate position 1c), quantum objects being thus «wave-like» and therefore fuzzy, the microscope experiment assumes that reality is formed of more or less classical particles, with sharply defined properties, but whose precise values are claimed to lie beyond experimental determination (position 1b). In his notoriously obscure texts on the completeness of QM, Niels Bohr also intermingled elements of both the ontic and the epistemic defences of completeness (positions 1c and 1b), as well as of anti-realism (position 1a).

Now the tenability of the central thesis of this article depends on the existence of good reasons for adopting position 1c. We believe such reasons *do* exist: strong objections can be raised to all the other alternatives.

Firstly, although most of the founding fathers of QM leaned towards one type or another of anti-realism (position 1a), we hold that the abandonment of the classical realist stand in

science is not forced upon us by QM, as they often assumed, and that a careful philosophical analysis of the issue favours realism instead (Chibeni 1999).

Secondly, concerning position 1b, even if positivism is taken for granted, the defence of completeness through the idea of a disturbance upon measurement has several irreparable conceptual shortcomings, as first shown by Popper in his *Logic of Scientific Discovery* (first German edition 1934). But Popper was swimming against the tide, and his criticism passed virtually unnoticed for more than two decades. It is now generally agreed, however, that his arguments were sound, and that his point can be supported by other, independent arguments as well. Details on this issue can be found elsewhere (Chibeni 2001).

Finally, we shall examine in a separate section the case against position 2 (incompleteness), as it deserves a more detailed attention.

4. Restoring ontic sharpness in microphysics?

By 1935 the completeness thesis was already prevalent, and not even EPR's and Schrödinger's powerful arguments for incompleteness changed the rapidly established orthodoxy. It is our opinion, however, that the weight of the original rebuttals to these arguments was overestimated, and that sound evidence for completeness did not arise until much later. Ironically, such evidence — which, given the analysis above, should also count as evidence for ontic vagueness — was finally obtained through a series of no-go results concerning the so-called *hidden variables* research programme, which aimed exactly to complete what was apparently missing in QM.

In the literature on the foundations of QM, the expression 'hidden variables' designates certain parameters, to be added to the quantum mechanical states in order that all measurable physical magnitudes of objects get a definite, sharp value. The first and most important *hidden variables theory* (HVT) was formulated by David Bohm in 1952. This theory is capable of reproducing all the quantum mechanical empirical predictions and, at the same time, of restoring sharpness in all the properties of quantum objects. As Bohm himself noticed, however, this achievement has a price: certain other theoretical and conceptual traits of classical theories are violated by the theory. More importantly, further theoretical and experimental research has revealed that not only Bohm's theory has to pay this price, but *any* other theory capable of completing the quantum mechanical property assignment must pay as well. We are here referring to the following three classes of results.

There is, first, a series of algebraic proofs, in the tradition of von Neumann's famous 1932 theorem, to the effect that completing the quantum states through hidden variables leads to inconsistencies (Gleason 1957, Bell 1966, Kochen and Specker 1967, Mermin 1990). Bohm's theory escapes inconsistency only by incorporating a form of «contextualism», roughly meaning that some properties assigned to the object somewhat reflects its «experimental context» in a thoroughly non-classical way.

Secondly, in 1964 John S. Bell proved that the most objectionable trait of Bohm's theory, *nonlocality*, must be present in any HVT reproducing certain quantum mechanical predictions concerning correlated, EPR-type pairs of objects.[10] These peculiar predictions have

[10] Roughly put, locality is the assumption — well backed by relativity theory — that all physical influences take finite time to propagate in space. For an exposition of the reasons that have led Einstein to maintain that this is a principle to which we should «absolutely hold fast» in physics (1949, p. 85),

subsequently been confirmed by several experiments, the most important of which being reported in Aspect et. al. (1982). *Any* empirically adequate HVT must, therefore, be non-local.

Finally, some authors succeeded, more recently, in bringing together these two classes of results, showing that the assumption of a local HVT also leads to mathematical inconsistencies (Heywood and Redhead 1983, Greenberger et al. 1989).

These results mean that although the restoration of sharpness in the dynamic properties of quantum objects is possible, as clearly shown by Bohm, the price may be too high. They form, thus, the basis of a strong argument for taking quantum mechanical fuzziness as being much more than a peculiarity of a *specific* theoretical representation of reality (QM).[11]

5. Vague quantum objects: three simple examples

The boundaries concerning which objects are regarded as vague or precise are often taken as their spatio-temporal boundaries.[12] This restriction is by no means necessary, and tends to bias the issue against ontic vagueness, especially when the theories used to analyse the notion of a material object are classical theories. It is important to bear in mind that it can be discussed, for instance, if an object possesses a definite energy, or spin component, or polarization, and these are not spatio-temporal properties.

QM offers plenty of examples of objects lacking properties of both kinds. According to QM this is indeed the *rule*, not the exception, at least in the case of the fundamental entities forming the material world, such as electrons, photons, neutrinos, protons, neutrons, quarks, etc. We shall now give three straightforward examples, taken from non-relativistic quantum mechanics. None of the conclusions drawn depends on this or other simplifications, which are made solely for the sake of mathematical simplicity.

a) Let us begin by considering the simplest example possible, that of a single particle with one degree of freedom not subject to the action of forces. Take, for instance, an electron allowed to move along a straight line, and let us concentrate on two of its dynamic properties, position and momentum (mass times velocity). At any given instant its quantum mechanical state will then be the complex-valued function of the spatial coordinate x

$$\Psi(x) = (2\pi\hbar)^{-½} \int_{-\infty}^{+\infty} \phi(p) \exp(ipx/\hbar) dp$$

see Fine 1986 and Howard 1985.

[11] French and Krause seem to be the first who have drawn attention, if only *en passant*, to the no-hidden-variables results in connection with the issue of ontic vagueness. In their (1996), for instance, they remark that «the force of Bell's Theorem lies in its generality, and it is this which renders the vagueness ontic in the sense that it is not dependent upon a *particular* representation» (p. 25). See also French and Krause 2003.

[12] See e.g. Burgess (1990, 263), Keefe and Smith (1997a, 50), Sorensen (1998).

where $\hbar$ is the reduced Plank constant, p is the momentum coordinate and $\Phi(p)$ is a complex-valued function of p. Although $\Psi(x)$ has no straightforward meaning as a «wave» in ordinary 3-d space, by the Born rule its modulus squared gives directly the probability of getting a result between x and $x + dx$ in a position measurement: $P(x)dx = |\Psi(x)|^2dx$. In typical situations, this quantity has non-zero value over large regions of space. Now if the wavefunction $\Psi(x)$ is taken as embodying maximal information on the object prior to measurement, it is unavoidable, upon a realist construal of the wavefunction, to conclude that before measurement the object lacks a precise spatial localization. In other words, on the assumption that QM is complete the above probabilities cannot be understood epistemically, i.e., as reflecting lack of precise knowledge (as they could in classical physics). Thus, if we indulge at all to consider the real object and its properties, we must think of it as something «spread» over space, and lacking a spatial boundary. The exact «shape» of this fuzzy entity will, of course, depend on $\Phi(p)$, which in turn depends on the details of the actual physical circumstances involving the object. Here, it is enough to remark that $\Phi(p)$ itself is a wavefunction, related to another property of the object, its momentum. Once again, this relation is indirect: the (generalized) Born rule says that $|\Phi(p)|^2\, dp$ gives the probability of finding, in a momentum measurement, a value in the interval $[p, p + dp]$. Exactly the same analysis holds for p and for x, that is, p is also typically a fuzzy property of the object. It is worth mentioning that the «amount» or «degree» of fuzziness in these two properties is precisely determined by mathematical analysis of the wavefunctions: the more «diffuse» is the object in space, the less «diffuse» it will be in momentum, and vice versa. This is just one way of analysing the contents of the Heisenberg principle. It should be clear, however, that in this interpretation the principle is not at all about «uncertainties» (as the usual name «uncertainty principle» implies), but about lack of definiteness of properties. (See Chibeni 2001 for more details on this distinction.) Notice, finally, that one of the most important unsolved problems in the foundations of QM is just to understand why and how such fuzzy properties become definite upon measurement. The proponents of the orthodox interpretation of QM famously claimed that the state transition induced by measurement (from wavefunctions such as Ψ and Φ to eigenfunctions of the measured quantity, affording precise values to it) should be introduced *ad hoc* and *post factum*. For a realist this position is totally unacceptable. He wants to understand what is really happening, in terms of physical properties and interactions; furthermore, he cannot get along with the subjectivist idea that this process is peculiar to *measurements*.

b) Our second example results from the first by considering that the object is now bounded between two impenetrable barriers, lying at $x = - a$ and $x = + a$. Although in this case too we have fuzziness in x and p, we shall concentrate our attention on another property, kinetic energy. Restricting, again, the analysis to a given instant, we have that the time-independent Schrödinger equation has two possible classes of solutions:

$$u_n(x) = (a)^{-½} \cos (n\pi x/2a) \text{ (for } n = 1, 3, 5, \ldots)$$

and

$$v_n(x) = (a)^{-½} \sin (n\pi x/2a) \text{ (for } n = 2, 4, 6, \ldots).$$

An important new feature of this example is that if the object is in one of these states (for a particular n) it will have a definite energy, given by

$$E = \pi^2\hbar^2n^2/8ma^2$$

where *m* is the object's mass. Furthermore, only these values can be found in energy measurements. The energy spectrum is thus discrete, or *quantized*. Also, the above states — called the energy *eigenstates* — are stationary, i.e., if the object is put in one of them, it remains indefinitely in it, unless something interferes with the object. However, the superposition principle allows any linear combination of eigenstates as *bona fide* states. Thus, the general state of our bounded object will be

$$\phi(x) = \sum_{n=1}^{\infty} A_n u_n(x) + B_n v_n(x)$$

where A_n and B_n are complex coefficients, and where the terms $A_n u_n$ and $B_n v_n$ exist only for odd and even *n*, respectively. If more than one of the coefficients is non-vanishing, the state will not be an energy eigenstate. In this case the theory does *not* ascribe a precise energy to the object: energy becomes fuzzy. Furthermore, the state will no longer be stationary, which implies that the exact «shape» of this fuzziness will vary with time. By the Born rule we have that $|A_n|^2$ and $|B_n|^2$ give the probability of getting the corresponding energy eigenvalues in an energy measurement, when the object is in the state $\phi(x)$. Again, on the assumption of completeness these probabilities cannot be interpreted epistemically. Therefore, the energy property of the object is vague, except when the state is an energy eigenstate. As in the case of position and momentum, this vagueness disappears when an energy measurement is made, and it remains a mystery why and how this can happen.

c) For our third and last example we shall take a specifically quantum mechanical property called *spin*. The name might suggest some resemblance to the spinning movement of an ordinary body, but this suggestion is misleading. Like ordinary angular momentum, quantum spin is a vector; but contrary to the angular momentum vector, its magnitude is fixed: spin is indeed one of the static properties of elementary particles, by which they are classed. The square of the spin magnitude is conventionally written as $\mathbf{S}^2 = s(s+1)\hbar$, where *s* is a positive integer or half-integer. Thus, for electrons, protons and neutrons, for instance, we have $s = ½$. For simplicity, one says that these are 'spin-½ particles'. One may now define a related set of properties, the *spin components*. These properties depend, of course, on the spin, but they are *dynamic* and, in contrast with the components of ordinary vectors, they are always quantized. For spin-½ objects, for instance, there are just two spin components along any spatial direction *z*: $s_z = +½\,\hbar$ and $s_z = -½\,\hbar$. Take now one of these objects, an electron, say. Its complete quantum mechanical state must include, besides a wavefunction like those of our examples (a) and (b), a «part» related to spin. Disregard, for now, the wavefunction. The spin part of the state is represented by a vector in a two-dimensional Hilbert space (notice that this is *not* the spin vector!). For any direction *z*, two states are of special interest: the eigenstates of s_z, associated with the eigenvalues $+½\,\hbar$ and $-½\,\hbar$. These can be abstractly designated by $|+\rangle_z$ and $|-\rangle_z$. Thus, if the electron is in state $|+\rangle_z$, its spin component along *z* is $+½\,\hbar$; if it is $|-\rangle_z$, this property has the value $-½\,\hbar$. There is no vagueness here. If however, as allowed by the superposition principle, we take a state which is not one of these eigenstates then the spin component becomes blurred. In fact, a general spin state for a spin-½ object can be written as

$$|\Xi\rangle = \alpha\,|+\rangle_z + \beta\,|-\rangle_z$$

where α and β are complex coefficients. When none of these coefficients is zero, the state is such that there *is* ontic vagueness in the property s_z. Once again, the modulus squared of the coefficients gives the probabilities of getting the results $+\frac{1}{2}\hbar$ or $-\frac{1}{2}\hbar$ in a measurement of s_z, and these probabilities are not related to our ignorance as to the real properties of the object, on the orthodox assumption of completeness.

6. Conclusion

After arguing that what science tells us about matter should be taken into account in the debate on ontic vagueness, we remarked that the philosophers' current disinclination to believe in vague objects is partly due to their implicit adherence to superseded classical theories. We showed, both by a general theoretical analysis and by some concrete examples, that our best contemporary theory on the structure of mater, quantum mechanics, clearly ascribes fuzzy properties to objects. The examples were chosen so as to avoid several unnecessary complications inherent in the example proposed by Lowe in his much-discussed 1994 article. Furthermore, we pointed out that several theoretical and experimental results in microphysics afford very strong evidence for the existence of vague objects, as they prove that any theory purporting to restore sharpness in the properties of quantum objects will meet with severe constraints.

We stress that our case for ontic vagueness obviously presupposes a commitment to at least a mild version of scientific realism. But, as we remarked in section 3, QM or, more generally, microphysics does not represent a direct threat to this epistemological stand. In particular, we do not think Putnam is right in holding that quantum vagueness indicates that «something seems to be wrong with metaphysical realism» (1983, 274). Here, we fully agree with French and Krause, who hold, to the contrary, that «one way to maintain a form of realism in the quantum context is to take vagueness seriously».[13] We would go even further than these authors, and say that this is the *best* way of introducing realism in microphysics.[14] If, for good physical and methodological reasons, we shun contextual, nonlocal HVTs, we should accept the challenge of devising an ontology for the micro-world which takes at face value what QM says, and this includes vagueness.

Given the discouragement imposed on a whole generation of students by the leaders of the orthodox, «Copenhagen» interpretation of QM, it comes as no surprise that the search for a quantum ontology is still «in its infancy» (Krause 2000, 164). However, some progress has been made in recent decades, and it is only to be hoped that further research will shed more light on this challenging issue.

Acknowledgement

I would like to thank Mark Colyvan for his detailed and useful comments on a previous version of this article.

[13] French and Krause (1996, footnote 1).

[14] A clear, if bizarre, alternative would be the so-called many-worlds interpretation of QM; see Geroch 1984 and other papers in the same issue of *Noûs* for more details.

References

Aspect, A., J. Dalibard, and G. Roger: 1982, 'Experimental test of the Bell's Inequalities using time-varying analysers', *Physical Review Letters* **49**, 1804-1807.

Bell, J. S.: 1964, 'On the Einstein Podolsky Rosen paradox', *Physics* **1**, 195-200.

Bell, J. S.: 1966, 'On the problem of hidden variables in quantum mechanics', *Review of Modern Physics* **38**, 447-452.

Bohm, D.: 1952, 'A suggested interpretation of the quantum theory in terms of «hidden variables», parts I and II', *Physical Review* **85**, 166-179 and 180-193.

Burgess, J. A.: 1989, 'Vague identity: Evans misrepresented', *Analysis* **49**, 112-119.

Burgess, J. A.: 1990, 'Vague objects and indefinite identity', *Philosophical Studies* **59**, 263-287.

Chibeni, S. S.: 1999, 'Le réalisme scientifique face à la microphysique', *Revue Philosophique de Louvain* **97**, 606-627.

Chibeni, S. S.: 2001, 'Indeterminacy, EPR and Bell', *European Journal of Physics* **22**, 9-15.

Chibeni, S. S.: 2004, 'Holism in microphysics', *Epistemologia* **27** (2).

Colyvan, M.: 2001, 'Russell on metaphysical vagueness', *Principia* **5**, 87-98.

Dummett, M.: 1975, 'Wang's paradox', *Synthese* **30**, 301-24.

Dummett, M.: 1978, *Truth and Other Enigmas*, London, Duckworth.

Dummett, M.: 1981, *The Interpretation of Frege's Philosophy*, London, Duckworth.

Einstein, A.: 1949, 'Autobiographical notes' (transl. P. A. Schilpp), in P. A. Schilpp (ed) *Albert Einstein: Philosopher-Scientist.* 3rd ed, La Salle, Open Court, pp. 3-94.

Einstein, A., B. Podolsky, and N. Rosen: 1935, 'Can quantum-mechanical description of reality be considered complete?', *Physical Review* **47**, 777-780.

Evans, G.: 1978, 'Can there by vague objects?', *Analysis* **38**, 208. Reprinted in Keefe and Smith 1997b, chap. 17.

Fine, A.: 1986, *The Shaky Game. Einstein and the Quantum Theory*, Chicago, The University of Chicago Press.

French, S. and D. Krause: 1995, 'Vague identity and quantum non-individuality', *Analysis* **55**, 20-26.

French, S. and D. Krause: 1996, 'Quantum objects are vague objects', *Sorites* **6**, 21-33.

French, S. and D. Krause: 2003 'Quantum vagueness', *Erkenntnis* **59**, 97-124.

Garrett, B. J.: 1991, 'Vague identity and vague objects', *Noûs* **25**, 341-351.

Geroch, R.: 1984, The Everett interpretation, *Noûs* **18,** 617-33.

Gleason, A. M.: 1957, 'Measures on the closed subspaces of a Hilbert space', *Journal of Mathematics and Mechanics* **6**, 885-893.

Greenberger, D. M., M. A. Horne, and A. Zeilinger: 1989, 'Going beyond Bell's theorem', in M. Kafatos (ed) *Bell's Theorem, Quantum Theory and Conceptions of Universe*, Dordrecht, Kluwer, pp. 69-72.

Hawley, K.: 1998, 'Indeterminism and indeterminacy', *Analysis* **58**, 101-106.

Heisenberg, W.: 1949, *The Physical Principles of the Quantum Theory* (transl. C. Eckart and F. C. Hoyt), Toronto, Dover.

Heywood, P. and M. L. G. Readhead: 1983, 'Nonlocality and the Kochen and Specker paradox', *Foundations of Physics* **13**, 481-499.

Howard, D.: 1985, 'Einstein on locality and separability', *Studies in History and Philosophy of Science* **16,** 171-201.

Johnsen, B.: 1989, 'Is vague identity incoherent?', *Analysis* **49**, 103-112.

Keefe R. and P. Smith: 1997a, 'Introduction: theories of vagueness', in Keefe and Smith 1997b, pp. 1-57.

Keefe R. and P. Smith (eds.): 1997b, *Vagueness: A Reader*, Cambridge, MA, MIT Press.

Kochen, S. and E. P. Specker: 1967, 'The problem of hidden variables in quantum mechanics', *Journal of Mathematics and Mechanics* **17**, 59-87.

Krause, D.: 2000, 'Remarks on quantum ontology', *Synthese* **125**, 155-167.

Lewis, D.: 1988, 'Vague identity: Evans misunderstood', *Analysis* **48**, 128-130. Reprinted in Keefe and Smith 1997b, chap. 18.

