Issue 2(2)

ISSN 1748-9393 (Online)

British Journal of Undergraduate Philosophy

Editor: Andrew Stephenson
University of Oxford

Journal of the British Undergraduate Philosophy Society

British Journal of Undergraduate Philosophy

Journal of the British Undergraduate Philosophy Society

bjup@bups.org
www.bups.org/BJUP

Consultant Publisher: Robert Charleston
robert.charleston@bups.org

ISSN 1748-9393

Contents

Editorial

You might be forgiven for thinking that there is nothing less complicated, nothing less dangerous and unsure, nothing more downright secure and run-of-the-mill than a band's second single release from their second album. Sure, there has been some shuffling in the line-up, but surely this is commonplace in pop. And after all, their first album did well, and the last single promises good and interesting things. But here lies a problem: promises have to be kept. Sometimes of course this is not a problem at all – I can quite happily vow to put the kettle on. Only when the precedent is high and the promise great does there arise even a semblance of a problem. Unfortunately, or rather fortunately, here the precedent *is* high and the promise *was* great. And so, although it might correctly be supposed that our BJUP has comfortably settled-in for its second year, this victory was not bloodless. Problems, however spectral, call for solutions, and in this issue I have taken the liberty to shake things up… just a little.

Some of you may remember that in my first editorial I made unsubtle hints at our desperate desire for book reviews, and I am very pleased to say that henceforth I will be confident in the efficacy of such ungraceful begging. In this issue we present a veritable triumvirate of book reviews, all on the philosophy of religion and all by BUPS committee members. This section of the journal is very important to us for several reasons, all of which mirror the reasons it is important for other journals, ones aimed at practising academics. Not only do book reviews offer authors a chance to be published with a work that requires slightly less intense commitment, they also offer readers a chance for a break from the very serious attention required from full philosophical papers. But light-relief is without a doubt a book review's secondary function. Just as for academics book reviews can present a manageable way to keep up with and assess for importance and relevance the volumes upon volumes of material being printed, so for undergraduates they can give a glimpse into subject areas which might come across on their particular syllabus but which nevertheless can contribute to the more *general* philosophical understanding that is often so useful. And one should never underestimate the role serendipity can play in shaping work and making it original. So book reviews can do all this as well as convince you that you need or needn't read this or that book!

The first portrait in our triptych is of a book by Professor Robin Attfield on creation, evolution, and meaning, and it is painted by Craig French. I am especially excited about this review because it is our first commission, in this case from Ashgate. French engages with Attfield's book at a thoroughly philosophical level, subjecting proposals and arguments to close scrutiny. Particularly, French is respectful but critical of Attfield's approach to the reconciliation of science and religion via a pseudo-mediaeval philosophy of language. For example, Attfield exploits distinctions between temporality and atemporality, between transcendence and immanence, and between the employment of key terms analogically, equivocally, and univocally. But French sees unresolved coherence-tensions in this attempt at reconciliation taken as a whole.

Similarly in the next review, this time of a book by Reverend Canon Brian Hebblethwaite, Carl Baker tackles an attempt to reconcile philosophical theology with the commonly misrepresented Christian doctrines, such as the trinity and the incarnation. Baker fully endorses the book's aim to show that such doctrines are not obviously meaningless or incoherent, as is often supposed, but he does gently question whether this project could ever lead to conversion. Perhaps no academic book can fulfil this role, and Baker thinks that Hebblethwaite's collection is certainly of the right kind and at the right level for undergraduates, whether studying philosophy or theology.

Finally, Andrew Turner reviews three of the many books entrenched in what I have called on the contents page 'the Dawkins debate'. Again this review is specifically tailored to its audience. Turner concludes that whilst this debate, and the books therein, are intensely interesting and relevant for many important issues in today's society, it does not (and nor do the books therein) live up to the rigour or speciality of what we have come to expect from philosophical texts – the juxtaposition is that of the philosophical with the sociological.

Although all three reviews are in varying degrees and on various points critical and praising, a single and rather heart-warmingly united message comes forth: philosophy of religion is very worthwhile.

Having waxed so lyrical on my joy with the book reviews I will be overly brief in my summary of the papers. We have works in the philosophy of language, epistemology, metaphysics, political philosophy, and phenomenology. We have works firmly rooted in the Continental tradition and we have works firmly rooted in the Analytic tradition, and we have work that spans both traditions. We have work by British students, foreign students, and visiting students. We have long works and we have short works. All in all, we have lots of work, and I am sure you will all be delighted to get your teeth into it.

To kick us off we have the winner of the 2007 BJUP essay contest (more of which below). David Birch asks what we have learned about belief since the famous interjection of scientific realism and Kripkean theories of fixed reference into the philosophy of mind, language, and epistemology. Birch argues that we have indeed been shown an inconsistency in folk psychology, but interestingly that this fact is rather irrelevant, or at least inconsequential. Then Levno Plato tests the endurance of utilitarianism when faced with difficult counter-examples. Due to the mass of literature and the blanketing nature of the terms there is a lot of groundwork to be done here, and Plato does it with a confident clarity that any philosophy student will be admiring and perhaps envious of. In a wonderfully concise paper, Alexis Artaud de La Ferrière-Kohler shows us what is wrong with Locke's account of personal identity – illuminating examples and analogies abound. Next is Keith Wilson's much longer piece on the phenomenology of attention. Wilson's excellent paper is a phenomenological investigation into the nature of attention and its role in human perceptual awareness. Attention, you will see, is very important indeed. Our fifth paper is another shorter piece, wonderfully free of pomp and circumstance. Alex Rubner's paper is also of extremely contemporary relevance, dealing as it does with issues about belief, knowledge, and assertion that Timothy Williamson and John Hawthorn are currently publishing on. Then Jessica Woolley asks a question, and just like that we are transported away from the dreaming spires of Oxford and to the dreaming minds of the Continent. By way of answer to her arresting question, Woolley contends that Laing's notion of ontological security is plagued by problems, not all of which can be conclusively overcome. And last but not least, so far as the papers go, Mirja Holst informs us that Plato's beard is not generally misdirected; that is, for all that Willard Van Orman Quine has to say about it. If you have no idea what this means, then turn to page 186 and start reading.

Separate mention is required of Andrew Bacon's introductory article on formal metaphysics, which follows the papers and precedes the reviews in this issue. The BJUP has an established habit of including such articles where other journals would refer them to textbooks or companions. In the past we have published introductory articles – written *by* undergraduates, but more importantly, *for* undergraduates – on topics as various as Wittgenstein, formal logic, and the divide between Continental and Analytic philosophy. Bacon's article is very much part of this tradition, but it is also unique in two ways. First, it is much longer, more detailed, and more involved. This is in turn a result of the fact that it is more difficult than the others have been and requires some background in formal logic and a tiny bit of set theory. The hope is that for all the readers that are excluded by this requirement, there will be several more to take their place who find the article invaluable. And of course such a background is only required at all if you want to understand *everything* in the article – it will still be incredibly interesting and informative for everyone. Second, it acts as a sequel that should perhaps have been a prequel, like the new Star Wars movies, or perhaps like the song on the second album that extrapolates on the chorus to a song on the first album. The article introduces the mereological concepts of formal metaphysics that Bacon used to explore the viability of endurantism a couple of issues ago. The paper and the article are literally made to be read together.

And so to this small, even tiny extent, I have taken the liberty to shake things up, to ensure the fulfilment of our promise, and to produce a hit.

But before I leave you to your reading and your thinking, it is my pleasure to announce the winners of the 2007 BJUP essay contest! As I mentioned above, David Birch nabbed the first prize with the paper that follows this editorial. He wins two hundred pounds and a subscription to *The Philosopher's Magazine*. Jack Farchy won second prize with an original diagnosis of the ailments of existential predicates and the statements that contain them. He wins fifty pounds cash and fifty pounds in book tokens. And Reema Patel gets thirty pounds in book tokens and a year's subscription to our very own BJUP with the third prize. Her paper pursued Bernard Williams' influential ethical internalism. We look forward to publishing Farchy's and Patel's papers in future issues. For now, please turn to the back of this issue – page 237 – to see a winners announcement page, where you will find further details about our winning authors, their papers, and the contest's sponsors, who helped make these prizes possible.

Without further ado, I hand you over to the capable and more philosophical hands of our undergraduate authors, without whom, without a doubt, *none* of this would have been possible *at all*.

British Journal of Undergraduate Philosophy, 2(2), July 2007

What can Putnam and Burge tell us about belief?