Lowe, E. J.: 1994, 'Vague identity and quantum indeterminacy', *Analysis* **54**, 110-114.

Lowe, E. J.: 1997, 'Reply to Noonan on vague identity', *Analysis* **57**, 88-91.

Lowe, E. J.: 1999, 'Vague identity and quantum indeterminacy: further reflections', *Analysis* **59**, 328-330.

Markosian, N.: 2002, 'Sorensen's argument against vague objects', *Philosophical Studies* **97**, 1-9.

Mermin, N. D.: 1990, 'Simple unified form for the major no-hidden-variables theorems', *Physical Review Letters* **65**, 3373-3376.

Merricks, T.: 2001, 'Varieties of vagueness', *Philosophy and Phenomenological Research* **62**, 145-157.

Noonan, H. W.: 1995, 'E. J. Lowe on vague identity and quantum indeterminacy', *Analysis* **55**, 14-19.

Odrowaz-Sypniewska, J.: 2001, 'Quantum indiscernibility without vague identity', *Analysis* **61**, 65-69.

Over, D. E.: 1989, 'Vague objects and identity', *Analysis* **49**, 97-99.

Parsons, T. and P. Woodruff: 1995, 'Worldly indeterminacy of identity', *Proceedings of the Aristotelian Society* **95**, 171-195. Reprinted in Keefe R. and Smith, P. 1997b, chap. 19.

Pelletier, F. J.: 1989, 'Another argument against vague objects', *Journal of Philosophy* **86**, 481-492.

Popper, K. R.: 1968, *The Logic of Scientific Discovery*, 5th ed., London, Hutchinson.

Putnam, H.: 1983, 'Vagueness and alternative logic', in *Philosophical Papers*, vol. 3, chap. 15, Cambridge, Cambridge University Press.

Russell, B.: 1923, 'Vagueness', *Australasian Journal of Philosophy and Psychology* **1**, 84-92. Reprinted in Keffe and Smith 1997b, chap. 3.

Sainsbury, R. M.: 1989, 'What is a vague object?', *Analysis* **49**, 99-103.

Schrödinger, E.: 1957, *Science, Theory and Man* (transl. J. Murphy), London, George Allen and Unwin. (Originally published in 1935.)

Schrödinger, E.: 1980, 'The present situation in quantum mechanics' (transl. J. D. Trimmer), *Proceedings of the American Philosophical Society* **124** (5), 323-338. (Originally published in 1935.)

Sorensen, R. A.: 1998, 'Sharp boundaries for blobs', *Philosophical Studies* **91**, 275-295.

Williamson, T.: 1994, *Vagueness*, London, Routledge.

Zemach, E. M.: 1991, 'Vague objects', *Noûs* **25**, 323-340.

Silvio Seno Chibeni

Universidade Estadual de Campinas

Cx. Postal 6110, 13083-970 Campinas, SP, Brazil

chibeni@unicamp.br

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp. 42-49
Roman Suszko and Situational Identity

ROMAN SUSZKO AND SITUATIONAL IDENTITY

Charles Sayward

---

1. Introduction

At the start of his first paper dealing with the quantified sentential calculus with identity, Roman Suszko writes (Suszko [1968: 8-9]):

> Consider …
>
> (1) Some situations are not facts
>
> It has the same grammatical structure as the following sentence:
>
> (2) Some philosophers are not logicians
>
> Both sentences (1) and (2) are existential sentences. But there is a very deep difference between them. The terms «philosopher» and «logician» in (2) are unary predicates. The terms «situation» and «fact» are not predicates. They are unary sentential connectives like the word «not» which converts any sentence ϕ into a negation: not-ϕ. To see this, let us make the first step in formalizing (1) and (2). We write:
>
> (3) for some *p*, *Sp* and not-*Fp*
>
> (4) for some *x*, *Px* and not-*Lx*.
>
> The letter *p* is a sentential variable and the letter *x* is a nominal variable. They are bound above in (3) and (4) by the existential quantifier: «for some». The symbols *P* and *L* are unary predicate and the letters *S* and *F* are unary sentential connectives like the connective «not». The difference between sentences (sentential variables, sentential formulae) and names is very deep and fundamental in every language. It must be observed in any rigorous thinking. However, natural language leads sometimes to confusion on this point. Having in mind the categorical difference mentioned above we consider the sentence (1) quite as legitimate as the sentence (2). Moreover, both sentences (1) and (2) are true because:
>
> 1. It is a situation that London is a small city but it is not a fact.
>
> 2. Dr. B. W. is a good philosopher but he is not a logician.

The quantifier 'for some' is used in accord with certain inference rules. It is regrettable that we have fallen into the habit of calling that quantifier the *existential* quantifier, for this makes it look as if logic settled something ontological. Some authors, sensitive to this circumstance, have spoken instead of the *particular* quantifier, characterized in terms of its role in inference. That seems to me to be a sound way of speaking, and it is one I shall throughout adopt.

Suszko is inclined to contrast two particular quantifications

(∃x)(x is a philosopher & x is not a logician),

(∃p)(it is a situation that p & it is not a fact that p)

by pointing to the difference between a sentence and a proper name.

But isn't this only a *grammatical* difference?

Quine would say these sentences *both* assert the existence of some object or entity, arguing in something like the following way: Since the result of putting 'Dr. B. Wolniewicz' in for 'x' in 'x is a philosopher & x is not a logician' is an *instance of* and *logically entails* the first quantification, 'Dr. B. Wolniewicz' there occurs *referentially*, its *referent* being some *object*. Similarly, since the result of putting 'London is a small city' in for 'p' in 'It is a situation that p and it is not a fact that p' is an instance of and logically entails the second quantification, 'London is a small city' there occurs referentially, its referent being some object. So, though 'Dr. B. Wolniewicz' is grammatically a proper name and 'London is a small city' is grammatically a sentence, each is *semantically* a term of reference with some object as its referent. Quine [1961: 118].

The guiding principle of this argument is that terms accessible to positions accessible to a variable of quantification are all *referential* and their referents, if any, are *objects*.

What *supports* this principle?

Consider the following answer: What shows that each of these terms (despite their having grammatically different types of occurrence) *refers* to some *object* is this: the kind of *semantical account* we must provide for the quantifications which bind variables in positions available to these terms.

For example, one type of semantical account would assign e.g., human beings as referents for names and nominal variables and situations or propositions or truth-values as referents for sentences and sentential variables. Thus, Ryszard Wójcicki provides such an account for Suszko's quantified sentential calculus with identity. Wójcicki [1984: 326, 333-335]. And M. J. Cresswell provides such an account for propositional identity. Cresswell [1967: 284].

But *how* is the possibility of this sort of semantical account of quantification supposed to show that in its occurrence in 'If it is a situation that London is a small city, then (∃p) it is a situation that p' 'London is a small city' *refers to* something — be it a proposition or a situation or a truth-value or anything else?

One answer would be that it is possible to give an adequate semantical account of quantification *only* in referential terms, and that there *must* be an adequate semantical account of quantification.

Particular quantification is a certain sort of operation within language. The question is when *that* operation forms sentences in assenting to which we commit ourselves to the existence of things.

One answer that suggests itself is to go beyond the inference role of the particular quantifier and find out its semantical role. Its semantical role should be made clear in a truth theory of the language under discussion.

So, I want to ask whether the general sort of semantical account of quantification I have sketched is the *only* one possible. Why not a *non*-referential semantical account of quantification of those quantifications which bind variables in for sentences? I here propose such an account for the quantified sentential calculus propounded by Suszko. I think that such an account is more in keeping with Suszko's remarks quoted at the beginning of this paper

than is Wójcicki's semantical account. Suszko was attempting to set out the ontology of the *Tractatus*. For such an end it would not do to treat grammatically proper names and sentences the same way semantically. For Wittgenstein, while situations are made up of objects, they are not objects themselves. I do not think Suszko's commentators have appreciated the point that for Suszko grammatically proper names and sentences ought not to be treated the same way semantically. For instance, Grezegorz Malinowski writes, «The identity connective … was introduced by Suszko to express coincidence and other referential relations between sentences.» Malinowski [1985: 21].

Quine observes that there are non-referential accounts available for such varieties of quantification. For example, quantifications binding variables in for sentences and clauses whose only modes of combination are truth-functional can be introduced, and thus also eliminated, in terms involving no binding of variables in positions apt for sentences. Hence, an account of such quantification need involve only such elements of reference as may be required by an account of the underlying truth-functionally composed sentences and clauses. Quine [1961: 118]. Also see Church [1956: 151-154]. Further, in some cases we may so quantify as to find it *necessary* for the truth of a particular quantification that it have some true instance. Such quantification is called *substitutional* and its account need involve only such elements of reference as may be required by an account of the sentences which serve as instances of quantifications.

I do not here controvert these claims. But I do note that the first claim is inapplicable to the sort of system Suszko investigates. For that system includes '=' as a sentential connective. And that connective is not truth-functional. As to the second claim I note that those of our particular quantifications which bind variables in positions appropriate for sentences include ones for the truth of which we do *not* require that they have some true instance. For example, 'Some situations are inexpressible' or '(∃p) (it is a situation that p & that p is inexpressible)'.

So when I consider quantifications of the type which bind variables in for sentences, I will take it that such quantifications are not to be construed substitutionally.

2. Some Theses of the Tractatus

Suszko was attempting to set out the ontology of the *Tractatus*. The language Suszko used included *(a)*the ordinary truth-functional calculus, enriched with quantifiers binding variables standing for sentences, and with an identity-function with sentences as arguments; *(b)* the ordinary theory of quantification applied to the special quantifiers; and *(c)* ordinary laws of identity applied to the special function. I propose to give a semantical account of the special quantifiers and the special function. I intend it to be non-referential. I do not intend to follow Suszko in setting out the ontology of the *Tractatus*. But I do intend it to incorporate some of that work's theses.

Here are some theses from Wittgenstein's *Tractatus Logico-Philosophicus*:

1. There are elementary sentences.
2. Elementary sentences are combinations of words.
3. If a word can occur in an elementary sentence, then it names some object.
4. The meaning of a word that can occur in an elementary sentence is the object it names.
5. An object named by a word that can occur in an elementary sentence is *simple*.

6. No elementary sentence names anything.
7. Each elementary sentence *says* that something is the case.
8. It is by means of the fact that its names stand to one another in certain relations within the elementary sentence that an elementary sentences *says* that something is the case.
9. An elementary sentence is true if what it *says* is the case, *is* the case, and is false if what it *says* is the case, *is not* the case.
10. What an elementary sentence *says* is the case *can* be the case, and also *can* fail to be the case.
11. If a sentence says something that *can* be the case, and which *can* also fail to be the case, then it is a truth-function of elementary sentences.
12. A non-elementary sentence that *fully* articulates what it says consists of names of the kind which can occur in elementary sentences combined into a whole with terms of logic.
13. No logical term names anything.
14. If a sentence does not say which *can* be the case, and which *can* fail to be the case, then it says nothing at all.
15. What a sentence says is something which can be expressed in various ways corresponding to various attitudes and interests, for example, as when we *ask whether* or *order it to be the case that* or *assert that* the door is closed.

I do not think any semantical account, even one which incorporates *Tractatus* elements, should try to incorporate all of these ideas. Thesis 11 is definitely false. Thesis 14 leads to the idea that neither necessary truths nor necessary falsehoods say anything. Since a true (false) sentence is one which says something true (false), this idea is incoherent. (This is argued by Hugly and Sayward in [1999]).

What of the other theses? First, there has been much thoughtful criticism of idea that the meaning of a name is the object it names. Second, since predicates must occur in elementary sentences, predicates must be names. How plausible is this? Finally, can anyone come up with a single example of an elementary sentence? Or a *simple* object? These are dubious ideas which Wittgenstein himself effectively criticized in the first sixty sections of the *Investigations*.

If we drop these dubious ideas we are still left with a lot. Every atomic sentence of a truth-functional language says something. What is said at this level determines what gets said by every sentence which is a truth-functional compound of the atomic sentences. A sentence is true if what it *says* is the case, *is* the case, and is false if what it *says* is the case, *is not* the case. If we now add a non-extensional connective such as 'John believes that' to the language, the result of appending that connective to a truth-functional sentence is true just in case John believes what is said by that sentence. These are ideas suggested by the *Tractatus*.

One of Suszko's thoughts is to add '=' to the language as a sentential connective. Then, for any two truth-functional sentences ϕ and ψ, $(\phi = \psi)$ is true just in case what is said by ϕ is the same as what is said by ψ. A second of Suszko's thoughts is to add sentential quantifiers to the language. Since sentences are not names (a thought shared by Wittgenstein and Suszko and many others), the sentential quantification introduced is not to be treated as

referential quantification. Finally, Suszo reads the sentential variables as taking situations as values.

Now the task is to put all of this to work to get a semantics for the resulting language.

3.Semantics

In what follows upper case letters 'S' and 'T' will function as sentential variables in the metalanguage, ranging over situations. The lower case letter 'α' will function as a variable in the metalanguage which ranges over sentential variables of the object language.

The language of truth-functional logic will be extended by the addition of the particular quantifier '∃' which binds the sentential variables and by the addition of two sentential connectives, '=' and 'B', which are not truth-functional. The first of these is a binary connective read as 'That __ is the same situation as that …'. The second is a unary connective read as 'John believes that __'.

Truth-functional formulas are understood in the usual way, using brackets, sentential variables, and a binary connective for the stroke function.

The formulas of the extended language are as follows (bold faced type functions as quasi-quotation):

(a) Any truth-functional formula is a formula.

(b) For any truth-functional formulas ϕ and ψ, **(ϕ = ψ)** is a formula.

(c) For any truth-functional formula ϕ, **Bϕ** is a formula.

(d) For any formulas ϕ and ψ, **(ϕ | ψ)** is a formula.

(e) For any variable α and formula ϕ, **∃α ϕ** is a formula.

(f) Nothing else is a formula.

A model for the language is an assignment of a situation to each variable of the language. More precisely, for each sentential variable α there is one situation S such that a model M of the language interprets α to say that S.

Each truth-functional formula says something relative to M (M-says something): (i) for any situation S, a sentential variable M-says that S if and only if M interprets it to say that S; (ii) for any situations S and T, and for any truth-functional formulas ϕ and ψ, **(ϕ | ψ)** M-says that neither S nor T if and only if ϕ M-says that S and ψ M-says that T.

Truth in M then runs as follows:

(g) A truth-functional formula ϕ is true in M if and only if, for some situation S, ϕ M-says that S and it is the case that S.

(h) For any truth-functional formulas ϕ and ψ, **(ϕ = ψ)** is true in M if and only if, for some situation S and for some situation T, ϕ M-says that S and ψ M-says that T and the situation that S is the same as the situation that T.

(i) For any truth-functional formula ϕ, **Bϕ** is true in M if and only if, for some situation S, ϕ M-says that S and John believes that S.

(j) For any formulas ϕ and ψ, $(\phi \mid \psi)$ is true in M if and only if neither ϕ nor ψ is true in M.

(k) For any variable α and formula ϕ, $\exists\alpha\,\phi$ is true in M if and only if ϕ is true in some model M* which differs at most from M in the interpretation of α.

The stipulation that, for any sentential variable, a model assigns *one* situation to that variable, assures that no truth-functional formula is ambiguous. That is, it follows that, for any truth-functional formula ϕ, there is one situation S such that ϕ says that S. This is a necessary requirement. For suppose that, for some truth-functional formula ϕ, ϕ said something that is the case and something else that is not the case. Would ϕ be true or false?

It may be objected that I have used connectives for situational identity and belief to give a semantical account of a language which includes such connectives. But that would be like objecting to the use of a connective for disjunction to give a semantical account of a language which includes a connective for disjunction.

It may be objected that I have used sentential quantification to give a semantical account of sentential quantification. But why is this objectionable? Referential quantification is used to give a semantical account of referential quantification in first order logic. Nobody objects. So what is objectionable in using sentential quantification to give a semantical account of sentential quantification?

4.Ontological Commitment

What it is for someone to be a realist about electrons, or numbers, or properties, or situations, and so on? I find it natural to answer this question thus: It is for that person to be committed to the existence of electrons, or numbers, or properties, or situations.

I agree that existence is expressed by such particular quantifiers as

There are…

Something…

There is something which…

There exists something which…

I disagree that whenever one uses such quantifiers to make assertions one asserts existence.

There is at least a *prima facie* case that not all uses of ‘there is’ and the like are ontologically committal. Consider the following sentences:

There is a question whether you are reluctant to speak up in this class.

There is more to life than philosophy.

There is a mistake in the derivation!

There is a way of putting the point that won’t hurt his feelings.

That a person assents to such sentences as these does not *itself* show that he or she is ontologically committed to the existence of questions, that which is more to life than philosophy, mistakes in arguments, and ways of putting points. There are indefinitely many such examples.

The question is when the operation of particular quantification forms sentences in assenting to which we commit ourselves to the existence of things.

I suggest the following answer: a particular quantifier asserts existence just in case it is a referential quantification (a quantification whose bound variable takes terms of reference, for example, names, as substituends).

Given (k) of the truth definition and the fact that neither '=' nor 'B' are truth-functional, sentential quantification is neither substitutional nor eliminable. Is it referential, then? Do the sentence letters of the object language take objects as values thereby becoming terms of reference? Unless there is some good argument that sentential quantification must be referential if not substitutional or eliminable, I have no reason to suppose the sentential variables of this object language take objects as values or function as terms of reference.

At the end of the second section I said that Suszko reads the sentential variables as taking situations as values. I intended the semantical account to incorporate this feature. But I see nothing about this that has ontological relevance.

We are inclined to speak of the values of the variables, and to speak of what the variables range over. In this connection we distinguish the values of a variable from its substituends, the expressions which can replace it.

In one case the values of the variables are people and the variables range over people. In another case the values of the variables are situations and the variables range over situations. In both cases the variables are letters. Now what is it to have a value — to have a person or a situation as a value? And is having a value the same in both cases?

In terms of models and truth relative to models, the idea of *ranging over* is expressed by the role variant models play in defining truth relative to a model for the quantifications. A variable has one value relative to one model, another value relative to one of its variants, a third value relative to another of its variants, and so forth. This holds independently of what it is for a variable to have a value or a value of this or that sort. So *ranging over* can be said to be the same for sentential variables and variables for people.

A natural picture for *ranging over* would be that of a gesture of hand in respect to a group of people. And that carries with it a picture of having a value: we think of the people as values, and the gesture over them is like an ever so rapid pointing at each. But the sober fact is that there are many variant models and in each there is a particular variable with a particular value. That is all there is to *ranging over*.

REFERENCES

Church, A. (1956). *Introduction to Mathematical Logic, Vol. 1* Princeton, New Jersey: Princeton University Press

Cresswell, M. J. (1967). 'Propositional Identity'. *Logique et Analyse* 40: 283-291.

Hugly, P. and Sayward, C. (1999). 'Null Sentences'. *Iyyun, The Jerusalem Philosophical Quarterly* 48: 23-36.

Malinowski, G. (1985). 'Non-Fregean Logic and Other Formalizations of Propositional Identity'. *Bulletin of the Section of Logic* 14: 21-29.

Quine, W. V. O. (1961). *From a Logical Point of View*. Second Edition, Revised. Cambridge, Massachusetts, London: Harvard University Press.

Suszko, R. (1968). 'Ontology in the Tractatus of L.Wittgenstein'. *Notre Dame Journal of Formal Logic* 9: 7-33.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus.* Introduction by Bertrand Russell. Translated by C. K. Ogden. London: Routledge & Kegan Paul.