Winner of the 2007 BJUP essay contest

David Birch
University of St Andrews
db41@st-andrews.ac.uk

In light of the arguments of Putnam and Burge, some theorists have made the distinction between two types of content: broad and narrow. These categories designate content which is individuated with respect only to the individual (narrow), and that which is individuated with respect to the individual taken in a certain context (broad). Analogous are the distinctions of *de re* and *de dicto* belief ascription. In *de dicto* ascriptions the semantic content of the subject's belief are taken privately, characterising the belief of the subject through her own eyes.[1] On the contrary, *de re* ascriptions take the semantic content of the subject's belief publicly, such that the belief of the subject is related to her context (believing *of* as opposed to *that* – see (1) and (2) below)). In the arguments of Putnam and Burge (henceforth 'Purge') *de dicto* ascriptions are taken broadly, which results in an apparent tension with the private nature of *de dicto* semantics. In this discussion I shall be attempting to resolve this conflict by searching for a type of content that can maintain the solipsistic[2] nature of *de dicto* semantics; that is, I shall be looking for a workable account of narrow content to function as belief content. Finding this project untenable, I shall suggest that the Purge considerations expose an inconsistency in Folk Psychology. This

[1] Generally speaking, this means that what might ordinarily be taken as synonymous cannot necessarily be taken as such (and so cannot be substituted in) when we're dealing with contents of the subject's belief. This is because these terms may not be taken as synonymous by the subject herself. In other words, the *de dicto* belief will express the *de re* belief in the subject's own terms.

[2] Rather lazily I will talk of 'private semantics' and 'solipsistic semantics' interchangeably. I take them both here to amount to the same thing; namely, the content of belief as based in the subject's world-view irrespective of how the world actually is.

inconsistency, I suggest, illuminates the pragmatic nature of Folk Psychology which, in turn, illuminates why we should approach a science of behaviour as eliminative materialists.

I shall proceed by giving an exposition of Purge's arguments and their bearing on *de re/de dicto* (I), searching for content narrow enough for *de dicto* belief in the manner of (a) descriptivism and (b) phenomenalism (II), questioning the prospects and reasons for eliminativism (III), and finally concluding (IV).

I

Putnam (1975) asks us to imagine a world just like Earth (Twin-Earth) except that the chemical composition of the watery stuff there is XYZ, not H_2O. NN is a resident of Twin-Earth, and on Earth is her doppelganger N; both are ignorant of the molecular composition of the watery stuff around them. When N thinks 'water is wet', it is intuitive to say that her thoughts are about H_2O and not XYZ; and vice versa for NN. This being the case, we must hold that the content of one's thoughts and beliefs are *not* wholly determined by one's internal properties, but that one's environment plays a role in shaping one's mental content – we must individuate mental content with respect to one's natural environment.[3]

For Burge's thought-experiment (1979) we are asked to imagine that N and NN occupy different linguistic communities. N is an English speaker with many true beliefs about arthritis as well as the belief that she has it in her thigh. NN's beliefs are homonymous, though in her community 'arthritis' denotes a condition which blankets both arthritic conditions and certain muscle conditions. We cannot attribute to NN the belief that she has arthritis in her thigh since this would make her true belief false. However, we would say that N believes that she has arthritis in her thigh. Thus, the contents of one's beliefs cannot wholly be a matter of one's internal properties – we must individuate mental content with respect to one's linguistic community.

[3] Putnam's initial story focuses on meaning; it was later applied to mental content.

What is the bearing of these arguments on *de dicto* belief?[4]

It is often held that the mark of a *de dicto* ascription is that it precludes substitution *salva veritate*. For example, if N does not know that $x = y$, substituting y for x in (*) yields a false sentence:

(1) N believes that *Fx*.

De re ascriptions on the other hand exhibit no such semantic feature. Fr example:

(2) N believes of x that *F*.

Here '*x*' is not featuring in the singular term ('that *F*') which refers to N's belief, so it can be substituted without upsetting that reference. It is sometimes taken that, with *de dicto* ascription, a disquotation principle can determine its truth-conditions; namely, that N believes that p iff N assents to 'p'. Since *de dicto* is meant to deal in private content, this principle is then meant to serve as the means of determining that solipsistic aspect of N's belief which properly characterises N's conception of the world. The Purge arguments have important consequences for this. In these arguments N and NN would both assent to 'p'. However, this would mean something different in their individual contexts, and so they would be taken to have different belief contents. However, as the arguments stipulate, they are intrinsically identical and so, intuitively, they conceive of the world identically. This suggests that the disquotation principle is inadequate for its task, it fails to cut beliefs fine enough to properly characterise one's private belief state. Our interest here then is to work for an account of that content