Wójcicki, R. (1984). 'R. Suszko's Situational Semantics'. *Studia Logica* 43:323-340.

Charles Sayward

UNL (University of Nebraska Lincoln)

csayward@unlserve.unl.edu

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp. 50-55
Miller's Defence of Bartley's Pancritical Rationalism

MILLER'S DEFENCE OF BARTLEY'S PANCRITICAL RATIONALISM

Armando Cíntora

To the memory of Karl Popper in the centenary of his birth.

I. Argumentative Background

Popper characterized uncritical rationalism as follows:

> Uncritical or comprehensive rationalism can be described as the attitude of the person who says 'I am not prepared to accept anything that cannot be defended by means of argument or experience'. We can express this also in the form of the principle that any assumption which cannot be supported either by argument or by experience is to be discarded. Now it is easy to see that this principle of an uncritical rationalism is inconsistent; for since it cannot, in its turn, be supported by argument or by experience, it implies that it should itself be discarded. (…) Uncritical rationalism is therefore logically untenable… (Popper, 1945, p. 217.)

Uncritical rationalism can be analysed in terms of the following two injunctions:

C1: *Anything* (belief, action or aim) justified by non-viciously circular argument is to be accepted as rational.

C2: *Only* that which can be justified by non-viciously circular argument is to be accepted as rational.

Now, from C2 it follows that if rational one should justify anything one accepts. In particular, one should justify C1 and C2, but C1 cannot be justified without presupposing it, because if one were to try to justify C1 by arguing in its favour, we would be begging the question, that is, we would be presupposing C1. In other words, one would be presupposing that argument and experience are rational justificatory strategies, precisely what C1 says.

Therefore, C1 and C2 cannot both be true.

Let's then first assume we discard C1 and try to keep C2. Then C2 should be justified in a non-viciously circular way, but C2 cannot be justified by non-viciously circular argument, because if we offer an argument in its favour we are again presupposing that argument is valuable as a rational justificatory strategy, something that C2 assumes. Therefore, C2 rejects itself and uncritical rationalism is logically impossible.

Another option would be to discard C2 and keep C1 just by itself, this alternative doesn't lead to a contradiction it just says that justified beliefs are rational, but it doesn't require us *qua* rationalists to justify all our accepted beliefs, actions or goals. Rationalism without C2 would not aim to be comprehensive and it would be a very weak form of rationalism, so weak that according to C1 a non-justified (or an unjustifiable) absurd belief could still be rational.

Because of all these problems with uncritical rationalism, Popper proposed an alternative: critical rationalism,

> (…) whoever adopts the rationalist attitude does so because **without reasoning** he has adopted some decision, or belief, or habit, or behaviour, which therefore in its turn must be called irrational. Whatever it may be, we can describe it **as an irrational *faith in reason*.** Rationalism is therefore far from comprehensive or self-contained.
>
> (…) **a critical form of rationalism**, one which frankly admits its limitations, and its basis in an irrational decision, and in so far, **a certain priority of irrationalism.** (Popper, 1945, p. 218; emphasis added.)

The critical rationalist and the irrationalist differ in having different irrational commitments, but they also differ in that the critical rationalist intends to minimize her irrationalism; thus, the critical rationalist argues that a minimal form of irrationalism is *morally* preferable to exuberant forms of irrationalism.

Thus, Popper claims that critical rationalism with his minimum of irrationalism is preferable, because,

> The choice before us is not simply an intellectual affair, or a matter of taste. It is a moral decision.
>
> … It is my firm conviction that this irrational emphasis upon emotion and passion leads ultimately to what I can only describe as crime. One reason for this opinion is that this attitude, which is at best one of resignation towards the irrational nature of human beings, at worst one of scorn for human reason, must lead to an appeal to violence and brutal force as the ultimate arbiter in any dispute. (Popper, 1945, pp. 219-21.)

In other words, Popper believes that while the rationalist may not be moral,[1] the irrationalist is often immoral, and that this gives us one reason to prefer critical rationalism (with its minimal irrationalism) to full fledged irrationalism. Popper then transfers the decision for critical rationalism to the uncertain domain of moral judgement. On the other hand, Popper holds that moral judgements can be influenced, though not determined, by a *rational* analysis of the practical consequences of our moral decisions, and by contrasting these practical consequences with the prescriptions of our conscience. Popper says,

> (…) **a rational analysis** of the consequences of a decision does not make the decision rational; the consequences do not determine our decision; it is always we who decide. But an analysis of the concrete consequences, and their clear realization in what we call our «imagination», makes the difference between a blind decision and a decision made with open eyes (…) in the case of moral theory, we can only confront its consequences with our conscience. (Popper, 1945, p. 220.) (Emphasis added.)

Notice that such a rational analysis already presupposes a favourable valuation of an argumentative or rational attitude, therefore a pondered moral decision about whether to be critically rational or not will itself *presuppose a favourable valuation of an argumentative attitude*. Then a moral decision in favour of critical rationalism would be circular, therefore, a moral judgement in favour of critical rationalism is in the end justificatory useless, but this is all as well, since the critical rationalist openly admits that she has to assume dogmatically her high valuation of argument.

On the other hand, it could well happen that some consciences would prefer irrationalism even when fully aware of its probable immoral consequences, since it is almost a truism that

---

[1] The rationalist, however, may need to be moral in a minimal sense of morality, given that rationality is a cooperative process of inquiry which values dialogue, intellectual honesty and humility, respect for other people's arguments, etc.

different individuals or communities often don't agree on what constitutes a crime. This becomes clear when one considers the examples of many Nazi SS who would not consider the Auschwitz camp as criminal, or of many XVIII th century slave traders or owners who neither would deem his metier as criminal. And given this possibility one can only hope that the conscience of most of us will as a matter of fact side with critical rationalism, but if this were not to happen, that would be the end of the matter for a Popperian.

A crisis of integrity arises, however, for the critical rationalist, since her rational identity requires a leap of faith, which by her own lights is irrational or at least non-rational, and this, then provides a rational excuse for all kinds of irrational commitments; it supplies the irrationalist with the *tu quoque* argument, an argument that says:

> … (1) because of logical reasons, rationality is so limited that everyone must make a dogmatic irrational commitment; (2) therefore, the irrationalist (Christian, or whatever) has a right to make whatever commitment he pleases; and (3) therefore, no one has a right to criticize him (or anyone else) for making such a commitment… (Bartley, pp. 272-3.)

The *tu quoque* tells us that given that rational argument about ultimate commitments is impossible, then any commitments are rationally possible.

II. Bartley's Pan Critical Rationalism

W. W. Bartley thought it was possible, however, to reform Popper's critical rationalism into a consistent and comprehensive theory of rationality («pan critical» rationalism: PCR, also called comprehensive critical rationalism: CCR.) Bartley claimed that it was possible to reform critical rationalism into a theory that allegedly does not lead into a fideism of ultimate commitments. Bartley proposed a new rational identity one that allegedly does not lead into conflicts of rational integrity. Bartley's pan critical rationalist can be characterized as one,

> … who is willing to entertain any position **and holds all his positions, including his most fundamental standards, goals, and decisions, and his basic philosophical position itself open to criticism; one who protects nothing from criticism by justifying it irrationally**; one who never cuts off an argument by resorting to faith or irrational commitment to justify some belief that has been under severe critical fire; one who is committed, attached, addicted, to no position. (Bartley, p. 118; emphasis added.)

This pan critical rationalist justifies nothing and allegedly criticizes everything, even his own rational attitude or position, he is not committed to any position, not even to a belief in the value of argument. This doesn't mean that the PCrationalist is without convictions, but only that he is willing to submit his convictions to critical consideration. PCR, however, leads to logical paradox, thus consider the following argument, due to Bartley himself and inspired by a critique of J. F. Post, an argument that Bartley finds unobjectionable:

(A) All positions are open to criticism.

And because of PCR's intended comprehensiveness it then follows,

(B) A is open to criticism. And,

> Since (B) is implied by (A), any criticism of (B) will constitute a criticism of (A), and thus show that (A) is open to criticism. Assuming that a criticism of (B) argues that (B) is false, we may argue: if (B) is false, then (A) is false; but an argument showing (A) to be false (and thus criticizing it) shows (B) to be true. Thus, if (B) is false, then (B) is true. Any attempt to criticize (B) demonstrates (B); **thus (B) is uncriticizable, and (A) is false.** (Bartley, p. 224.) (Emphasis added.)

Hence, PCR is refuted and this conclusion is a result of the self-referential character of PCR — a theory that intends to be a theory of all theories itself included, and it recalls the

logical difficulties of classical rationalism, which also wanted to be comprehensive. Bartley claims that the paradoxical nature of PCR could be dealt,

> …by type and language-level solutions, Zermelo-type solutions, category solutions, radical exclusion of all self reference… (Bartley, pp. 219-20.)

But, this is too vague, mere possibilia.

III. Miller's Defence of PCR

David Miller outlaws self-reference (and in this way he avoids PCR's paradoxical nature) by distinguishing between positions and statements and declaring that (B) is just a statement — and not a position — and as such it is not in the domain of (A), that is, Miller claims that (B) — being a mere statement — doesn't have to be criticizable on its own,

> …I reject the thesis that criticizability is an automatic property of all statements. It is not an intrinsic property of statements at all, but an honour that must be bestowed on them by the development of appropriate methods of criticism. How, it may be asked, is this to be done? In many cases the answer can be only: **by a consideration of the problems that provoked them**…
>
> **CCR must not be understood to hold that every statement that a comprehensively critical rationalist counts as true (rationally accepts) is on its own criticizable.**
>
> …As far as statements… are concerned, what is important for the rationalist, I suggest, is that each statement that he accepts either is itself criticizable **or follows from a statement** that he accepts that is citicizable. Any position adopted must be criticizable, but it is no concession to the irrationalist to allow that some logical consequences of the position may not be criticizable. (Miller, pp. 86, 89.) (Emphasis added.)

Miller's proposal is motivated by the fact that falsifiable statements can entail unfalsifiable ones,

> All falsifiable hypotheses have amongst their consequences a host of unfalsifiable statements (ranging from tautologies and unrestricted existential statements to meaty metaphysics) that enter science as it were on the coat-tails of their parents. (Miller, p. 10.)

Metaphysical determinism is, for example, a consequence of Newtonian theory and in the case at hand, *mutatis mutandi*, an uncriticizable statement is allegedly entailed by a criticizable position. Thus,

> [(A)] is a position that Bartley recommends that we adopt, and it is essential that it be criticizable. But [(B)] is just a consequence of it — an interesting consequence, in the light of what CCR says, and (one hopes) a true consequence; but it cannot be taken up as an independent position … although [B] is a possible position on its own (though a strange one), **it cannot be adopted at the same time as [(A)] is. Nor, of course, is there the slightest need for it to be adopted along with [(A)], which brings it along for nothing.** (Miller, p. 90.)(Emphasis added.)

IV. Critical Comments to Miller's Defence of PCR

1) Miller's demotion of (B), **a core** assumption of the PCrationalist, to statement status looks, however, suspiciously like an *ad hoc* manoeuvre introduced **only** to avoid logical paradox. If not, how does Miller's hypothesis increase the criticizability of PCR? Moreover, beyond the avoidance of paradox why should we not take (B) as a position simultaneously to (A)? In other words, is there a criterion that will allow us to distinguish between positions and statements? Or when do positions entail positions, and not just statements? And which statements are criticizable on their own, and why?

If there is no adequate answer to these questions, then it looks that the distinction between positions and statements will be settled by someone's (either an individual's or a community's) idiosyncratic judgment.

Miller has not offered any theory of criticizability that would provide some sort of demarcation criterion between positions and non-criticizable statements. Except for a hint at a pragmatic criterion, thus, Millller claims that in many cases the criticism of statements can only be done by a «consideration of the problems that provoked them.» (Miller, p. 86, quoted above.)

In contrast, Popper proposed a theory of falsifiability in which, for example, «pure existential», «all-some» and tautological statements were unfalsifiable due either to their syntactical or semantical structure.

2) Furthermore, when Miller says that there is not «the slightest need for (B) to be adopted along with (A), which brings it along for nothing», he seems to be giving up the comprehensive intention of PCR, a critical comprehensiveness which was PCR's main claim and aim and which distinguished it from the older non-comprehensive Popperian critical rationalism. If PCR abandons its comprehensive character it is giving up what was allegedly its great virtue, what was its main advantage over critical rationalism. This diminished PCR could then be confronted by a new *tu quoque,*

   If you pancritical rationalist don't hold as criticizable some of your **core** rational assumptions (or if you prefer statements), then neither I, an irrationalist, need open to criticism some of my **core** dogmatic suppositions or statements.

Now, in the case at hand, **the problem that motivated PCR** and in particular A and therefore B was the *tu quoque* argument made possible by the leap of faith required by critical rationalism. Therefore, once one realizes what B says (i.e., once one realizes that B is a core assumption of PCR and not just any statement inferred from A) and if we are to avoid a new *tu quoque*, we should hold B open to criticism, if not, B would have a dogmatic character.

3) Finally, Miller claims, that,

> The rationalist is not at all obliged to try to criticize all the consequences of his ideas, and if, as may happen, some of them are not open to criticism at all — which may mean no more than that **he cannot think of any way, even potentially, of criticizing** them – that need not disturb him… (Miller, p. 90.) (Emphasis added.)

Notice, however, that the paradoxical nature of PCR is an objective property of PCR, i.e., a property PCR has in Popper's world III. Therefore, PCR's paradoxicalness cannot be avoided by the pan critical rationalist's critical attitudes or thoughts (or lack of these attitudes or thoughts) concerning A and/or B, since these thoughts and attitudes are in Popper's world II while PCR is an autonomous object of world III. In other words, the paradoxical nature of PCR wouldn't disappear even if (as recommended by Miller) all of the PC rationalists were to decide not to hold A and B as simultaneously criticizable.

V. Conclusion

Because of the paradoxicalness of PCR (plus other various well known difficulties of this doctrine, cf., for example, Watkins) it seems that the best option *malgré tout* is Popper's critical rationalism with its minimum of irrationalism, with its dogmatic faith in reason.

Now, if Popper's critical rationalism requires a leap of faith in favour of reason why not accept also other leaps of faith, for instance, a leap of faith in favour of induction? In other words, the question now arises of how minimal can and should be the irrationalism required by critical rationalism. For example, if Popper's methodology requires a «whiff of inductivism» (Popper in Schlipp, p. 1192), then why not accept more than a mere whiff? J. Worrall has argued, for example, that a minimal methodological dogmatism (beyond Popper's faith in reason) is logically unavoidable in science.

References

Bartley, W. W. 1984: *The Retreat to Commitment*. Open Court. Illinois.

Miller, D. 1994: *Critical Rationalism. A Restatement & Defence*. Open Court, Illinois.

Popper, R. Karl. 1945: *The Open Society and its Enemies*. Routledge. London.

Post, F. J. 1971: 'Paradox in Critical Rationalism and Related Theories'. In Evolutionary Epistemology, Rationality, and the Sociology of Knowledge, edited by G. Radnitzky and W. W. Bartley III. Open Court. Illinois. 1988.

Post, F. J. 1987: 'A Goedelian Theorem for Theories of Rationality' in Evolutionary Epistemology, Rationality, and the Sociology of Knowledge, edited by G.Radnitzky and W. W. Bartley, III. Open Court. Illinois. 1988.

Radnitzky, G. & Bartley, W. W., III, eds. 1987: *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*. Open Court.

Schilpp, P. A. (Ed.) 1973: *The Philosophy of Karl Popper*. 2 vol., Open Court Publishing Co., Illinois.

Watkins. J. W. N. 1987: Comprehensively Critical Rationalism: A Retrospect. In «Evolutionary Epistemology, Rationality, and the Sociology of Knowledge.» In Radnitzky and Bartley eds. Open Court, 1987.

Worrall, John. 1988: 'The Value of a Fixed Methodology'. *Brit. J. Phil. Sci*. 39.

Worrall, John. 1989: 'Fix it and be damned: A Reply to Laudan'. *Brit. J. Phil. Sci*. 40.

Armando Cíntora

Departamento de Filosofía, UAM-I, México City

cintora@prodigy.net.mx

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp. 56-66
On Quine's Arguments Concerning Analyticity

ON QUINE'S ARGUMENTS CONCERNING ANALYTICITY

Shaun Baker

In his *Two Dogmas of Empiricism*, Quine argues that the traditional distinction between statements that are analytically true and statements that are true by virtue of matters of fact is a metaphysical article of faith. He argues that the distinction between the two sorts of statements has never been adequately set out. In this paper, I will attempt to present Quine's arguments and provide evaluations.

Quine sets up the problem of analyticity as being a question of coming up with an adequate account of how one problematic class of statements that have traditionally been considered analytic can be reduced to another less problematic class of statements.

The first class is typified by the statement No bachelor is married, the latter class by No unmarried man is married. The initial stages of argumentation seem to be as follows:

1. Statements traditionally considered analytic come in two general varieties. They are either (a) logically true, or (b) statements that can be turned into logical truths by substituting synonymous expressions.

2. Statements of type (a) are formally true. Any interpretation of the non-logical terms will give a reading that remains true. The non-logical terms in fact drop out of the explanation of these statements' truth.

3. Statements of type (b) can be turned into statements of type (a) only by substituting a term that means the same thing as (that is synonymous with) one of the non-logical terms. That term which is substituted must have the untouched term as a logical constituent.

4. In (a) type statements, the meaning of the non-logical terms are identical, and drop out as irrelevant to the truth of those statements. But since we are left with the meaning relations between the logical constituents of the sentences, these statements can be considered analytic of the non-logical terms that they contain.

5. In (b) type statements, the meanings of the non-logical terms are relevant to the truth of those statements. Only if we can say that one of those terms is synonymous with some other term that upon substitution would produce an (a) type statement, can we say that a (b) type statement is analytic.

With this as background, Quine moves into the next phase of argumentation. His overall goal is to show that we cannot give an account of analyticity. The form of his argument is disjunctive. He presents several possible accounts of analyticity. On the assumption that his disjunctive premise is exhaustive, he concludes that no account of analyticity is possible, and for us to continue to believe there are such things as analytic truths is to engage in blind faith.

His first few disjuncts are based upon the initial argumentation above. He argues that no account of synonymy is adequate. If that is true then no account of analyticity in terms of synonymy is going to be adequate. He then argues that no account of analyticity that is carried out independently of an account of synonymy will succeed either. I will now examine his arguments concerning synonymy. After that, I will look at the argument that is concerned with synonymy independent accounts of synonymy.

Synonymy

The terms in need of explication in (3) above are 'same in meaning' and 'synonymous'. Quine argues that we cannot give an adequate account of these terms. He argues disjunctively: he offers two possible ways to account for synonymy between terms: The first possibility is that synonymy comes about by one of three sorts of definition. The second possibility is that synony my between terms is simply the f act that they are interchangeable in all sentential contexts *salva veritate*. He argues that neither of these approaches will give an adequate account of what synonymy is. Assuming that the disjunctive premise is exhaustive, and he is correct about each of the disjuncts, he concludes that there can be no adequate account of synonymy. Since there can be no adequate account of synonymy, there can be no adequate account of how statements of type (b) are analytic if that account must depend on an account of synonymy.

In regard to the three types of definition, Quine argues that two types in some way presuppose synonymy without explaining what it is.

The third sort of definition does indeed create genuine synonymy relations, but Quine seems to think that these cases are in some way importantly different from most of the alleged cases of synonymy.

Therefore, this sort of definition cannot serve as the basis of a general account of synonymy.

In regard to the interchangeability thesis, Quine argues that it ultimately must involve some sort of presupposition of analyticity to work out. The interchangeability thesis fails because it ultimately has to make use of the notion of analyticity in order to make sense of a language that uses the modal operator 'it is necessarily the case that'. That language is apparently the minimal necessary language in terms of which the interchangeability *salva veritate* account of sameness in meaning can be insulated from counterexamples based upon extensionally equivalent terms.

I start by presenting the initial steps of Quine's argument. I will include all the disjuncts he considers in the article. I will however first concentrate on the portion of his argument concerning definitions, interchangeability *salva veritate*, and synonymy. Later, I will look at his argumentation concerning a synonymy independent account of analyticity:

1. Analytic truths of type (b), if they are to exist must be such that they can be turned into analytic truths of type (a) by making use of pairs of synonymous terms.

2. Synonymous terms are synonymous by virtue of either (c) one term being defined in terms of the other, or (d) being interchangeable *salva veritate* in all sentential contexts, or, (e) by being true according to the semantic rules of the language of which they are a part.

3. If terms are synonymous by virtue of (c), the definition has to be one of three types: lexicographic, or explicative/ampliative, or conventional.

4. If the definition is lexicographic, it is a report of pre-existing usage.

5. The lexicographer's report of pre-existing linguistic behavior cannot serve as the explanation or ground of that pre-existing behavior. His report merely records the pre-existing behavior.

6. Therefore, terms cannot be synonymous in virtue of lexicographic definition alone.

7. If the definition is explicative/ampliative, it is a modification or explication of some pre-existing synonymy relation. The person or persons making this definition supplement or refine the pre-existing synonymy for some purpose or another.

8. But if what (7) says is true, then some synonymy is already being made use of. The explicative/ampliative definition is dependent for its existence upon that pre-existing synonymy.

9. If (8) is the case, then the explicative/ampliative definition in question does not introduce the pre-existing synonymy, or account for its existence.

10. So, there is at least one synonymy whose existence cannot be accounted for in terms of this one explicative/ampliative definition.

11. But suppose that the pre-existing synonymy itself was the result of some earlier explicative/ampliative definition.

12. If this earlier event was indeed an explicative/ampliative definition, then it too must have been based upon some earlier synonymy relation.

13. If explicative/ampliative definition is to be used to account for all synonymy relations, then we land in an infinite regress.

14. So, explicative/ampliative definition cannot do as a general account of how synonymy relations come into being. Explicative/ampliative definition cannot give an account of synonymy.

15. If the definition is conventional, then we have a case where two terms are set as equal, one term being introduced as abbreviatory of the other. E.g., 'e.g. =df for example'.

16. If we have a case where two terms are being set as equal, one being abbreviatory of the other, then we have a case where a synonymy relation has been created by definition.

17. Conventional definitions create genuine synonymies.

Even though he allows that conventional definitions can create synonymous pairs of terms, Quine still thinks that an adequate general account of synonymy has not been produced. It seems that he want an account that can be made use of in explaining synonymies that exist in natural languages. He does not think that this can be done by conventional definition alone. Perhaps it can be carried out via some sort of combination of conventional and explicative/ampliative definition. It remains to be seen if this is true. I will spell out a way this might be done after I have presented Quine's argument concerning the impossibility of a synonymy independent account of analyticity. However, for now I want to look at the argument concerning the inadequacy of interchangeability as an account of synonymy.

Once Quine dismisses definition as an adequate account of synonymy, he moves on to interchangeability *salva veritate*.

Interchangeability *salva veritate* by itself is not sufficient for sameness of meaning for two reasons: One can substitute extensionally equivalent terms and preserve truth, while one does not preserve meaning. So, if we have a language that deals only with one place properties, two or more place relations, and contains the truth functional operators, singular terms, variables for those singular terms, and quantifiers, we will not have enough to guarantee that if we are to replace one term with another, while preserving truth in some statement, that we will have also preserved the original meaning of that statement. It may be true to say that all creatures with hearts live on earth, true that all creatures with hearts are creatures with kidneys, and thereby true that all creatures with kidneys live on earth, but the first and latter sentences do not seem to mean the same thing. So it seems relative to this sort of language, interchangeability in not going to be sufficient for synonymy.

But perhaps we can enrich our language. We can add what Quine calls 'intensional operators' to our language. These would be things like cognitive operators, modal operators, and the like. Quine points out the operator most likely to help us in g iving an account of analyticity: the modal operator 'it is necessarily the case that'.

Necessary truths are those that are true 'come what may'. Leibniz took this to mean that they are those statements that are true in every possible world. We can think of possible worlds as possible circumstances. So necessary truths are those that are true in all possible circumstances. this would seem to indicate that their truth is only an internal matter. The circumstances simply do not matter. So it would seem that necessarily true statements are true by dent of a feature havin g to do only with themselves. T hat feature cannot be the extensions of the necessarily true st atements, should they have them. Therefore, it must be their meanings that are the key feature.

So maybe if we work with a sentential matrix like 'Necessarily, if x is an A, then x is a B', as a sort of test for synonymy we can find those synonymous pairs of type (b) after all. For only if the non-logical terms are exactly the same in meaning will they drop out and become irrelevant to the truth preservation of the statements we are using as test cases. If we compare the statement Necessarily, if x is a creature with a heart, then x is a creature with a kidney, with the statement Necessarily, if x is three in number, then x is odd, we see how plausible this sounds. This sort of approach also seems to work when we plug in bachelor, and unmarried.

However, Quine thinks that doing this ultimately lands us in an account of analyticity through interchangeability that involves us in some sort of circular account. His argument is difficult to follow, but it seems to be the following:

Interchangeability

1. In the simpler first order language it is possible for universal claims to preserve truth upon interchange of terms, even when we would say that the meanings of the terms are different. All that needs be the case is that the terms are extensionally equivalent: e.g.; The true statement 'All and only creatures with hearts are creatures with hearts' goes into the equally true statement 'All and only creatures with hearts are creatures with kidneys'.

2. So, for any pair of universal claims that are both true, and one of which is the result of such an interchange of terms, it is possible that they differ in meaning while they agree in extension. They could both be true by accident, by dint of a matter of fact.

3. 'All and only bachelors are bachelors,' and 'All and only bachelors are unmarried males,' are a pair of universal claims, both true, and one of which is the result of interchange of the terms 'bachelor' and 'unmarried male.'

4. It is possible that the pair of universal statements in 3 differ in meaning, while agreeing in extension. It is possible that they are both true by accident, by dent of a matter of fact.

5. If it is possible that of a pair of universal statement which satisfy the description in (2) that one of the pair is true by dint of a matter of fact, then it is possible that they are not both analytic truths.

6. Only if it is not possible for a pair of universal statements like those described in 2 to differ in meaning, while agreeing in extension, and thus it is not possible for one of the pair to be true by dent of a matter of fact, will both statements be analytic truths.

7. Perhaps the following is true. It does seem that Quine's (a) type example and the three and oddness example successfully meet the following condition: Only synonymous terms will fill in a matrix like 'Necessarily, if x is an A then x is a B'. For example, if we put in 'creature with a heart' and 'creature with a kidney' for A and B, respectively, then we create a false statement from a true one. It is logically and physically possible for a creature with a kidney to have no heart, and for a creature with a heart to have no kidney. We can conceive of such beings, and we are familiar with people on dialysis machines, or who have artificial organs. On the other hand, if we substitute 'creature with a heart' for both A and B, we get a true statement. No matter of fact can refute that statement, and we cannot conceive of a matter of fact that could refute it.

8. But the proposal will work only if the operator it is necessarily the case that does not in some way presuppose a notion of analyticity.

9. But to say that something is necessarily the case is to say that it could not have been otherwise.

10. To say that something could not have been otherwise is to say that it would have been the case no matter what other facts might obtain.

11. To say that something would have been the case no matter what other facts might obtain is to say that the other facts are irrelevant to its truth.

12. But if something is the case regardless of all other facts, then it must be the case because of facts internal to it.

13. Therefore, necessarily true statements are true only because of facts internal to them.

14. Statements have extensions and intension s. These are the only 'facts' cognitively relevant to statements.

15. Whether or not a statement has an extension is a matter of whether or not its constituents have extensions. Whether or not words have extensions depends on what other facts obtain. Therefore, whether or not a statement has an extension depends on what other facts obtain. These other facts are not internal to the statement.

16. Therefore, necessarily true statements are not true by virtue of their extensions.

17. Intensions are the linguistic meanings of words.

18. Words can have linguistic meanings even when they lack reference.

19. Therefore, intensions are something a word can have independently of any matter of fact external to their own meaning.

20. But if linguistic meanings are something words can have independently of any matter of fact external to their own meaning, then statements can have meanings independently of matters of fact.

21. So, the only types of statements that can be necessarily true are those that are true only by virtue of the meanings they have.

22. But the only type of statement that is true only by virtue of the meanings they have are analytic statements.

23. Therefore, to say «Necessarily A is a B» is to say something like «It is analytically true that A is B».

24. But we are presupposing analyticity in order to be able to make sense out of the sentential operator it is necessarily the case that, which we were to use in a non-circular manner, in order to sift out analytic truths.

25. Therefore, we cannot produce an adequate account of analyticity in terms of interchangeability of terms *salva veritate*.

Semantic Rules

Quine now abandons the attempt to produce an account of synonymy, and attempts to give an account of analyticity in terms that are independent of synonymy. He makes use of the notion of semantic rules of languages. To simplify matters he considers artificial formal languages.

The general idea is that the analyticity of a statement is seen as relative to the semantic rules of the language of which it is a part. We might use an example: In pure propositional logic, the semantic value of a statement is one of two truth-values : 'true' or 'false'. There are simple statements, and compound statements.

Compound statements are created by concatenating simple statements using truth functions. (i.e., the connectives &, v, ~, etc.). Each truth function is assigned a truth table. The truth tables give semantic rules by which one can determine the truth-value or semantic value of any truth functionally compound statement that is created by use of the connectives. Analytic statements could be specified as those that describe applications of the semantic rules (the truth tables) which come out true on every possible concatenation of semantic values possible for the propositional variables in the statement. By that rule 'p' in not an analytic statement of propositional logic, but 'p∨~p' is.

Here is what Quine says about this:

> A semantical rule of this … type, a rule of truth, is not supposed to specify all the truths of the language; it merely stipulates, recursively or otherwise, a certain multitude of statements which, along with others unspecified, are to count as true. Such a rule may be conceded to be quite clear. Derivatively, afterward, analyticity can be demarcated thus: a statement is analytic if it is (not merely true but) true according to the semantic rule.

Quine now complains that in such a definition, one unexplained term is being substituted for another. Assuming 'truth' is not a problematic term, 'analytic' has been replaced by 'true

by virtue of a semantical rule'. Using the example, we can give a semantic rule for propositional logic as follows: All statements that are of the form 'p v ~p' are true. We might be inclined to say that this is a statement of a semantic rule of propositional logic. But Quine tells us that this statement is most generally described as a statement which says of a certain set of statement types that they are true. But not every statement that says of a certain set of statement types that they are true is happily described as a statement of a semantic rule. I might say, (and it might be the case that) all the predi ctions of the oracle of Delphi are true, but this would not seem to be a statement of a semantical rule of any language. Neither would my saying this sort of thing make it true that the oracular utterances are all analytically true. If all statements which say of a certain set of statement types that they are true were semantic rules, then all truths would be analytic. Clearly, it must be the case that some subset of these sorts of statements are statements of the semantic rules of languages and some are not. But what property sets off the favored subset from its brethren? Maybe that property is the fact that the favored statements point out truths that result only from the primitive postulated semantic rules (such as the truth table of disjunction in our example). The 'analytic making property' would be something like 'being a sentence form that receives a semantic value assignment true each time it is evaluated using the primitive semantic rules, and receives this assignment regardless of the truth value assignments of the atomic statements. Quine is not satisfied with this sort of move, and thinks that ultimately it cannot be used to give an account of analyticity. His reason seems to be the following: When we set up a formal language like propositional logic, we start by postulating some set of semantic rules as basic. Others can be defined in terms of the postulates. In formal languages, we are interested in statements in so far as they can be derived form other statements in accord with transformation rules. But it is open to us which semantic rules we treat as the basic set. E.g., we can treat rules involving negation and disjunction as the primitive semantic rules, and can then define other semantic rules involving, for example, conditionals in terms of the primitive rules we have chosen.

We might even go the other way, treating as basic the conditional, and negation, and define disjunction in terms of these. It is also possible that we could try and start with one semantic rule, perhaps the 'Sheffer stroke' and attempt to define a large set of connectives in terms of it. Conversely, we could start with a rich language, containing many basic semantic rules, and define certain long formulae using the Sheffer stroke in terms of the many postulated semantic rules.

In the former kind of case, the non-basic semantic rules are derived. They can be described as abbreviatory conventions. The language could do without them. It would just be more cumbersome without them. In those sorts of languages, it seems that the derived rules are not really in an ultimate sense basic semantic rules of the language. They are not among the postulates. But since it is up to us which semantic rules will be basic, there is no sense in saying that there is in any sort of task independent sense, a privileged set of statements that follow from the basic semantic rules of propositional logic. Relative to task A, semantic rule x can be treated as being a postulate, but relative to another task B, x may be treated as being defined in terms of and being dependent upon other semantic rules that are being used as postulates. T here are many possible purely logical tasks that can be carried out using a formal language. So if 'x is analytic' means 'x is invariably true by the basic semantic rules of L', means 'x's invariable truth is the result only of an application of a postulate of L', then any invariable truth of propositional logic could conceivably be the result merely of an application of some postulate of some L which is family related to propositional logic. It would thereby be an analytic truth by the basic semantic rules of that L.

So it seems that the following is the case: If the analyticity of a formula of some formal language L is to be defined in terms of whether or not it is (1) a statement which is true only because of the fact that it turns out true on every possible assignment of truth values to its propositional constituents, and (2) it uses only the primitive semantic rules of that language to determine if (1) is the case, and (3) what rules are treated as primitives, or postulates is something that is relative to the tasks that the creator of that system has in mind, then (4) it is true that analyticity, so defined is relative to languages. However, this does not seem to show that there are no analytic truths, as Quine seems to maintain.

But how can these considerations be applied to ordinary language? I think they can be applied in such a way as to throw doubt on Quine's strong position. Earlier, I mentioned in connection with the discussion of definition, that there might be a way of combining conventional definition, and explicative/ampliative definition so as to give some account of analyticity. Now I will try and sketch this out.

Because the relative richness of the concepts used in a language are in some way relative to tasks undertaken via that language, what may be a deductive consequence of the meaning of a word used to designate a concept of one language may not be a deductive consequence of the meaning of that same word as used in a richer language. But if there are deductive consequences of the meanings of words, then there are analytic truths. True, there cannot be analytic truths in some sort of absolute sense, but this does not seem to count against the thesis that there are at least some analytic truths. Quine claims that there are no analytic truths. To prove, as his semantic rules argument seems to, that analytic truths, if there are any, are in some way language dependent, and task dependent does not establish the stronger point. It still seems that relative to a given task, and a given way of conceptualizing a situation, that there will be some deducible consequences of that conceptualization. If the deduction of such consequences is not an analysis of the postulates, or conceptual underpinnings of that language, then what else could it be? This theory of language relative analytic truths may allow us to deal with the alleged counterexamples that are aimed at traditional examples of analytic statements. One such traditional example is the statement All bachelors are unmarried male humans of marriageable age.

Psychologists have found that people, if asked, will say that the Pope is not a bachelor. This is true despite the fact that he is an unmarried male human of marriageable age. People will also refuse to label a man who has lived in the same house as a woman with whom his is not wed with the term. Also, extremely old single men are not counted as bachelors (the Pope is once again a good example of this).

All of this is supposed to count against reading the universal generalization above as an analytic truth concerning the word 'bachelor'.

It seems that there are two distinct ways to respond to this claim.

(1) 'Bachelor' just means *unmarried human male of marriageable age,* and the counterexample shows that within that broad category there are subspecies. People may have in mind some rather typical examples of the species when they are asked to answer the questions concerning atypical examples. Because they have these typical examples in mind, more so than examples of the atypical types, they make these judgments. The empirical results do not show that there are no analytic truths concerning the term 'bachelor', they just show that people can be led into error by psychological factors. It seems that empirical results could be produced that would corroborate this view. If the psychologists had asked their subjects to

think carefully and tell them whether or not 'strictly speaking' the Pope etc. were bachelors, it seems to me that they would have received affirmative answers.

(2) The second type of approach to these alleged counterexamples would be more in line with the way Quine looks at formal languages.

According to that view, the word 'bachelor' can be seen as a symbol that is shared by various languages, each of which is a part of a motley collection called 'English', or 'natural language' or something of the sort. These languages are collected together by the fact that they are used by at least some people in our society at any one time.

Some languages are proper subsets of others, some languages share terms, or conceptual underpinnings, but are otherwise independent, and others might be completely independent of one another. Some terms are shared by various members of the motley crew, but vary in meaning either through variations in relative conceptual richness, or complete difference in meaning. Different languages or sublanguages can be roughly delineated by different tasks for which they were more or less consciously designed. So terms shared by distinct languages or sublanguages will vary in meaning according to the task or tasks for which the language or sublanguage exists. Being members of the overall society that makes use of this hodgepodge, we more or less pick up and use the members of the hodgepodge. Our problems with the term 'bachelor' are reflective of this situation. It may be an analytic truth of language A that an unmarried male of marriageable age is a bachelor, and an analytic truth of A that an unmarried male of marriageable age who is shacking up with a woman is also a bachelor. Yet language A may be some sort of a sublanguage of a larger language, which also has language B as a part. Language B has some task different than that of A, and according to it, only unmarried males of marriageable age who are in some sense of the word eligible are bachelors. So according to this hypothetical language B the Pope does not make the cut, and neither does our shacking up guy. We might imagine language A to be used by the legal community, or by the IRS, and language B to be used by people more or less interested in who stands a realistic chance of getting hitched.

Within each language, there are certain things that are taken for granted. In the IRS language, the universe of discourse simply consists of unmarried adult males and females, and married adult males and females, and bachelor simply means unmarried adult male, because relative to the task of determining what tax rate an adult male gets, whether he is married or not is one of the relevant characteristics he may have. Whether he is eligible is irrelevant to the purposes of the IRS. In the matchmaker's language, the universe of discourse consists of unmarried viable males and females, unmarried unviable males and females, and married males and females, and bachelor means unmarried adult male who is viable husband material, because relative to the task of determining which males are possible hitchees, not only is being unmarried a relevant property, but viability (broadly construed) is a relevant property.

In general, Quine thinks that it is troubling for those that are committed to the existence of analytic truths, that all the attempted explications of what analyticity is somehow land the believer in a closed apparently circular definition of analyticity. It is explained in terms of synonymy, and interchangeability, and these themselves ultimately depend upon the notion of analyticity themselves. Now it is not clear exactly what we should take away from these states of affairs, even if we grant that they are true. Quine seems to allow that we can create, by conventional means, some analytic, or definitional terms. Yet, he thinks that aside from this, we cannot point to analytic truths of ordinary language. Yet, he also feels that logical laws,

such as the law of excluded middle, are open to empirical falsification. He has in mind the particle wave duality of particle p hysics. So it seems that these to o are synthetic, or have some empirical element. But if we try to define analyticity in terms of any of the notions canvassed above, we will find ourselves explicating this family of terms by other terms in the family. Quine thinks this is a fatal flaw. It is fatal because it is circular.