[4] Both arguments conclude that the contents of our beliefs are not wholly determined by (should not be individuated with respect to) our internal properties. Burge's case does, however, give rise to a more pervasive phenomenon. For Putnam, the implication is that our mental contents depend on the nature of the natural kinds in our environment. Burge's case carries none of the implicit metaphysical baggage and, furthermore, will seemingly apply to any term of the language. Burge's point is more easily taken: we don't have to imagine what the content would be in light of the way the world is, but only in light of our linguistic community and in terms of our actual practices – a far more tangible thing. Despite these differences, for our purposes we can just bunch them together as 'Butnam'.

which the disquotation principle was intended to safeguard. Why should we think that there is content narrower than that which N and NN assent to?

Holding that all belief is broad conflicts with our understanding of the type of things that beliefs are; indeed, it conflicts with that very understanding which underpins *de dicto* opacity. Holding that all content is broad further appears to falsely attribute irrationality. Shaping this point into the form of Moore's paradox, let us imagine that N has acquired the term 'H_2O' but doesn't know that it and 'water' are coreferential. N then asserts:

(3) Water is wet, but I don't believe that H_2O is.

On the broad-only reading, this statement is as paradoxical as Moore's original example; but the statement seems reasonable given N's epistemic state. The broad theorist might just bite the bullet and say that N is being inconsistent, but N's assent to the law of non-contradiction is compatible with an assertion of (3). Clearly N is rational, and it seems that this could be accounted for if we were to consider N's private belief content. In what ways, though, does a broad-only account conflict with our understanding of belief?

Two features of our concept of belief look threatened on a broad-only reading: (i) self-knowledge and (ii) the relation between belief and action. These two related aspects inform the view that belief *de dicto* deals in private semantics. The point of self-knowledge is linked to privacy in that our total belief state is taken to consist in how we conceive of the world, and so to properly characterise our beliefs one is required to go through a semantics indexed to this conception. To say that we have access to our beliefs is just to insist (somewhat tautologously) that we have access to the way we conceive the world – as Wittgenstein says: 'One can mistrust one's own senses, but not one's own belief' (1953, p.162). Furthermore, relating (i) to (ii), it can be noted that were we to lack such self-knowledge our actions would appear mysterious to us. This is an interesting point. Loar (1988) has argued for narrow ('psychological') content through the observation that we can understand an expressed explanation of an action without knowing its context of origin; Wilson (1995, pp.104-5) has, however, contended that understanding in such cases is still reached through a broad reading, albeit one more vaguely understood.[5] Whether we accept Wilson's point or not, it cannot be extended to the first-person case. We can imagine that N has an incomplete grasp of her words (though enough of a grasp to get by). But in this case *no* broad reading is available to her, even though presumably she can understand *her own* actions.[6]

Both (i) and (ii) have been argued to be compatible with broad content. Lepore and Loewer (1986, p.611) argue that N could know the contents of her own thoughts whilst being ignorant of the semantically pertinent features of her context. Since she knows that 'water is wet' is true iff water is wet, she knows the content of her belief that water is wet. However, we can apply similar considerations to see how this approach does not work. Since knowing the truth-conditions of an assertion is tantamount to knowing its meaning, the knowledge needed to understand its truth-conditions exposes the semantic knowledge one has in relation to that assertion. What N knows is the truth of 'the proposition 'water is wet' is true iff water is wet', but the knowledge needed here is just knowledge of a general semantic principle.[7] Thus knowing its truth does not amount to knowing the meaning of either side of the biconditional. N still lacks knowledge of the contents of her thought since understanding the truth-theoretic relation between what is used and mentioned does not amount to understanding that which is being used and mentioned. With respect to (ii) it has been argued by Stalnaker (1989) that we can explain, say, N's going to get the mop through the *broad* belief that there is water in the basement. Here we can agree but also ask that the intimately related notions of causation

[5] For example, suppose that we read in N's diary, 'Arthritis in thigh, went to doctor'. Wilson claims that we can understand this, regardless of N's context, through taking 'arthritis' in a wide, albeit less fine-grained sense and coupling this with the generalisation that if a person has a disease and believes that a specialist can treat it, then, *ceteris paribus*, that person will see a specialist. We simply take N to be such a person.

[6] Perhaps this is put too strongly. I want to avoid the implication that our epistemic relation to our own actions is ideal. Put more weakly we can just note that her actions wouldn't become more comprehensible if she became more semantically aware.

[7] There will also need to be an understanding of certain syntactic principles.

and explanation be distinguished. We happen to explain actions through belief because we understand that beliefs cause actions,[8] but explanation might become modally detached from causation. N would have got the mop even if there was XYZ in the basement, and so (tentatively) going by a counterfactual account causation, the belief that there is water in the basement cannot have played the requisite causal role. (i) and (ii) look to be substantive issues, so what we are looking for is an account of content narrow enough to play the role in *de dicto* belief such that it: (i*) respects self-knowledge, (ii*) maintains the causal link between belief and action, and (iii) issues in sensible judgments of rationality. Our project is thus one of aiming to reconcile various conflicting elements in our belief-concept – conflicting elements which the Purge arguments have drawn our attention to.

II

McDermott's (1986) suggestion that narrow belief be taken as *de re* beliefs about our inputs and outputs will not do for us. *De re* beliefs fail to satisfy our criteria (namely (i*)). But more than this, we should be highly suspicious of accounts of belief that break out of intentional vocabulary in this way (for example by severing the link with propositions).[9] Unless we fix identity conditions through related intentional categories, we risk just changing the nature of our concept rather than reconciling its *prima-facie* conflicting elements.

It is often understood that with *de dicto* belief, a proper name can be substituted for a definite description. This relates to *de dicto* belief dealing in private semantics: the relation between the object and content of belief is mediated by a mode of presentation.[10] It would be a natural thought to try and extend this to our case. We might try and take N and NN to share the narrow belief that 'the boat-ridden,

[8] More precisely, that belief has a *causal role* in actions.

[9] He concedes that for these *de re* beliefs, there are no corresponding *de dicto* counterparts. Later considerations will give us reason to find such suggestions very fishy indeed.

[10] For example, N might think of Joseph Conrad as the author of *Lord Jim*, and so in terms of a semantics indexed to her total belief state, 'Joseph Conrad' means 'the author of *Lord Jim*.'

sometimes salty, transparent stuff...[around here] is wet' (where the indexical is needed to determine the correct broad content in their respective contexts). However, as Lepore and Loewer point out, the language used in the description itself looks broad. Clearly the language of narrow content will have to be semantically immune to contextual variation. Fodor has suggested that content expressed in those terms denoting phenomenally accessible properties might do the trick.[11]

Fodor's thought seems to be that such terms will be synonymous across contexts and serve as means of trimming content fine enough to stay constant from N to NN. For example, 'water' would then be characterised in terms of its phenomenal properties such as being transparent, odourless, etc.[12] However, it has been argued that even terms such as these are not immune to Twin-Earth treatment. For example, on Twin-Earth atmosphere might alter the wavelength of light such that things on Earth that are red look green on Twin-Earth. Consider NN saying 'Roses are red'. How do we translate 'red'? Both Lepore and Loewer and McDermott suggest that we should translate it as 'green', otherwise we will end up attributing many false beliefs to NN because those things that NN thinks are red are actually green. Thus, although N and NN may both look at a rose and be neurophyisologically identical, they will have different belief content. The objection, however, is questionable.

We understand that the colours of objects can appear different in abnormal conditions, but an object has a true colour (the predicate of which has a judgment-dependent extension) which obtains under normal conditions. Thus our colour ascriptions are elliptical. For example, when we say that 'the book is red', we mean that 'the book is red [under normal conditions]'. However, 'normal' is an indexical better expressed as 'that condition which *commonly* obtains'. On Twin-Earth the condition which commonly obtains is one by which there are

[11] Unfortunately the suggestion is in an unpublished manuscript I've been unable to get my hands on. Should the reader feel particularly determined, what you're after is called 'Narrow Content and Meaning Holism.'

[12] Obviously such terms will not avoid Burge's case; however, begging the reader's indulgence, we should think in terms of communities that lack the semantic deference involved with Burge's story.

certain atmospheric conditions. To then translate their term 'red' as 'green' would, in fact, attribute Twearthians with many false beliefs since, for example, their roses are not red under *their* normal conditions. Perhaps one would want to translate their 'normal conditions' into ours, but this is poor play. The truth-conditions of their claims should be taken from their assertoric context, just as I relate the implicit 'here' to the context of my Californian grandfather when he tells me 'it's sunny'. This point will apply to all phenomenal properties since they all change under abnormal conditions. That problem aside, there is a more serious concern for Fodor.