However, there are other families of terms each member of which finds its meaning explained in terms of other members of the same set. Consider the terms 'father', 'mother' 'child'. A child is the result of genetic contributions of a male and female human (a mother and a father), who account for the child's existence. A father is a male human, who along with a female human (that would be a mother) has contributed genetically to a third human (that would be a child) accounting for that human's existence. A mother is a female human who along with a male human (father) has contributed genetically to a third human (child again) accounting for that human's existence.

Does this relation between these terms throw us into grave doubts as to the viability of familial discourse, and the very possibility of making meaningful utterances about children, fathers and mothers? Does it lead us to think their is no distinction between things familial, and things non-familial? No. Even in logic, (as Quine points out in his essay), the truth functional terms are defined in terms of each other. This fact does not lead us to abandon logical discourse, or proclaim that there can be no satisfactory account of the logical connectives. It does not lead us to claim that there is no distinction between logical truths and truths of other types.

In general, if we can countenance such families of related terms, and can establish membership in such families, then there will be analytic truths. Those truths will explicate the conceptual structure, indeed the identity of those families. Similarly, if we can countenance the family of terms that Quine presents, and their conceptual interrelations, then there will be such properties as analyticity, synonymy, and necessity.

Philosophy is replete with such families of terms. The family that Quine explores is one. Another is {knowledge, truth, justification}. Another is {good, obligatory, permissible}.

In general, Quine has this problem: If we are to take interdefinability as a fatal flaw, and as an indicator that an area of discourse is either impossible, or ultimately meaningless in some way, it seems we will have to throw out not only philosophical discourse, but much discourse that has to do with matters of fact. But we (and presumably Quine) do not want to abandon the latter sort of discourse.

Why abandon the former? Perhaps there are some practical considerations. A reason that Quine has for adopting his 'web of belief' view is that it is supposed to be a tonic against dogmatism.

Philosophical discourse might degenerate into dogmatism, and people will not keep their minds open, if they are convinced that there are truths that are immune to empirical falsification. If we were convinced web of belief theorists, this would be less likely to occur.

But the possible empirical falsification of logical laws does not lead Quine to aba ndon the practice of that discipline. Why then should he abandon, as impossible, the possibility of conceptual analysis in general, which is in effect exactly what he is doing? Perhaps his overriding concern is the spec ter of a recalcitrant dogmatism.

Concerning the worry over dogmatism, I think that way lies a two-edged sword, which can with equal justice be wielded against web of belief theory. If applied consistently, web of belief theory can and should land one into a firm acceptance of the alleged fact that all statements are in some way synthetic, and that there are no conceptual truths, and that even the laws of logic are (even if only slightly) empirical, and open to falsification. This would tend to degenerate into dogmatic relativism, and a quick dismissal of views of a more traditional nature. But, this would be to take up a position that the Quinean position is itself somehow independent of the web of belief, and privileged in that it is immune from empirical falsification. To take up this sort of position is just as dogmatic as is the position that claims that the Pope type examples do not show that there are no analytic truths, only inadequately grasped conceptual structures. So if there is no virtue in the one camp, then there is none in the other. It may be that there are analytic truths.

We should not dismiss that possibility

Bibliography

Willard Van Orman Quine, *From a Logical Point of View*, Chapter 2, pp. 20-46, Harvard University Press, 1953.

Shaun Baker

University of Michigan-Dearborn

shaunbak@umd.umich.edu

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp. 67-72
Against Compatibilism: Compulsion, Free Agency and Moral Responsibility

AGAINST COMPATIBILISM: COMPULSION, FREE AGENCY AND MORAL RESPONSIBILITY

William Ferraiolo

Free agency is incompatible with compulsion. Free agency is also necessary for moral responsibility. Acting under compulsion, therefore, entails the absence of free agency and moral responsibility. One cannot be legitimately judged morally blameworthy (or praiseworthy) for any act that is compelled by conditions that lie beyond one's ultimate control. Compatibilists will agree that free agency is necessary for legitimate ascriptions of moral responsibility, but will also argue that we often act freely and can be held morally responsible for our behavior because we often act in the absence of constraint or compulsion. Even in the face of causal determinism (i.e. the hypothesis that all events are subject to causal determination by antecedent conditions and laws of nature), compatibilists claim that we can and do, nonetheless, act freely.

In this paper, I will argue that free agency and moral responsibility are incompatible with causal determinism precisely because causal determinism entails that all human choices and actions are ultimately compelled by originating conditions beyond the agent's control. If causal determinism is true, then a complex chain (or web) of causal antecedents and laws of nature nomologically necessitate all deliberation, choice and action. If conditions beyond the agent's control ultimately determine the choices that the agent makes and the behaviors that result from them, then it follows that ascriptions of *moral* responsibility are unjustifiable (though praise and/or blame may serve pragmatic functions — e.g. conditioning future behavior). The compatibilist position is, I claim, inconsistent. Behaviors that compatibilists identify as *free* are ultimately subject to compulsion and should, by the compatibilist's own account, be identified as *unfree*.

Causal Determinism

Both compatibilists and hard determinists (i.e. determinists who deny the compatibility of causal determinism and free agency) will assent to something like the following account of the nature of causal determinism:

> **D:** Any event, E, is causally determined, just in case, given the actual antecedent conditions and laws of nature, it is not nomologically possible for E not to have occurred.

In other words, if all events that have actually occurred prior to, and/or concomitantly with E, taken jointly with the relevant laws of nature, constitute a nomologically sufficient condition for producing E, then E is causally determined.

For example, given that I pick up a book, hold it above my desk, and then release my grip on it, then, given that the book is not otherwise supported and is subject to the earth's gravitational field, the book must fall to the desk. For the book not to fall, the antecedent conditions must be different than those described, or the laws of nature must be altered. The antecedent conditions for the book falling are many. They include: the formation of the planet earth, the evolution of human beings, my birth, my picking up the book, etc. The relevant laws of nature include: the law of gravity, those laws of neurophysiology governing my bodily behaviors in their relation to brain states and environmental stimuli, etc.

If the determinist's thesis is true, then all events in the universe are subsumable under the description D (i.e. all events are causally determined). Compatibilists claim that causal determinism does not preclude the possibility of morally significant free agency. One can be causally determined to perform some act and, nonetheless, perform that act freely. But how, according to compatibilists, can such an act be legitimately called «free»?

Free Agency

Both compatibilists and hard determinists will assent to something like the following as a necessary condition for free agency:

> **F:** If a person, P, acts freely, then, on the occasion in question, P must not be subject to compulsion — and her act must be causally determined by the exertion of her will.

Surely a «free act» requires at least this: the act in question must be attributable to the agent's will and not compelled by conditions beyond the agent's control, such as neurological disorders, muscle spasms, external force, etc. If this condition is not met, then a free act does not occur. Both publicly observable behaviors and also «internal» acts of will (e.g. choices, deliberations, etc.) must be attributable to the will of some agent if that agent is to be legitimately described as acting freely on the occasion in question. An agent that is compelled to choose, deliberate, or behave in a certain fashion does not do so freely.

The Compatibilist Position

Compatibilists claim that the hard determinist (as well as the libertarian incompatibilist) tacitly assumes a faulty account of free agency that does not accord with the common usage of the relevant terms (e.g. «We are *free* to travel whereas prisoners are not.») We can make sense of the distinction between actions that are compelled or constrained by forces beyond the agent's control and those that are not, so, the compatibilist insists, it follows that *free acts* are commonly understood to be voluntary acts or acts in the absence of constraint or compulsion. We all recognize and understand the distinction between voluntary and non-voluntary behavior, hence we should all recognize and embrace the phenomenon of free agency (causal determinism notwithstanding).

So, W.T. Stace argues that an act is free so long as it is causally determined by internal psychological states (e.g. beliefs and desires) of the agent performing the act and not by conditions external to the agent. In other words, voluntary acts are free. Involuntary behaviors, on the other hand, are not free because they do not proceed from the agent's will but result instead from external compulsion or constraint. Stace suggests that we compare sets of hypothetical conversations and ask ourselves if there is not a clear distinction between the causal determinants of the behavior in each type of case described, and a concomitantly clear distinction between descriptions of free acts and descriptions of acts that are not free

(according to common usage of «free» and related terms). The following is the first of the comparisons he offers, and it is, I believe, the most felicitous illustration of the bunch:

> Jones: I once went without food for a week.
>
> Smith: Did you do that of your own free will?
>
> Jones: No. I did it because I was lost in a desert and could find no food.

To be contrasted with:

> Gandhi: I once fasted for a week.
>
> Smith: Did you do that of your own free will?
>
> Gandhi: Yes. I did it because I wanted to compel the British Government to give India its independence.

After presenting a number of similar illustrations, Stace provides an analysis that he believes to be in accord with common usage of expressions such as «of his own free will».

> We have now collected a number of cases of actions which, in the ordinary usage of the English language, would be called cases in which people have acted of their own free will. We should also say in all these cases that they *chose* to act as they did. We should also say that they could have acted otherwise, if they had chosen. For instance, Mahatma Gandhi was not compelled to fast; he chose to do so. He could have eaten if he had wanted to. [1952: pp. 250-1]

So long as Gandhi's choice to fast is determined by his desire to force the British out of India and his belief that fasting may produce the desired result, his act is free (though causally determined) and is an expression of his internal psychological states. There is nothing more required for free agency. Jones, on the other hand, is compelled to fast by an external condition, namely, the absence of food in his immediate environment. Hence, Jones does not decide to fast (he is *compelled* to do so), and his doing so is not a free act. No further analysis of the causal antecedents in each case is required to understand the distinction between free and unfree acts. Anyone not embroiled in a philosophical debate over free will would say that Gandhi's act is free and Jones' act is not. Stace claims that the issue is thus resolved.

Similarly, A.J. Ayer (1954) argues that the absence of internal compulsion (e.g. neurosis such as kleptomania) or external compulsion (e.g. a gun to one's head) entails freedom of the will that can be translated into free agency. In other words, the absence of internal or external compulsion is all that is necessary for freedom of the will and morally significant free agency. But can this condition ever actually be satisfied in a deterministic world?

For the sake of argument, I am willing to grant Stace and Ayer the claim that the absence of constraint or compulsion is sufficient for free agency. However, I will argue that it is impossible ever to be free of constraint or compulsion in a deterministic world, at least insofar as one's will is concerned. Given that the agent's will is causally determined by antecedent conditions and laws of nature, the agent is ultimately compelled to will, to choose, to deliberate, etc. as the antecedent conditions and laws of nature have determined. But antecedent conditions and laws of nature are, ultimately, originating conditions external to the agent — though the agent may be «submerged in,» and part of, the causal stream that determines all deliberation, choice and behavior.

Compulsion and Moral Responsibility

If any event, E, is causally determined, then it cannot have not occurred — given the prevailing laws of nature and the actual causal antecedents of the event in question. Suppose the event in question is an «internal» act of will (e.g. a choice resulting from causally determined deliberation). In what sense is the agent not compelled to perform this act of will (i.e. make the choice in question), if the act is causally determined by the prevailing laws of nature, the agent's environment, the complex causal relationships between the agent's brain states, psycho-physical history, heredity, current environmental stimuli, and the natural laws that govern such complex relations? In Stace's scenario, the external world compels Gandhi to choose as he does, just as the absence of food in the desert compels Jones to fast. Causal determination of externally observable behavior *by* an act of will, does not preclude prior causal determination *of the act of will* by conditions that originate in the external world. Gandhi's beliefs and desires are determined by his heredity and environment — his initial cognitive endowment and all modifications thereof resulting from environmental impingements. These compel him to deliberate and choose accordingly. If Gandhi's choice is compelled, then it seems that the resulting behavior is compelled as well. The mere fact that one's body behaves in accordance with one's will does not imply that one's will is free.

The question is this: Can one, given the actual antecedent conditions and laws of nature, choose otherwise than one actually does? As C. A. Campbell (a libertarian incompatibilist) puts it:

> We do not consider the acts of a robot to be morally responsible acts; nor do we consider the acts of a man to be so save insofar as they are distinguishable from those of a robot by reflecting an inner life of choice. Similarly, from the other side, if we are satisfied that a person has definitely elected to follow a course which he believes to be wrong, but has been prevented by external circumstances from translating his inner choice into an overt act, we still regard him as morally blameworthy. Moral freedom, then, pertains to *inner* acts. [1957: p. 160]

The implications for moral responsibility are fairly straightforward. To hold any agent morally responsible for his «inner acts», they must not be the result of compulsion. Moral claims are ultimately normative. If any agent is morally obligated to choose in a certain way, that agent must have the capacity to so choose. One cannot be obligated to do that which is ultimately beyond one's power. If it is not within one's power to choose otherwise than as is determined by the relevant antecedent conditions and laws of nature, then one cannot be obligated to choose otherwise. Claiming that one *should* choose differently than one actually does is tantamount to claiming that one *should* have had a different history than one actually does. Campbell, again, nicely articulates the point:

> The proposition which we must be able to affirm if moral praise or blame of X is to be justified is the categorical proposition that X could have acted otherwise because — not if — he could have chosen otherwise; or, it is essentially the inner side of the act that matters, the proposition simply that X could have chosen otherwise. [1957: p. 164]

If determinism is correct, the claim that I could have behaved otherwise if I had chosen otherwise, is vacuously true because I never *could* have chosen otherwise given the actual causal antecedents of my choice. The claim that I *should* have chosen that which I *cannot* have chosen (not to mention the moral blame attending that claim) is indefensible. It is no different than suggesting that I *should* have been born with blue eyes and blaming me for my brown eyes.

Compatibilist Reply

Ayer complains that this type of incompatibilist argument conflates the concepts of causation and compulsion:

> That all causes equally necessitate is indeed a tautology, if the word 'necessitate' is taken merely as equivalent to 'cause': but if, as the objection requires, it is taken as equivalent to 'constrain' or 'compel', then I do not think that this proposition is true. For all that is needed for one event to be the cause of another is that, in the given circumstances, the event which is said to be the effect would not have occurred if it had not been for the occurrence of the event which is said to be the cause, or vice versa, according as causes are interpreted as necessary, or sufficient, conditions: and this fact is usually deducible from some causal law which states that whenever an event of the one kind occurs then, given suitable conditions, an event of the other kind will occur in a certain temporal or spatio-temporal relationship to it. In short, there is an invariable concomitance between the two classes of events; but there is no compulsion, in any but a metaphorical sense. [1954: pp. 281-2]

But it is not metaphorical compulsion that results from the conjunction of one's brain states and the laws that govern the translation of such states into acts (including «inner acts» such as deliberation and choice). In such matters, the distinction between «necessitate» and «compel» is a distinction without a difference. If causal determinism is true, then one is quite literally compelled to act as is nomologically necessitated by one's overall neurophysiological state, and one is quite literally constrained by the relevant laws of biology, chemistry, physics, etc. Furthermore, the fact that the agent's brain states and other physiological states are «internal» to her is irrelevant to ascriptions of freedom and moral responsibility. A kleptomaniac or a person with a gun to her head behaves as her brain states determine that she must, but Ayer lists these as agents who do *not* behave freely (because they are compelled by forces beyond their control). The statistical abnormality of the causal determinants of psychological states in these cases is irrelevant. Compulsion via quotidian psychological processes is no less necessitating than compulsion via psychosis or threat. Sanity does not liberate the agent from the laws of nature, but merely consigns her case to the «usual» mechanisms of necessitation — and compulsion.

Again, Ayer claims that the incompatibilist position assumes an unsatisfiable account of free agency that is not in accord with common usage, and that misses the real intention behind common ascriptions of freedom and moral responsibility:

> …to say that I could have acted otherwise is to say, first, that I should have acted otherwise if I had so chosen; secondly, that my action was voluntary in the sense in which the actions, say, of the kleptomaniac are not; and thirdly, that nobody compelled me to choose as I did: and these three conditions may very well be fulfilled. When they are fulfilled, I may be said to have acted freely. [1954: p. 282]

What Ayer ignores is the fact that in a deterministic world, the third condition (which clearly should state that *nothing*, rather than «*nobody* compelled me…») cannot be satisfied. He also ignores the aforementioned vacuity of the first condition (one *cannot* have chosen otherwise). Perhaps not some*one*, but some*thing*, namely, the antecedent conditions and laws of nature (taken jointly), always compel one's choice in a deterministic world. The laws of nature and the antecedent conditions determine every one of the agent's internal psychological states, and these in turn determine each of the agent's behaviors. Ultimately, therefore, all causal power is rooted in conditions that originate externally to the agent. If forces beyond our control compel us to believe, desire, and choose as we do, then our beliefs, desires and choices are not free. Hence, if our internal states are causally determined, then there is no meaningful sense in which we can be ascribed ultimate responsibility for the behaviors that

result from them. The fact that causal mechanisms proceed, in part, through the medium of our beliefs and desires, in no way diminishes our compulsion or endows us with freedom.

Conclusion

Even if we assume the compatibilist account of free agency, we must conclude that free will is an illusion in a deterministic world. If determinism is true, human behaviors, including *acts of will*, are compelled by antecedent conditions and laws of nature, and none of them could have been avoided — unless the world had turned out differently. But the world is, of course, as it is and not as it might have been. Similarly, the agent's causal history and environment are fixed antecedents and concomitants of each of the agent's internal states. These conditions compel the agent's choices. Compelled choices compel behaviors. If determinism is true, we are not free and we cannot be legitimately held morally responsible for our actions.

References

Ayer, A.J. (1954). *Philosophical Essays*. New York: St. Martin's Press.

Campbell, C.A. (1957). *On Selfhood and Godhood*. New York: MacMillan Company.

Stace, W.T. (1952). *Religion and the Modern Mind*. New York: Lippincott.

William Ferraiolo

San Joaquin Delta College. Social Sciences Division

Stockton, CA 95207

bferraiolo@deltacollege.edu

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp. 73-75
Mad, Martian, but not Mad Martian Pain

Mad, Martian, but not Mad Martian Pain[1]

Peter Alward

Lewis (1980) has argued that neither the identity theory nor functionalism can accommodate the possibility of both mad pain and Martian pain. Functionalism cannot accommodate the possibility of mad pain — pain whose causes and effects diverge from those of the pain causal role. This is because what it is to be in pain according to functionalism is simply to be in a state that occupies the pain role. And the identity theory cannot accommodate the possibility of Martian pain — pain whose physical realization is foot-cavity inflation rather than C-fibre activation (or whatever physiological state occupies the pain-role in normal humans). After all, what it is to be in pain according to the identity theory is to be in whatever state that occupies the pain role for us.

Lewis attempts to solve this difficulty by combining functionalism and the identity theory in the following way: he gives a functionalist account of pain *for a population*, and gives an identity theoretical account of pain *for individual members of a population*. According to Lewis, a state S is pain for a population P if and only if, with few exceptions, whenever a member of P is in S, her being in S has the sorts of causes and effects given by the pain role. (Lewis, 1980, p. 113) As a result, C-fibre activation (or what have you) is pain for the human population, and foot-cavity inflation is pain for the Martian population. And an individual, X, is in pain if and only if X is in a state which is pain *for the appropriate population*. (Lewis, 1980, p. 113) Thus, any individual for whom the appropriate population is the human population is in pain just in case she is in a state of C-fibre activation. And any individual for whom the appropriate population is the Martian population is in pain just in case she is in a state of foot-cavity inflation.