The concern is whether we are going to be able to account for our concepts in purely phenomenal language. Part of our concept of water is that boats sail on it, it can be salty, that it is sometimes treated with fluoride, etc. To fully classify our water-concept (and so get the narrow content of belief right), given these conceptual-interrelations, will be a practically impossible task – in practice, we shall never be able to pin-down the narrow content of belief (if this is the case, how will such an account satisfy (i*)?). Even supposing that a phenomenal breakdown of these terms is possible, what about non-observationally derived beliefs like '2+2=4'? However this might be phenomenally accounted for, it will fail our conditions since we do not think (*de dicto*) of arithmetic in any such way.[13] Are there any last resorts?

Perhaps we could *think*, for example, 'water is wet' and name the thought assertion. Calling the assertion *F*, we could then say that 'N believes that *F*'. Since we are taking the content of belief experientially, N and NN can share the same belief in this sense since their experiences will be identical – they will both believe that *F*. Furthermore, other people can believe the same thing since they can instantiate type-identical states. But what are the identity conditions for these states? When N thinks that *F* she might be looking at a leaf or any thing else – how are we to isolate the relevant features of her state to be that which is essential for the state *thinking that F*? To make rigorous this idea, it

[13] This might go by taking a set-perceptual account. If our knowledge of arithmetic is derived by set-perception, then sets will denote phenomenally accessible properties such that a phenomenal breakdown of 1+1=2 might be {Ø} {Ø}={Ø,{Ø}}.

seems we are going to have to specify the identity criteria of the state in terms of the thought sentence ('water is wet') – this seems fine as long as we stay metalinguistic. However, this identity criteria is too general, all those that think 'water is wet' will not be all those we would want to ascribe the same narrow content. There is the further problem of non-English thought – to relate it back to the stated identity criteria it seems we will have to go semantic, and so self-defeatingly go to the object level. I think our inability to find a sustainable account of narrow content was inevitable.

III

We were trying to give a solipsistic breakdown of the content of our beliefs. Fodor's suggestion was to pick those terms of public language which are contextually immune and so serve as public correlates to a private language. This is shaky ground. The picture this approach is working by is one of semantic hierarchies such that one linguistic level can be semantically accounted for through a more primitive one. This reductionism cannot hold (that is, if we want to keep our distance from a 'language of thought'). In use, acquisition and meaning language is too eclectic to allow that some aspects of it stand semantically prior to others. Without reductionism, however, it would seem that our private languages can only be expressed through public means. If this is the case, then it seems there is no scope for a workable account of narrow content. But are we resting too heavily on a propositional construal of belief content?

Conceptual roles and mappings have been appealed to in order to play the role of narrow content, thus jettisoning a propositional account. I advise suspicion. If we can see why we have a propensity to think of belief propositionally, we might see why giving alternate accounts of belief might be wrongheaded. Why might belief be this way? Folk Psychology arises from our interactions with others; it is a model by which we can understand the behaviour of others. To reach this understanding we must first feel that we inhabit an intersubjective world, that our experiences are congruent. This congruence is established through language – communication presupposes intersubjectivity; this is why we explain the behaviour of others through language (through beliefs *that*) since this is the point of established experiential congruence. We domesticate the behaviour of others by relating their actions back to our mutually understood forms of behaviour; namely, language.[14] If belief is bound-up in this way with language, why should accounts of narrow content that are not propositional warrant being regarded as pertaining to belief? Those components of our concept that we were out to reconcile with contextual individuation of beliefs should have further included their propositional nature. If we have to drop one of these aspects along the road to narrow content (as in mapping accounts), we have failed in our aim. But some have loftier aims than ours; some want to find narrow content for the sake of science – but have our reflections not shown that this attitude is also wrongheaded?

The Purge considerations have revealed an inconsistency in Folk Psychology. By the nature of this inconsistency and those considerations I have propounded, I think we can see the essentially pragmatic nature of Folk Psychology. In explaining its origins we can get a grip on this aspect of its nature. So, tentatively, we need a theory of mind in order make sense of the behaviour of others. Language is a means of affirming intersubjectivity and so, as we have seen, we posit beliefs that relate the actions of others to language. In relating actions back to language (beliefs *that*), we further relate that language back to our community and so make sense of the agent's actions in relation to the shared community (these Purge considerations are a more fine-grained instance of comprehension-through-intersubjectivity). So far, then, we have the propositional nature of belief and the social nature of propositions. I would now like to reverse what was said earlier about causation: we infer that beliefs cause actions because they happen to sufficiently well explain actions. What about self-knowledge? This occurs when we turn that model by which we understand others onto ourselves; to cohere with our sense of authorship with regard to our actions, we infer that we must have direct access to our own beliefs (since beliefs cause actions). The conjuring trick is then switching the

[14] We might think that insofar as a private language is possible, it would have no word for 'belief', since self-understanding comes unmediated – without a community there would be nothing in need of explanation for which beliefs would serve the purpose.

belief-model to explain our own actions. This move coupled with our sense of authorship makes it seem that we must have direct access to our belief content. However, this illusory basis for propositional self-knowledge shows how Folk Psychology is not in the business of being representational, but in the business of being sufficiently self-consistent to provide a satisfactory *working* model – one whose inconsistency (as revealed by Purge) never had occasion to be resolved for it never manifested in practice. Where there is this governing pragmatism, it makes little sense to revise or develop Folk Psychology in anyway (for example, through narrow content), for revision is only needed when the mechanics go awry. However, its inconsistency has never been manifested and so has never hindered us from getting along. Really, we just cannot say that it is broken, and if it is not broken, why fix it?

The above observations suggest that there can be no question of Folk Psychology being 'saved' (Fodor:1987, p.2). Nor can there be any question of a conflict between Folk Psychology and some, say, neurophysiological account, since what is governing their respective criteria of acceptance is different. Folk Psychology's domestic nature means that it can only be displaced by a model whose everyday use is simpler and more familiar (certainly not a neurophysiological account).[15] But why, in science, leave Folk Psychology behind? Well, why not? It has great explanatory success, but then so too does Folk Physics *at the level in which we engage with the world*. We have no reluctance in parting with Folk Physics, why should we treat Folk Pschology any differently? Perhaps it is felt that we have a qualitatively different relation to the phenomenon of Folk Psychology than we do with Folk Physics – we can see its truth from the inside. This forthright Cartesianism is unsettling. Certainly we might say that we have some

[15] Though I am an eliminativist I am not sympathetic to the way in which Churchland (1981) treats the issue. By taking Folk Psychology as any other theory whose fate could well be the grave (like alchemy, for example) Churchland misrepresents what a neurophysiological account of behaviour will do. By suggesting that Folk Psychology is the type of thing that can be *replaced*, Churchland poorly advertises eliminativism. The familiarity and comfort of Folk Psychology makes us want to grip onto it, but if Churchland emphasised that nothing will be taken from us, we would see the eliminativist cause as it is – just the conviction that science will part with our everyday conceptions in behaviour just as it has done in all other areas of enquiry.

direct relation to our phenomenal states, but this does not transmit to a direct relation to our constitutive nature – we should expect to be just as surprised by our own nature as we are by the nature of, for example, matter. Allowing the Cartesian grip to be loosened will lead to a science of behaviour whose explanatory capacities far exceed those of Folk Psychology, and will radically alter our self-image… for a few minutes. At the level at which we engage with world, our folk sciences are what frame our everyday conceptions. The eventual displacement (if it can be called that) of Folk Psychology will no more bother us than the fact that, say, space is not really Euclidean (or some such analogy).

IV

The Purge considerations have revealed an inconsistency in Folk Psychology, one which we have failed to resolve. By reflecting on the nature of *de dicto* belief we have taken this inconsistency as indicative of Folk Psychology's essentially pragmatic character. As such, that its pieces fail to fit together is not a concern; a full science of behaviour will have no room for our intentional categories. It seems it was some such lesson Wittgenstein tried to teach us long ago – these categories are essentially founded upon the way in which we engage with the world. To expect something more, perhaps some reductive or rigorous account, is to misunderstand their place in our 'forms of life'. I hope I have succeeded in propounding an interesting and different way of framing a similar point. But what of our starting problems? The various aspects of our belief-concept do not fit, but they fit enough to get along, and get along we shall. But why do we not find N irrational in light of (3)? I am sure that we have some loose notion of narrow content – nothing susceptible to theoretic formulation, but enough to get along; that is, with 'one eye on the background facts' (Loar (1988, p.574)), such as N's epistemic state. As to Davidson's (1987) objection to Putnam's spatial metaphor, we can actually concur with Putnam that 'beliefs ain't in the head' – they ain't anywhere. Cutting the pie the way we have, we see that beliefs are not the sorts of things to which we should expect ontic correlates.