The problem that arises for Lewis' view concerns his account of under what conditions a population P is appropriate for a given individual X. He provides us with the following four criteria:

(1) P is the human population.

(2) X is a member of P.

(3) P is a population in which X is unexceptional.

(4) P is a natural kind.

Lewis does not give much explicit guidance as to how to apply these criteria, so it is best to look at how he does so in particular cases. Consider the following:

---

[1] Thanks are due to the student in my philosophy of mind class at the University of Lethbridge in the fall of 2001 whose struggles with Lewis helped me «see the light.»

(a) If X is our Martian, we are inclined to say that he is in pain when the cavities in his feet are inflated; and so says the theory, provided that criterion (1) is outweighed by the other three, so that the appropriate population is taken to be the species of Martians to which X belongs. (Lewis, 1980, p. 113)

(b) If X is our madman, we are inclined to say that he is in pain when he is in the state that occupies the role of pain for the rest of us; and so says the theory, provided that criterion (3) is outweighed by the other three, so that the appropriate population is taken to be mankind. (Lewis, 1980, p. 113)

The suggestion seems to be that each criterion is of relatively equal weight and, so, the population that satisfies the greatest number of criteria for a given individual is the appropriate population for that individual.

But note that given this characterization of the procedure for applying the criteria, Lewis' application of them is in both instances slightly askew. Consider case (a). The Martian population does satisfy criteria (2), (3), and (4) for a Martian for whom foot-cavity inflation occupies the pain-role. After all, it is a population which is a natural kind and of which our ordinary Martian is an unexceptional member. But the Human population does not merely satisfy criterion (1), but criterion (4) as well. And so while Lewis's theory does imply that the appropriate population for our ordinary Martian is the Martian population, it has to be because (2), (3), and (4) outweigh (1) and (4) together, and not simply (1) by itself as Lewis suggests. Similarly, in case (b), Lewis's theory implies that the appropriate population for the mad human is the human population rather than the mad human population. But this has to be because (1), (2) and (4) outweigh (2) and (3) together, and not (3) by itself as Lewis claims.

The real trouble arises, however, when we consider the case of the mad Martian. Lewis gives the following analysis of this case:

(c) If X is a mad Martian, I would be inclined to say that he is in pain when the cavities in his feet are inflated; and so says our theory provided that criteria (2) and (4) together outweigh either (1) or (3) by itself. (Lewis, 1980, p. 114)

Now I agree that the Martian population does satisfy criteria (2) and (4) for a Martian for whom foot-cavity inflation fails to occupy the pain-role. But note: the human population does not merely satisfy criterion (1) for our extraordinary Martian, but criteria (1) and (4). And the mad Martian population does not merely satisfy criterion (3) in this case, but criteria (2) and (3). The upshot is that Lewis view does not imply that the appropriate population for our extraordinary Martian is the Martian population as opposed to the Human or mad Martian populations. As it stands, each of the three populations has an equal claim to be the appropriate one in this case. And so it seems that Lewis's theory fails to accommodate the possibility of mad Martian pain.

The source of this difficulty is the fact that criteria (1) and (3) have been inappropriately given the same weight as (2) and (4). The reason less weight ought to be given to criterion (1) is that, given that the human population is a natural kind, it automatically satisfies two of the criteria for being the appropriate population for any arbitrary individual. And the reason less weight ought to be given to criterion (3) is that for any physical realization of any given causal role, there will be a population of individuals who are similar with respect to the realization of the role, but many of whose (at least possible) members will be otherwise

physiologically very different.[2] And some such motley population will always satisfy two of the criteria for being appropriate for any given individual.

In light of this, one possible emendation to Lewis' view would be simply to dispose of criteria (1) and (3) altogether. This approach, however, ought to be rejected. Suppose, for example, humans were an exceptional species with respect to the physical realization of pain within a larger a larger genus, itself a natural kind. The wholesale dismissal of (1) and (3) would have the effect of giving the human population and the more inclusive genus equal claim to being the appropriate population for any given ordinary human.

A better suggestion, at first glance, would be to retain (1) and (3) but give them lower weight than (2) and (4). This would yield the desired results regarding both sane human and mad Martian pain. But optimism here may well be premature. This is because there seem to be good reasons for demanding both that criterion (1) outweigh criterion (3) and that criterion (3) outweigh criterion (1), surely an untenable result. First, unless (1) outweighs (3), there will be no grounds for taking the appropriate population for Lewis' madman to be the human population as opposed to a more inclusive genus in which C-fibre activation is the occupant of the same causal role it occupies for the madman. But second, unless (3) outweighs (1), there will be no grounds for taking the appropriate population for an artificially intelligent robot to be the population of similarly designed robots (no natural kind) as opposed to the human population. Lewis' view, it seems, may ultimately be unsalvageable.

References

David Lewis (1980), «Mad Pain and Martian Pain», in *Readings in the Philosophy of Psychology, Vol. I*, N. Block, ed., Harvard University Press, 1980, pp. 216-222. Page references to *Problems in Mind*, J. Crumley, ed., Mayfield Publishing Company, 2000, pp. 110-117.

Peter Alward

University of Lethbridge. Canada

peter.alward@uleth.ca

[2] Note: I am assuming that what criterion (3) requires is that an individual have the same physical realization of (relevant?) causal roles as members of a population.

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp. 76-86
The Veil of Perception and Contextual Relativism

THE VEIL OF PERCEPTION AND CONTEXTUAL RELATIVISM

Dimitris Platchias

The question that the present paper addresses is what are the proper objects of perception. By this one might mean something like what would be the immediate objects of perception, the objects of direct acquaintance or that which can be perceived.[1] Answers to this can be divided into two broad categories. The first has it that we are immediately acquainted with mind-independent existing objects such as tables, chairs, cars etc. or to put it differently we directly perceive objects or acquire information about objects located in the external world. The second claims for a need of an intermediary where our acquiring of information about external things depends upon immediate acquaintance with some mind-dependent entities (private mental objects), that is percepts; and further, only derivatively- through these objects- we acquire information about mind-independent objects located in the external world.

Both claims are a species of Realism in that they both claim an external world containing mind-independent physical objects. The first is broadly called 'direct Realism'[2] and the second 'indirect or representative Realism.' It is clear that what distinguishes the two kinds of realism is the veil of perception existent in the latter. It is that veil that led Berkeley to repudiate Locke's scientific Realism based on the impossibility of justifying a correlation between the percept and the mind-independent object that causes that percept — there can be

---

[1] I'll concentrate on *that which can be perceived* i.e. that which literally can be *seen*, *heard*, *touched* etc. Objects and more particularly properties of objects and not events or situations. And moreover 'seeing' and not 'seeing that'. To this one might reply with reference to D.M.Armstrong's twofold claim (for a reference see my bibliography) that mediate perception involves inference whereas immediate perception involves no element of inference and that all perception is the acquiring of beliefs. Whereas there is a whole lot more to be said as regards the distinction I will only point out two things. First, not all perception involves belief as the cases of walking (e.g. side-stepping of trees) and driving show, where the subject may be unaware of what she has perceived. Answers in terms of the presence of unconscious inference or of inclinations to believe do not seem persuasive enough. And even if one accepts that such cases can be somehow explained by the presence of unconscious inferences or beliefs there are cases where there is an illusory appearance *p* even when the subject has a justified belief that not-*p* (e.g. Muller-Lyer case). Second, The misapprehension of the distinction between 'seeing' and 'seeing that' (the belief about perception or the interpretation of *that which is perceived*) or between cognitive (conceptual) content and content that involves necessarily belief is the reason for postulating a necessary involvement of inference in the case of mediate perception whereas in the case of immediate not.

[2] It is claimed however to be subdivided to a scientific form of direct realism and to a very close view labelled naïve realism or common-sense realist view.

no resemblance between them, an idea can be like nothing but another idea[3] (Hume has also claimed for the inescapability of a veil of perception pointing out the impossibility of making a step beyond appearances). Thus, in *The First Dialogue*, Philonous claims:

> How can that which is sensible be *like* that which is insensible?[4]

Berkeley is seen as the predecessor of Phenomenalism where one of its propounders, J.S.Mill, has claimed of material objects as being permanent possibilities of sensory experiences. The above views summarized can be restated thus (where R1 stands for a causal etc. relation between *x* and *y*):

i. $(\exists x)\ (\exists y)$ [*x* is mind-independent & *y* is a property of *x* & A sees *y*],

ii. $(\exists x)\ (\exists y)$ [*x* is mind-independent & *x* R1 *y* & *y* is mental & A sees *y*],

iii. $(\exists x)\ (\exists y)$ [*x* is mind-dependent & *y* is a property of *x* & A sees *y*],

(with respect to direct realism, indirect realism, and phenomenalism), and hence, that which can be seen is *y* (a perceived entity), that is, properties such as colour, shape, extension and texture. As it is apparent, there has to be drawn a distinction between sense-data and percepts. I follow Don Locke's idea[5] that sense-data is a theory-neutral term. Thus, the formulation of the question should be whether speaking about sense-data is reducible to speak about percepts and if so in what way. Otherwise the question of whether phenomenal properties can exist unperceived would have no value whatsoever. In that case (if sense-data are percepts) they cannot.

If these are the only options, then we either perceive mind-independent objects in a direct way and thus we acquire information about these objects in an immediate manner or we are caught in a veil of perception and we are only acquainted with it, thus acquiring information for what there is beyond that veil (mind-independent objects) only derivatively, which is always forcing one to formulate a plausible account of the nature of the correlation (causal etc.) between appearances and the external world. Moreover, one must also provide a rationale for explaining the presence of a mind-independent world beyond that veil in that otherwise it will lead unavoidably to scepticism about the external world (to a refutation of the presence of a mind-independent reality beyond that veil). But why should we claim the presence of such private mental objects leading to a presence of a veil of perception in the first place?[6] What would a postulation of such entities serve?

A common-sense realist view of perception holds that secondary qualities are no less real than primary qualities.[7] Secondary or sensory qualities of objects are mind-independent. That is to say, colours, temperature, sounds, texture etc. are *properties that physical objects possess independently of the existence of a perceiving subject.* Moreover, the claim of course entails

---

3 In Berkeley's *The Principles of Human Knowledge*, section 9.

4 In Berkeley's *Three Dialogues Between Hylas and Philonous*, p. 169.

5 For a discussion see Locke's *Perception and Our Knowledge of the External World*.

6 Derivatively, what would be the properties of such entities?

7 I say more on the distinction later on.

that the primary qualities of an object such as, shape, size, solidity, position, are of this kind. Thus, Aristotle claimed that the subject directly perceives properties possessed by objects and moreover,

> With regard to all perception, we must take it that the sense is that which can receive perceptible forms without their matter, as wax receives the imprint of the ring without the iron or gold, and it takes the imprint which is of gold or or bronze.[8]

It is clear however that such a contention faces severe problems. Objects in different contexts appear different. Their phenomenal properties change. It is implausible for one to maintain that he is directly acquainted with the actual properties that objects really possess (properties that appear different in different contexts have equal right in being considered as real since — according to Aristotle — we *receive perceptible forms* as wax receives the imprint of the ring without the iron or gold). Why should we therefore favour any of the received (perceived) properties?

Thus, besides the problems that the direct realist account faces with cases such as when one's perceptual error is due to a *subjective standing* of his perceptual state, and also, in cases of hallucination, where there might be no presence of an external object whatsoever,[9] there are cases where there is an *objective standing* of phenomenal properties, that is, in these cases as in the cases of veridical perception there is an interpersonal convergence in judgment. The case of the bent stick in the water, the white expanse of the wall that under certain conditions of illumination looks red, or the case of the rectangular surface of a table which, viewed from a certain angle, its two opposite lines appear to converge, are cases of such an interpersonal convergence in judgment.

Now, the stick cannot be both straight and bent, the expanse of the wall both white and red and the surface of the table both rectangular and not-rectangular. There are however, no distinguishing marks between the two perceptual states in each case. The objective phenomenal standing of both cases (veridical perception and perceptual error) enables one to suggest that with what one is directly acquainted is not a property of the object i.e. of an aspect of the world, but rather a mental item. Hence, the postulation of such entities, private mental objects, serves to explain the case of the perceptual error. With what one is directly aware (or acquainted) is appearances, and not the real properties of objects. This objection, especially when it comes to perceptual errors with an objective standing, looks devastating to a naïve realist analysis of perception.[10]

The direct realist next move is to appeal to modern science. The claim is that the 'real' properties of objects are not appearance-determined or perceiver-dependent. The identification of the 'real' properties of objects has to be done not by appealing to the perceptual state of the perceiving subject but rather by an appeal to the real character of the properties of objects as it is identified by science. Modern science tells us that the constitution of colour is light

---

[8] Aristotle, *De Anima*, p. 42.

[9] The direct realist here appeals to the distinction between sensation and cognition but that is an issue that needs separate discussion. Later on, I separate illusion into four categories.

[10] See third case of illusion at the end of paper.

of a certain wavelength, of heat is molecular motion etc. A closer look to what modern physics has, as regards light and matter, is of importance.

Light is referred to as electromagnetic radiation because the nature of light is based on tiny electromagnetic fields, called photons. These photons of light can have many different energy levels or wavelengths, which are expressed in nanometers (nm). Light has properties both of a particle and a wave. In the case of the visible wavelengths, every wavelength is represented by a different colour. Visible light takes up only a tiny fraction of the electromagnetic spectrum, that between ultraviolet and infrared light (heat). Now we see things as a colour because objects typically absorb all light except one particular wavelength of light or colour. That wavelength (colour) is reflected off the object and absorbed by our eyes, hence we see the object as being the reflected colour.[11] As regards matter, all matter is made up of various combinations of three elementary particles: protons, electrons and neutrons. Electrons have wave properties similar to those of light. All matter is in a state of perpetual activity. There is no dividing of matter and force (energy) into two distinct terms, as they are one (the matter/energy distinction of a given phenomenon is relative to the inertial frame of the observer).

Now if one points out, when referring to an expanse of the wall with which one is directly acquainted and which looks red, that 'this expanse is not really red', one would not be taken as uttering an inconsistency. There may be bizarre conditions of illumination of which the agent is aware. Similarly, a distant object looks small and the shape of an object varies according to the conditions of observation. Moreover, objects in the everyday life are often obscure, or only partly visible. Thus, different perspectives commit one to different verdicts concerning an object's shape, extension and texture. Indeed, if we are to employ a microscope the texture of the object looks different. Moreover, viewed from different angles the shape of an object looks different. Therefore one's claim that 'the surface of this table is actually rectangular' when its two opposite lines appear to converge makes perfect sense.[12]

Phenomenal properties of an object therefore do change with respect to different contexts. In addition, the potential phenomenal character of these qualities such as colour, shape and texture is intrinsically the same. This is to mean, as in the case of colours that microphysical

---

[11] Light is attracted by gravity, it is deflected. The deflection of light is immediately related to the decrease of its speed. Mass and energy are not two separate things but rather they could be traded one for the other (the conversion of mass to energy could account for the enormous energy output of the stars). Moreover, nowadays, there are several physical processes that can accomplish that, that is, to realize this potentiality; and further, the light can provide detailed information about distant stars and planets, e.g. distance, speed, rotation, chemical makeup etc. as also sounds (being wave motions), can give us information not only on the sources from which they originate, but also on the bodies through which they pass, and against which they are reflected or deflected.

[12] What one perceives in both cases, namely, the 'convergence of the two opposite lines of the table' and the 'red expanse of the wall' do not differ in nature. In both cases things appear to be different than they really are. The subject perceives them both (as in the case of the stick looking bent in the water) as being the case. Special conditions (specified differently in each case) lead a putatively unaware subject, or one with no past experience, to believe that the two opposite lines of the table do converge, the expanse of the wall is red and the stick is indeed bent (a subject unaware of '*p* is not true because of *q*' or of 'perceptual experience *p* misrepresents reality for the reason that conditions *q* obtain' would take these perceptual experiences as being the case). That is, all three perceptual states do not differ in nature.

structures of the objects have a permanent disposition of reflecting light of certain wavelengths respecting to different contexts, and they therefore appear differently coloured, in the case of shape, size and texture they do also have a permanent disposition to appear in a certain manner as regards to different contexts. As nothing can be round and square at the same time, nothing can be blue and red at the same time. The colour of an object, (the colour it appears to be), is not due to chance. The colours (visible wavelengths) that it absorbs or reflects are determined by the atomic and molecular structure of its surface.

The claim that we can measure (and thus confirm) the, say square, shape of an object, whereas in the case of light of a certain wave length we cannot verify its redness is evidently false. It is that which differentiates colour judgments from moral judgments. A utterance of the kind '*x* is red' by normal humans in the same context requires an interpersonal convergence in judgement (with no deliberation whatsoever). It is self-explanatory. He, who cannot discern that '*x* is red', in a certain context, *ipso facto* has a certain kind of deficiency. Moreover, to assert that a colour is *actually* light of a certain wavelength (respecting each colour) is to imply at the same time that there is a scale upon which colours can be placed according to different wavelengths and in case of the application of a form of measurement of light (e.g. spectrophotometers are used for colour measurements), or of the arrangement of the microphysical structure of the surface of the object, science can identify the 'real' colour of an object.[13]

'Colours as seen' exist only when perceived. Light of a certain wavelength appears to one red. An animal may perceive it differently. The snake 'sees' the heat of the body of living creatures. That is, as in the case of the colour temperature in humans, it perceives heat colourfully. A close examination to a green dress will show that the dress is not really green (the range of the colour spectrum is expanded). In addition, the texture of an object if we employ a microscope looks different. A modern electron microscope can just about see individual atoms. Is this to mean that objects are really such as we see them without the microscope solid and rigid (in that we do not perceive the tiny particles that constitute them) whereas colours are not as we see them and actually objects are colourless? Why should we only withdraw the first claim and maintain that there are objects as we see them (though soundless, colourless etc.) existing independently of a perceiving subject? If the world is constituted by tiny particles and what we perceive is objects, why should we not claim that there is indeed a veil of perception where, whereas there are such things as light of certain wavelength, tiny particles, electric charges, sound waves, molecular motion, we instead perceive colours, objects, sounds, heat and so forth, and not reality as such. If there were no perceiving subjects and thus there presumably be no red and only light waves why shouldn't also be no tables and cars but instead tiny particles and electric charges?

Strawson makes an attempt to reconcile the naïve realist view(stated above) with a form of scientific realism. He claims:

---

[13] Likewise, in the case of temperature as an alleged secondary quality, one may feel as having fever or high temperature, however the employment of a thermometer might show that *actually* his temperature is normal. Moreover, a claim that an object has a certain shape or size (as in the case of colour) can only be contingently true whereas a claim that there exists a non-dispositional molecular structure on the grounds of which, such dispositional properties of the objects are manifested, appears to be necessarily true.

> This method of reconciling scientific and common-sense realism requires us to recognize a certain relativity in our conception of the real properties of physical objects. Relative to the human perceptual standpoint the grosser physical objects are visuo-tactile continuants (and within that standpoint the phenomenal properties they possess are relative to particular perceptual viewpoints, taken as standard). Relative to the scientific standpoint, they have no properties but those which figure in the physical theories of science.[14]

Now, it is not hard to see why Strawson's attempt actually amounts to a potential reconciliation between common-sense realism and indirect realism. This can be illustrated more clearly if we state first what scientific realism holds.

For a scientific realist 'colours as seen' 'sounds as heard' etc.(appealing to the Lockean picture of secondary qualities) cannot exist in the absence of the perceiving subject. The claim (in a reductive form) identifies colours with distinct light waves in terms of a disposition of the microphysical structure of a surface to reflect light of a certain wavelength in certain conditions. That is to say that redness (similar considerations hold, as shown above, for sounds, heat, taste and also size, shape and texture) is a permanent possible cause of the experience of seeing red. I'm therefore able to perceive the red colour of the chair (appearing red to me) due to the manifestation of a certain disposition that the surface of the object possess, that is, according to the context in which I perceive the object in question, its surface reflects light of a certain wavelength.[15]

Now, what we are left with? What is that which ultimately remains constant in this object and which possess the dispositions to manifest all these properties? It is plain, that what is left of the object as we know it, is hard even to conceive. If therefore scientific realism claims of the object as being a set of dispositions to be manifested under certain conditions based on an unconceivable reality where there holds a correlation between 'seeing red' and the disposition of such a reality to cause that sensation, in what respect then would that account differ from an indirect realist account? It appears therefore that a claim of the kind '*y*s are to be identified as a set of dispositions manifested according to certain contexts' cannot escape of falling into (ii) as stated above. Such an attempt of reconciling direct realism with the indirect variant (by simply invoking different perceptual viewpoints) simply won't do. It is incoherent. Therefore clause (i) is rejected.[16]

I turn now to (ii). The Lockean variant of indirect realism (representationalism) holds that:

---

[14] And moreover: «'This smooth, green, leather table-top', we say, 'is, considered scientifically, nothing but a congeries of electric charges widely separated and in rapid motion'». In his *Perception and its Objects*, pp. 110-111.

[15] The dispositional theory of colour holds that for an «object to instantiate a colour property is for it to have a disposition to cause experiences as of an object having that property in normal perceivers in normal conditions». In C. McGinn's *Knowledge and Reality — Selected Essays*, p. 298. McGinn however rightfully suggests that phenomenology of colour perception contradicts such an identification, namely colour properties do not look much like dispositions to produce colour experiences and thus an error theory of colour perception comes to seem inescapable. I will come back to this point later on.

[16] It might be objected that Strawson's account can be put in terms of the Fregean sense-reference distinction. That is to say, in the case there is an absolute physical property of an object P we perceive it in a qualitative manner Q (e.g. 'colours as seen'). Q is not itself an object of perception but rather it is a mode of presentation of P. Thus, in terms of relativity, what Q presents depends on our occupying a certain point of view. The account however fails with respect to the third case of illusion (stated above) which is stated more clearly at the end of paper.

> Since the things the mind contemplates are none of them, besides itself, present to the understanding, it is necessary that something else, as a sigh or representation of the thing it considers should be present to it: and these are ideas…The ideas of primary qualities of bodies are resemblances of them, and their patterns do really exist in the bodies themselves, but the ideas produced in us by the secondary qualities have no resemblance of them at all. There is nothing like our ideas (of secondary qualities) existing in the bodies themselves.[17]

In short, representationalism holds that that which is immediately perceived is caused by that which is mediately perceived, and further, the former resembles the later (the Lockean primary qualities). However, the notion of such a resemblance is hardly graspable. Appearances, that is phenomenal properties, can be y_1 at t_1, y_2 at t_2, y_3 at t_3 and so forth (according to different contexts), where $y_1 \neq y_2 \neq y_3$. Is this to mean that the object has at t_1, t_2, t_3 different primary qualities? For the representationalist the resemblance holds only in the case of the veridical perception. As we saw earlier the argument from illusion and modern physics were the primary reasons for the direct realist analysis of perception to be rejected. That was also the reason for the postulation of mental objects.

But how is it that these objects resemble reality (external objects) in the case of veridical perception? The external object appears to be three-dimensional and the mental object is suggested to resemble spatial properties of the external object such as shape, size, distance and direction. Private mental objects cannot be three-dimensional. Can it be a case similar to the relation between a picture and its object? The picture shows how the object looks like, it does not resemble the object. What the picture resembles is the object's phenomenal properties not the primary ones. Therefore (ii) with the form of representationalism is rejected.

The veil of perception therefore consists of the appearances (the phenomenal properties) of objects, that is, in a reductive manner, of sense data (if we are to exclude the first two cases of illusion stated above). Sense-data alter with respect to a different context. Phenomenal properties of the objects must be these properties that change with respect to a different context and these properties cannot be the *intrinsic* properties of the object. That is to say, properties without which an object cannot exist, or properties without which, an object cannot be the same object. Thus an object, the same object, can have different colours (as we perceive them), different texture different shape etc, and it still remains the very same object.

In different contexts an object appears different. Sense-data differ, it is though evident that the object in question remains the same object. There has to be thus something that remains constant in this object or something which is intrinsic to it and which could identify this object in every physically possible context. We might claim then of secondary qualities as the properties that alter in virtue of a different context, and of primary qualities as the intrinsic properties of the object.[18] It is not the case that particles of light (photons) are seen in the everyday life as it is also not the case that electrons, protons etc. are seen either. In addition, to think of the colour of an individual atom is meaningless. This is roughly an order of

---

[17] John Locke, *An Essay Concerning Human Understanding*, pp. 69, 370.

[18] This applies to Berkeley's argument of the indiscernible of primary and secondary qualities (as having the same status) in his *Principles of Human Knowledge*, wherein not managing to show that these qualities cannot be mind-independent, arguing thus, enables one to ask how, provided that an object has certain qualities (such as a distinction of primary and secondary qualities is inapplicable), are we to identify that object in a different context where its qualities (or some of them) have changed? Further, by what means can one be aware of such an alteration? I meet that point later on.

magnitude too small for it to reflect light on its own — the light wave is too big to be reflected and show us colour. Furthermore, modern physics tells us that matter is capable of infinite subdivision. This is why Russell was so right when he claimed (with reference to an object's texture) that if we are not to trust our naked eye in what we see why should we trust the microscope.[19]

If however, there is nothing to remain constant or in the case that there is no intrinsic property that an object has then what could give justice to the claim that the object in question which appears different in different contexts, is that very object, and 'appears' is not to be considered as 'is'? But then what is still to perceive *x* if *y* varies? Is there any point at all to argue that *x* remains the same when *y* varies? That is to say, if, for instance, we accept that the real colour of an object is that under inspection in 'normal circumstances', e.g. green, and not the one that appears under a closer look where the range of the colour spectrum is expanded, then to hold that 'of an *x*, *y* may vary whereas *x* remains *that x*' would be absurd. Variability comes in two levels. For an object *x*, if *z* is its dispositional properties[20] and *y* that which is perceived (shapes, colours, heat etc) then *z*, *y* may vary. Veridical perception occurs when statements about *y* can be reducible to statements about *z*. I say more on this in the remainder of the paper.

Consequently, it appears that we are indeed inescapably caught in a veil of perception. The fact that one perceives only *y*s seems to show that one is actually caught inescapably in that veil. It certainly looks hard to consent to but appears that one has to. Again though, there appears the problem of identification. That is to say, what is for an *x* to be *that x* in every possible context. What is that which remains constant or to put it differently what is that makes statements of *x* to be of *that x* when *z* and *y* may vary without render these statements senseless? It is plain that there is no room for (iii) as a plausible candidate. It appears that the grounds for the identification of an object, in different contexts, can be given only by a realist analysis of a material object.

Perceptual experiences that may be caused in us differ from a permanent disposition to behave in a certain manner. Thus, manifestations of the dispositional properties of objects according to different contexts appear to function as identity determinants. That is, object B and object C manifesting dispositional properties γ and δ respectively, according to a certain context, where $\gamma=\delta$ is also B=C. That might lead us to provide an account of what is for an object to be identical with another. But our primary concern here rather is, what is for *x* to be *that x* in different contexts, or to put it differently, how a particular object is to be identified throughout different contexts. Before I proceed, it would be of importance to stress that according to what I've said so far, *z* is an objective standing property (not to be confused

---

[19] Notably, there is also the view that matter after a certain point of subdivision 'falls' inevitably into energy (quarks may be destroyed and hence no further subdivision). But even the very notion of a' quark' is characterized by its relations to other physical entities. The same holds for other physical properties (e.g. mass). That is, they are characterized by certain dispositional roles and relations to other physical entities. But physics says nothing about the intrinsic properties of such physical entities that could ground their dispositions. Hence, rightfully follows the question: Can a physical entity have a certain dispositional role (be characterized by that role) without at the same time having any intrinsic properties to ground its dispositions?

[20] *Z*s are to be taken into account as manifested properties (sensible properties e.g. light of a certain wavelength) and not as permanent causes of appearances.

with phenomenal property). It can exist independently of a perceiving mind. The fact that it is causally related to a non-dispositional ground is not to mean that (as it is shown) it cannot consist of a non-dispositional molecular structure. To put it differently, the fact that α_1 is causally related to α_2 is not to mean that they cannot have the same nature (an event is causally related to another event and matter is causally related to matter).

Now to our primary concern, namely, what is for *x* to be *that x*, in different contexts. I could perhaps suggest the following: An *x* (object) is *that x*, when with respect to a certain context x manifests z dispositional properties, perceived by *a* subject as *y* (appearances), on a basis of permanent non-dispositional properties *n*, where *n* is identifiable with *z* and vice versa. It could only be identified thus. To say that the world only contains *n*s is certainly inconceivable. In addition, to say that *z*s are sense-dependent is to be confused with *y*s. *Z*s are out there, ready to be perceived.[21] Furthermore, it will be of interest to attempt to provide an account of what is for A to perceive *that x* with respect to different contexts. Thus, for objects x_1, x_2, *x*v that cause phenomenal properties of an objective standing (veridical and illusory perception) y_1, y_2, y_v to A, with respect to a certain context,[22] if $y_1 \neq y_2 \neq y_v$ then it is also $x_1=x_2=x_v$. This cannot be other than a self-identity relation. It is plain that what is for A to perceive *that x* depends on what is for *x* to be *that x*.

For a more clear grasp some further remarks might be of some help. The key point to my considerations is *z*s. I have considered *z*s as objective standing properties (i.e. aspects of the world), and not as phenomenal. In addition, z*s* are manifested properties of the objects (sensible properties) and not permanent causes of appearances. Furthermore (as argued), they *can* have the same nature with the non-dispositional properties of an object (the intrinsic properties of it) *n*s, as stated above. As regards perception, whereas in Locke's representationalism is claimed that an object *x* stands in a relation to *a* subject (more precisely, in my account, *n*s R2 A), I suggest that the relation is between the subject and the mind-independent *z*s which are causally related to *n*s (*z*s R2 A).[23] The subject is not (directly) related to mind-independent — non-dispositional ground *n*, but rather to much more intelligible and comprehensible mind-independent (manifested) properties of the object, *z*. Now the four categories of illusion are as follows:

---

[21] That there are such things as molecules (constituents of objects), which have a certain arrangement has been very recently warranted by direct evidence (after the electron microscope invention), however this claim is a very old one, with an indirect theoretical establishment, that is by *inference*. Science has shown in the past, as in present, to be successfully led by inference from that which can be seen to that which cannot. To conclude:

- Statements about *y*s are always true.
- Statements about *y*s in the case of veridical perception are true if and only if are reducible to statements about *z*s. As one might point out an analogous reduction holds also in the case of illusory perception of the third kind (and therefore justifying its objective-veridical standing). In this case (especially where science remains silent) we must seek the answer to one's cognitive faculties (mental processes, conceptualisation and consciousness) as well as to one's past experience.

[22] With two provisions:

- A is aware of the context.
- A is conscious of been already acquainted with the object at least once.

[23] Notably, R2, stands for a functional, causal etc. relation between the subject A and *n*s and *z*s respectively.

- The case where there is a subjective standing of phenomenal properties. In this case A sees B and takes it to be C — there isn't an interpersonal convergence in judgment (*y*s are not reducible to *z*s).
- The case of hallucination where there might be no presence of an external object whatsoever (*y*s are not reducible to *z*s).
- The case where there is an objective standing of phenomenal properties. In this case A sees B & there is B & B misrepresents reality (*y*s are always reducible to *z*s) e.g. the bent stick in the water where there is B(light of a certain wavelength) which is veridically perceived. We might call it the case of veridical illusion.
- The case of contextual illusion where past experience and context commit one to perceptual error (e.g. 'T /-\ E C/-\T' is taken as 'THE CAT' where '/-\' is taken both as 'H' and 'A'(*y*s are not reducible to *z*s).

As a concluding remark, the identification of colour properties, is not to be done as in the case of the dispositional theory of colour with the disposition (with the microphysical structure of its surface) of the object to emit light of a certain wavelength but rather with the manifested property. In this case with the *light* emitted from the surface. That is to say, not with the dispositional properties of the objects but rather with the manifested properties (manifested properties though are not to be confused with appearances). With that formulation McGinn's point of the inescapable of an error theory of colour perception (concerning a dispositional theory of colour) is inapplicable.[24]

Bibliography

1. Aristotle. 1993. *De Anima*. ed. J.L. Ackrill and Lindsay Judson. Clarendon Press-Oxford.
2. Armstrong, D.M. 1968. *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.
3. Berkeley, George. 1965. *The Principles of Human Knowledge*, *Three Dialogues Between Hylas and Philonous*, in *Berkeley's Philosophical Writings*, ed. David M. Armstrong. New York: Collier Books.
4. Locke, Don. 1967. *Perception and Our Knowledge of the External World*. London: Allen and Unwin.
5. Locke, John. 1924. *An Essay Concerning Human Understanding*, ed. A.S. Pringle-Pattison. Oxford University Press.
6. Jackson, Frank. 1977. *Perception*. Cambridge University Press.
7. McGinn, Colin. 1999. *Knowledge and Reality — Selected Essays*. Clarendon Press-Oxford.
8. Russell, Bertrand. 1998. *The Problems of Philosophy*. Oxford University Press.

[24] It is worth to note cases where people who are born blind and recover their vision, experience colours which at first float and are gradually bound to objects.

9. Strawson, P.F. 1988. *Perception and its Objects*, in *Perceptual Knowledge*, ed. Jonathan Dancy. Oxford University Press.

Dimitris Platchias

University of Glasgow

dimitrisplatchias@yahoo.com

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
http://www.sorites.org
Issue #15 — December 2004. Pp. 87-93
Johnston on Fission

JOHNSTON ON FISSION

Brian Garrett

---

1. In this discussion paper, I want to evaluate some arguments of Mark Johnston's which appear in his articles 'Fission and the Facts' (FF) and 'Reasons and Reductionism' (RR). My primary concern will be with his description of fission cases, and his assessment of the implications of such cases for value theory.

2. Johnston presents the case of fission as a paradox. Following Parfit's (1971) discussion,[1] he begins by describing Sydney Shoemaker's example in which Brown's brain is transplanted into Robinson's debrained body.[2] Call the resulting person 'Brownson'. The dominant response is that Brown is the same person as Brownson. Brownson is a unique psychological continuer of Brown, and the psychological continuity has its normal cause (the continued existence of the brain).

3. Consider a different example. Suppose that Brown is one of the (few) actual people whose brain hemispheres are equipollent (i.e., they support the very same mental functions), and work in tandem. Suppose that Brown undergoes a hemispherectomy. His left hemisphere, say, is removed and destroyed, whilst his right hemisphere continues to function as normal. We would have no hesitation in saying that Brown survived the operation and now has only one hemisphere. The resulting person is psychologically just like Brown, and his psychology is supported by the very same physical organ which (in part) supported Brown's earlier psychology.

4. Given the above, if Brown's left hemisphere had been transplanted into Robinson's debrained body (and Brown's old body and right hemisphere destroyed), the resulting person would have been Brown. But now — this is the fission case — imagine that *both* of Brown's hemispheres are simultaneously transplanted into two debrained bodies. We now have two excellent, and equally good, candidates for being Brown. So it seems that we must conclude that Brown survives as two people. This conclusion is bizarre; yet each step in the reasoning seemed plausible — so we have a paradox.

5. As Johnston notes, the paradox arises conditionally upon the following five plausible assumptions:

(A) Persons are not spooky entities; persons are thoroughly material.

(B) The identity through time of persons is an intrinsic matter.

---

1 D. Parfit 'Personal Identity'.

2 S Shoemaker *Self-Knowledge and Self-Identity* (Cornell University Press, 1963) pp, 23 ff.

(C) A person never survives spatially separated from himself in the fashion of a universal.

(D) No two persons can be in exactly the same place at the same time.

(E) At no time is a person constituted by two independently functioning human bodies. (FF: 379)

6. Responses to the paradox typically involve rejecting one or more of these assumptions. Thus Swinburne rejects (A) — thereby allowing that Brown might survive as just one of the two off-shoots.[3] Parfit, and many others, have denied (B): the relation of personal identity is the relation of *non-branching* physical and/or psychological continuity.[4] Hence, acknowledging that Brown survives in a non-branching situation (such as the single hemisphere transplant case) does not force us to say that he survives in a branching situation (such as fission). On a view of persons as universals, the consequence that Brown has two instances should not strike us as absurd — hence we should reject (C). On David Lewis' view of fission, two people occupy Brown's pre-division body, and so share a common temporal part.[5] They subsequently become spatially distinct. On the four-dimensional view of persons, it ought to be no more remarkable for two people to share a common temporal part than to share a common spatial part. So we should reject (D). Finally, one might think it acceptable to suppose the products of fission to constitute one whole (big) person, and so reject (E).

7. I will not here be concerned with responses that involve giving up (A), (C), (D), or (E). Johnston has an alternative response, which I shall presently consider, but I want first to look at his reasons against the response which recommends jettisoning (B).

8. One point Johnston makes in passing, which might be thought to lend some support to (B), is this. He notes that, in the actual world, it suffices to determine who is who over time if we trace only lines of bodily continuity. «We do not look elsewhere and elsewhen.» (FF: 381) Only intrinsic factors are relevant to determining the identity over time of an actual person. As an attempt to support (B), this point is irrelevant. Just because, in the actual world, we can always conclusively determine whether or not Jones robbed the bank by tracing the world-line of Jones' body, it does not follow that personal identity can never be extrinsically grounded, in unusual contrary-to-fact circumstances. To think otherwise would be to let a mere contingency (the fact that fission of persons doesn't occur) determine a metaphysical necessity (the supposed necessity of intrinsic grounding). It would also be to let an epistemological claim settle a metaphysical thesis.

9. Johnston then suggests that if we embraced Extreme Haecceitism about persons (EH), we would have a reason to reject (B). EH is the view that «… facts of personal identity can float free of any other layer of facts, whether they be facts about the total world process, facts about the persistence of brains, bodies or minds, or facts about processes of intermediate extent.» (FF: 391).

First, EH is a very strong doctrine, with few adherents. Rejectors of (B) are certainly not committed to EH. Second, the Best Candidate theorist (who rejects (B)) has no reason to

---

3 See Swinburne's contribution in S. Shoemaker and R. Swinburne *Personal Identity.*

4 See *Reasons and Persons*, 266-72.

5 See his 'Survival and Identity'.

accept the doctrine. EH about persons is normally understood to be a view about *transworld* identity: the view that numerical identities or non-identities between persons in different worlds do not supervene on any impersonal identities or non-identities. (So, for example, according to EH, there is a possible world, qualitatively identical to this world — containing all the same atoms arranged in all the same ways – yet containing a numerically different set of people.) But the Best Candidate theory is a theory about *transtemporal* identity. Third, thus understood, there would — contrary to one of Johnston's claims (FF: 391) — be no embarrassment if a Best Candidate theorist chose to embrace EH. Perhaps there is a world like this one in all impersonal respects (e.g., in terms of the arrangement of matter), except only that the person occupying my body in that world is not Garrett, but, say, Pamela Anderson. It can still be true that, *within* any world, the identity of a person over time supervenes upon relevant lines of continuity.

10. Johnston's central objection to the rejection of (B) as a response to the paradox is that there is no reason why we should reject (B) rather than any of the other assumptions. We could reject any one of the assumptions in the face of fission, so why reject (B)? But. there are reasons why giving up (B) is the most plausible response.

11. First, on any plausible account of artefact identity we are committed to the extrinsicness of identity. In the Ship of Theseus story, the planks of the original ship *a* are continuously removed and replaced. The removed planks are used to construct a ship *b* which is plank for plank identical with the original ship *a*. At that later time a ship *c* exists which is spatio-temporally continuous with the original *a*. We have two later candidates for identity with the original ship. In this case our dominant reaction is to identify *a* with *c*: in the case of artefacts, the spatio-temporal continuity criterion outweighs the identity-of-original-parts criterion. But had *c* not existed (i.e., had the removed planks not been replaced), *a* would have been identical with the ship composed of *a*'s original parts.[6] (This last situation is not relevantly different from that in which a ship is dismantled, transported across land, and then rebuilt. In such a case, we have no hesitation in saying that the earlier ship is the later ship.) So we are committed to the extrinsicness of artefact identity. If we think that the identity of persons, like that of artefacts, is traced by lines of continuity, we should not find it surprising that the identity of persons can be determined extrinsically.

12. Second, giving up (B) is not as counterintuitive as giving up any of the other assumptions. Rejecting (A) conflicts with the commitment of many to materialism; and rejecting any of the others involves too great a distortion of our concept of a person. Aside from split-brain cases, it is hard to make much sense of giving up the principle **one person: one body**.

13. Third, the consequences of extrinsicness are not bizarre. There is no violation of the necessity of identity.[7] It does follow that properties like *being occupied by person B* are

---

[6] This ship, though qualitatively identical with *b*, is not numerically identical with *b* — on pain of violating the necessity and transitivity of identity. (Note that even if we judged that *a* is *b* and not *c*, we would still concede that, had *b* not existed, *a* would then have been the continuously repaired ship. Either way, we are committed to extrinsicness.)

[7] Why not? If A divides into B and C (where all designators are rigid), isn't the extrinsicness theorist committed to the claim that, e.g., although A is not C, had B not existed, A would have been C? No — the extrinsicness theorist accepts the necessity of identity and holds that A is not B and that A is not

extrinsic properties. Suppose that A is neither B nor C, but that had C not existed, A would then have been B* (i.e., the person then occupying B's body). In that case, whether a particular body has the property of being occupied by B, rather than by B*, is fixed by extrinsic factors (the existence or non-existence of C). But since the identity-involving property *being occupied by B* is not a causal property of a body, it should come as no surprise if its possession depends on what happens to objects which exercise no causal influence on it.[8] (The property of widowhood is extrinsic and non-causal, and whether a woman is a widow may depend upon what happens to a person who, at the relevant time, exercises no causal influence on her.)

14. Fourth, the consequences of extrinsicness are even more palatable if we are impressed by Parfit's Reductionist View of persons. On one way of understanding this view, facts about persons are 'conceptual' or verbal.[9] Once we know all the relations of continuity and connectedness holding between A at t_1 and B at t_2, it is a purely verbal issue whether we choose to call them 'the same person', just as it is a purely verbal decision, e.g., whether to call sea-sickness 'pain'. Suppose this view is correct. The extrinsicness of personal identity may seem less odd. All it amounts to is the claim that whether or not we *call* A and B 'the same person' depends upon whether there is an equally good or better competitor for identity with A.[10] This linguistic dependency seems less threatening than its ontological cousin.

15. What of Johnston's positive proposal — his solution to the paradox?

Johnston writes: «the fission case shows that we cannot subscribe to the unrestricted versions of all these principles. [(A) — (E)] There is no consistent way of extending to the fission case just the principles which hold up in everyday cases. So there is no determinate sense to be made of the idea of what we would say about fission …. …[W]e should regard the fission case as a case of indeterminacy, a case in which there is no fact of the matter about personal identity. …Personal identity is here an indeterminate matter.» (FF: 393)

16. So, for Johnston, if I divide into Lefty and Righty, it is indeterminate whether I'm Lefty and whether I'm Righty. That is, it is neither true nor false that I'm Lefty, and neither true nor false that I'm Righty. Is this a satisfactory response to the paradox?

17. First, and contrary to what he thinks, Johnston's response does involve jettisoning (B). Suppose I divide into Lefty and Righty. Then, for Johnston, it is indeterminate whether I'm Lefty. But, Johnston agrees, had Righty not existed, I would have been Lefty (and this identity

---

C; so he cannot accept that counterfactual. But he is committed to the following counterfactual: had B not existed, C wouldn't have existed. Some find the truth of this counterfactual hard to accept.

We can make a claim similar to the first counterfactual if we use non-rigid terms to refer to B and C. Suppose that 'Lefty' and 'Righty' are non-rigid terms referring to B and C, respectively. Then we can truly say: had Lefty not existed, A would have been Righty. (The terms 'Lefty' and Righty' should be so understood in what follows.)

[8] Lefty and Righty are in different rooms of the hospital, and exercise no causal influence on each other.

[9] See D. Parfit 'Who do you think you are?'

[10] This version of the Reductionist View will also have implications for value theory: how can the fact of a person's identity over time be significant if personal identity is a verbal matter?

is determinate). This means that whether it is determinately true that I'm Lefty *depends upon* whether or not Righty exists. But this is just to give up (B). So Johnston cannot consistently criticise the extrinsicness solution and endorse the indeterminacy response.

18. Second, there are cases of lop-sided fission to which Johnston's response is not applicable. One example is the Ship of Theseus. Another, involving persons, would be the following. Suppose that my hemispheres are not equipollent. My right hemisphere sustains all (or most) of my distinctive psychology. My left hemisphere sustains only basic psychological functioning. Suppose that my hemispheres are divided and transplanted. Righty has my distinctive memories, beliefs and character. Lefty is, in contrast, a very impaired person, with the mental age of a 4 year old. The most plausible description of this case is that I'm Righty and not Lefty. Righty is, by far, the best candidate for being me. Nonetheless, had Righty not existed, I would then have been Lefty — Lefty's candidature for identity with me is good enough in that circumstance. As in the symmetrical fission case, we have a commitment to the extrinsicness of identity. Since there is a best candidate in cases of lop-sided fission, the rationale for Johnston's line is missing (there is no indeterminacy in such cases). This is an objection to Johnston, I take it, because any account of fission should cover cases of lop-sided fission.

19. Third, as well as unwittingly violating (B), Johnston's solution also violates an analogue of (D). If I divide into Lefty and Righty, then it is indeterminate whether I'm Lefty and whether I'm Righty. Suppose that I exist at t_1, and that Lefty and Righty exist at a later time, t_2. How many people exist at t_1? At least one — me. But if it is indeterminate whether I'm Lefty, then it is indeterminate whether Lefty exists at t_1. The same is true of Righty. So it is indeterminate how many persons exist at t_1. But the thought underlying (D) is that one and only one person occupies the pre-fission body. Johnston's description violates this principle. In doing so, the implausibility of his description is revealed.

20. There is a fourth reason for dissatisfaction with Johnston's response. Cases of indeterminacy in identity over time are typically cases in which *something is missing*. For example, in *Star Trek* Teletransportation, it is plausible to hold that it's indeterminate whether Kirk 1 is Kirk 2 because one half of the two strands that make for personal identity in the normal case is missing (*viz.*, normal physical continuity). But in the case of fission, nothing is missing; on the contrary, everything is present, twice over. This underscores the oddness of Johnston's preferred description of fission.

21. Parfit has argued that if we change our metaphysics of persons (e.g., reject Cartesianism and embrace Reductionism), we should change our value theory (in particular, we should come to see that personal identity is not 'what matters').[11] One of Parfit's main arguments for the thesis that identity is not what matters relies upon the case of fission. I want here to focus on Johnston's criticism of this argument.[12]

22. In describing the case of fission in RR, Johnston rehearses points made in FF. But he also makes some new ones. He writes: «… relative to our practice as it stands the fission case (i) violates the ordinary presupposition of essential unity, (ii) is as a result an indeterminate

---

[11] See *Reasons and Persons*, esp. Chs. 13-15.

[12] My own criticism of it can be found in 'Persons and Values', *Philosophical Quarterly*, July 1992, pp. 337-344, and in Ch. 6 of my *Personal Identity and Self-Consciousness*.

case, and (iii) also violates a presupposition of our future-directed self-concern by providing more than one future person to continue an earlier person's mental and bodily life.» (RR: 603)[13]

23. In this quote, (i) is proposed as a reason for (ii). But what are we to make of (i)? Fission demonstrates that persons lack 'essential unity': «entertaining fission involves entertaining the idea that each person has two (or more) subparts such that the survival of either one of these subparts in the right environment can secure the survival of the person. Part of the importance of the fission case is that by imaginatively violating essential unity it illustrates how we might not be mental or physical substances.» (RR: 602)

24. It is the last step which is controversial. Why should fission be thought to undermine the 'essential unity' or substance-status of persons? Consider an obvious analogy. The human body has two kidneys, which work together in tandem. If one kidney is removed, the other takes over, continuing its normal functioning. No one thinks that these facts about our kidneys show that human bodies are not substances; and this would remain true even if most organs in the human body worked in pairs. So why think that analogous facts about the human brain show that persons are not substances? Failing an answer to this question, we should leave (i) to one side — it can provide no support for (ii).

25. Parfit's Argument from Fission runs as follows. Suppose that I am about to divide. The prospect of division is not as bad as that of ordinary death; so my relation to my off-shoots contains what matters. But I am not identical to either off-shoot. Hence, personal identity cannot be what matters.

26. Johnston concedes Parfit's central intuition about fission: «… I would not make a significant sacrifice to have someone intervene in my upcoming fission to ensure that only the transplanting of my left hemisphere proved viable. And this is my reaction even though I believe that only then would I determinately survive the procedure.» (RR: 610) Despite this, Johnston does not think it follows that personal identity never matters. He writes: «It is one thing to conclude that in the fission case (neurally-based) R [the relation of psychological continuity and/or connectedness] and not identity is the relation in terms of which one should extend one's special concern. But it is quite another to conclude that quite generally it is (neurally-based) R that matters.» (RR: 610)

27. The reason for this is that in fission cases two presuppositions of our special concern are violated: (1) determinate identity; and (2) having only one future person continue one's mental life (this was (iii) above). «When such presuppositions are violated, future-directed concern neither determinately applies nor determinately fails to apply. It is reasonable to try to find a natural extension of such concern for such cases.» (RR: 610) Hence, «… were we ever to face fission, it would be reasonable to care about our fission products as we would care about a future self. *But this is not because identity is never what matters. Instead, this is because caring in this way represents a reasonable extension of self-concern in a bizarre case*.» (RR: 611)

---

[13] As it stands, (iii) is not correctly formulated: cases of lop-sided fission do not violate any 'presupposition of future-directed self-concern' (since identity is preserved), yet in such cases more than one future person continues an earlier person's mental life.

28. The lynch-pin of this argument is the premise that fission violates two presuppositions of our special concern ((1) and (2) above). There are two ways of arguing against this premise: show that (1) and (2) are not presuppositions of our special concern or show that fission cases do not violate them. I have argued that fission does not violate (1). What of (2)? By definition, fission cases violate (2). Is (2) a presupposition of our special concern? What does this mean? Does it mean that, normally, when I have 'special' concern for a future person there is only one person for whom I have that concern? But then why suppose that having a unique continuer is a genuine *presupposition* (i.e., necessary condition) of special concern, rather than just a contingent accompanier of special concern in the actual world? We need an argument if we are to accept the former view — and that Johnston fails to provide. Parfit's Argument from Fission has not been confounded.

29. There is an additional worry about the claim that (2) is a presupposition of special concern. If 'special concern' means 'self-interested concern', then the claim is trivially true: by definition, I can only have self-interested concern for myself. If 'special concern' means 'strong concern I can have for myself and others', then the claim is false, but for reasons that have nothing to do with fission: I have strong concern for many people who are not R-related to me. The fact that Johnston's criticism is open to this dilemma reinforces my claim that Parfit's argument has not been properly addressed.

30. Johnston has advanced the following three claims:

(1) Rejecting (B) is an arbitrary response to the paradox of fission;

(2) Fission cases involve indeterminate identity;

(3) *Contra* Parfit, fission cases have no implications for value theory in the actual world.

I have argued that (1) and (2) are false, and that (3), if true, is not true for any reason that Johnston gives.

Bibliography

Garrett, B. J. Personal Identity and Self-Consciousness (Routledge, London & NY) 1998

Johnston, M.: 'Fission and the Facts' *Philosophical Perspectives* **3** (1989) 369-97

---: 'Reasons and Reductionism' *The Philosophical Review* **101.3** (1992) 589-618

Lewis, D : 'Survival and Identity' in A. Rorty (ed.) *The Identities of Persons* (Berkeley, University of California Press: 1976)

Parfit, D : 'Personal Identity' *The Philosophical Review* **80.1** (1971)

---: *Reasons and Persons* (Oxford, OUP: 1984)

---: 'Who do you think you are?' *The Times Higher* 11/12/1992.

Swinburne, R and Shoemaker S: *Personal Identity* (Basil Blackwell, Oxford: 1984)

Brian Garrett

Australian National University

Brian.Garrett@anu.edu.au

SORITES

An Electronic Quarterly of Analytical Philosophy

ISSN 1135-1349

COPYRIGHT NOTICE AND LEGAL DISCLAIMER

---

(01) ***SORITES*** is not in the public domain. It is the intellectual property of the **SORITES** team. In accordance with Spanish and international Law, all issues of ***SORITES*** are Copyright-protected throughout the Planet.

(02) Each issue of ***SORITES*** taken as a whole belongs to the electronic publisher, namely the SORITES team.

(03) However, the individual papers published in ***SORITES*** remain the intellectual property of their respective authors, with only these two restrictions: (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 mentioned as the primary source; and (ii) The authors agree to be bound by the other terms and provisions contained in this Copyright Notice, esp. to grant readers the entitlements warranted by clauses (06), (08), (09), and (19), and to abide by the electronic publisher's claims and commitments set forth in clauses (14) and (15).

(04) The authors of the included papers and the electronic publisher, the **SORITES** team — whether jointly or separately, as the case may be — hereby reserve all rights not expressly granted to other parts in this Copyright Notice.

(05) In compliance with Spanish Law, this issue of ***SORITES*** has been legally registered, three diskette-copies being deposited with the competent authorities, namely the «Deposito Legal» office of the Autonomous Community of Madrid, c/ Azcona 42. (Legal Deposit Registration: M 14867-1995.)

(06) A license is hereby granted without fee for anybody to freely make as many unmodified copies as they wish of this issue of ***SORITES*** **IN ITS INTEGRITY**, give such copies to anyone, and distribute this issue of ***SORITES*** via electronic means or as printed copies, **PROVIDED** no part thereof is altered or omitted, and especially **NEITHER THIS COPYRIGHT NOTICE NOR ANY OF THE COPYRIGHT BOXES ON TOP OF THE PAPERS IS REMOVED, AMENDED, OR OBSCURED**.

(07) In this context, *the issue of* ***SORITES*** *as a whole* is meant to consist in: either (i) a single file (be it its official version as a WordPerfect 5.1 document or any unofficial version released by the **SORITES** team as an undivided file); or (ii) a collection of files produced by slicing one of the entire-file versions in order to facilitate handling, browsing or downloading. In the latter case, the conveyor is bound to distribute the whole collection. (In this context printed copies of this issue of ***SORITES*** are taken to be equivalent to electronic copies, their distribution being subject to the same conditions.)

(08) This issue of ***SORITES*** may be sold for profit or incorporated into any commercial material only with the previous explicit consent granted by the SORITES team. Otherwise, no fee may be charged for its circulation. An exception is granted to non-profit

organizations, which are hereby authorized to charge a small fee for materials, handling, postage, and general overhead.

(09) Private copying of single papers by any lawful means is allowed when done in good faith and for a fair use, namely for purposes of teaching, study, criticism or review.

(10)

[i] No copy or partial copy of a detached paper may be publicly distributed or circulated without its author's acquiescence.

[ii] The SORITES team hereby allows any such distribution, subject to the author's consent, provided the copyright box on top of the paper is literally incorporated into the distributed copy or partial copy.

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

(11)

[i] No part of this issue of ***SORITES*** may be delivered to a plurality of individuals — whether in writing or by any other means — unless the source is clearly and explicitly acknowledged.

[ii] In particular, no part of this issue of ***SORITES*** 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 ***SORITES***, 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 website: **http://www.sorites.org**.

(12) Any use of copies or partial copies of this issue of ***SORITES***, or of papers therein included, not explicitly permitted in this Notice, will be a violation of the authors' and the electronic publisher's intellectual property; perpetration of, or complicity with, any such violation will be regarded as forgery or plagiarism — which, besides being, in any case a civil tort, may be a crime under current legislation.

(13) This issue of ***SORITES*** is provided «as is», without any guarantee of any kind. The electronic publisher, the **SORITES** team, disclaims all warranties, whether expressed or implied, including, without limitation, the implied warranties of fitness for any particular purpose with respect to the papers included in this issue. By furnishing this document, the **SORITES** team does not grant any license or endorses any commitment except in so much as explicitly set forth in the present Copyright Notice.

(14) The electronic publisher, the **SORITES** team, does not necessarily agree with the authors' views or arguments. The electronic publisher cannot certify the accuracy of any quotations or references contained in the papers.

(15) Each author vouches alone for the originality of the papers they submit to ***SORITES*** and for their compliance with established Copyright laws. Acceptance of a manuscript is done in good faith under the assumption the originality claim is truthful. The electronic publisher — i.e. the **SORITES** team — does not pledge itself for the accuracy of such declarations.

(16) Neither the **SORITES** team nor the authors are liable for any real or imaginary damages suffered as a result of downloading, reading, using, quoting or circulating any materials

included in this issue of ***SORITES***. The user assumes, at their own risk, full responsibility for the proper use of this issue of ***SORITES***.

(17) Downloading, reading or in any other way using this issue of ***SORITES*** or any part thereof entails full acceptance of the hereinabove stated terms and provisions. If, after downloading a file containing this issue of ***SORITES*** or a part thereof, a user fails to agree to the conditions contained in this Notice, they must discontinue using the material and irrecoverably erase or destroy the downloaded file, so as not to occasion any third-part's unfair use thereof.

(18) Although, thanks to a permission kindly granted by the system's administrators, *SORITES* is displayed at the internet host <www.ifs.csic.es/sorites/> belonging to the Spanish research institution CSIC, the magazine is not sponsored or endorsed by the CSIC, the only owner and publisher being the **SORITES** team, under the auspices of no academic establishment.

(19) A specific license is hereby granted for this issue of ***SORITES*** to be freely mirrored at any website, provided all conditions stated above are fully adheared to. No previous approval of the **SORITES** team is required for such a display.

Madrid. April 10, 1995

(Updated: December 31, 1996; April 27 1998; October 3, 2002; December 31, 2004)

The **SORITES** Team

RELEASE NOTICE

This issue of ***SORITES*** is made available in several formats, but its only official version is the WordPerfect 5.1 document released with filename:

sorite15.wp

Although each version, whether official or not — as initially released today (2004-12-31) 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.