Bibliography

Burge T., 'Individualism and the Mental', *Midwest Studies in Philosophy* 4 (1979): 73-121

Churchland P., 'Eliminative Materialism and the Propositional Attitude', *Journal of Philosophy* 78 (1981): 67-90

Davidson D., 'On Knowing One's Own Mind', *Proceedings and Addresses of the American Philosophical Association* (1987)

Fodor J., *Psychosemantics*, Cambridge MA: MIT (1987)

Loar B., 'Social Content and Psychological Content', Contents of Thought, ed. Grimm R. and Merill. D, Tuscon: University of Arizona (1988): 568-75

Loewer B. & Lepore E., 'Solipisist Semantics', *Midwest Studies in Philosophy* 10 (1986): 595-614

McDermott M., 'Narrow Content', *Australasian Journal of Philosophy* 64 (1986): 277-88

Puntam H., *Mind, Language and Reality*, Cambridge: Cambridge University Press (1975)

Stalnaker R., 'On What's in the Head', *Philosophical Perspectives* 3 (1989): 287-316
Wilson R., *Cartesian Psychology and Physical Minds*, Cambridge: Cambridge University Press (1995)

Wittgenstein L., *Philosophical Investigations*, Blackwell (1953)

British Journal of Undergraduate Philosophy, 2(2), July 2007

Does Broome's utilitarianism survive Diamond's objection to the sure-thing principle?

Levno Plato
University of Edinburgh
l.v.plato@sms.ed.ac.uk

As essentially consequentialist ethical theories, many variations of utilitarianism have been worked out. All have had to face objections of some kind, with most attacks referring to deontological ideas or intuitions engrained in common sense morality. Defenders of utilitarian theories have not tried merely to reject such criticism – they were surprisingly willing to alter their theories in order to accommodate such intuitionist worries. Utilitarianism has started opening the door to intuitionist and deontological worries by allowing actions that do not maximize the perceived good as long as these belong to a rule that maximizes it.

This essay will focus on a utilitarian theory put forward by John Harsanyi,[1] an intuitionist criticism against this theory presented by Peter Diamond,[2] and a defence of both Harsanyi's theory and Diamond's criticism by John Broome,[3] who pushes utilitarianism even further to include intuitionist values. I will consider problems to Broome's attempt and conclude that Broome fails to reconcile the two, which does not mean more than rejecting Broome's version of utilitarianism.

[1] Harsanyi, 1953, 1955. Harsanyi further developed his theory in subsequent works, which will hardly be considered here. It is not uncontroversial whether or not Harsanyi's theorem really entails utilitarianism. But for this paper I shall presume it does.

[2] Diamond, 1967.

[3] Broome, 1991.

Harsanyi developed and formulated a utilitarian social welfare theorem by combining expected utility theory (EUT),[4] a definition of rational behaviour in face of uncertainty and risk, and the claim that social welfare is the sum of individual utilities. In other words, Harsanyi asserts that if (i) individual and (ii) social[5] preferences satisfy the axioms of EUT and (iii) individual people's preferences determine which alternatives are socially preferable,[6] then social welfare can be represented by a function which is the sum of individual people's utility functions. This theorem presupposes that intra- and inter-personal comparisons of lives are the same and thus emphasizes the value of lives themselves rather than the relations of people to one another. Without going into the details of Harsanyi's three postulates I want to move on to Diamonds criticism, which will elucidate the relevant feature of this markedly composite theory.

Diamond criticises Harsanyi's second postulate because, he claims, it is inconsistent with some intuitions about justice. According to Diamond, intuition tells us that there is some element in the rationality of social decision making which is to differ from the rationality of individual decision making. People, he argues, also consider the '*process of choice*'[7] and not only the outcomes of choices when deciding about alternatives concerning social preferences. To clarify the criticism, Diamond gives an example which is conveniently illustrated by Broome[8] as a kidney transplant dilemma.

Imagine two identical persons, P and Q, who both need a kidney to survive. Only one kidney is available for transplant. Consequently, the person who gets the kidney lives, whereas the other dies. Who shall receive the kidney? Or, better, how is the medical practitioner going to

[4] EUT is basically an offshoot of Bayesian rationality.

[5] Harsanyi's notion of social or so-called moral preferences – i.e., preferences one has when choosing between alternatives concerning people in general rather than only oneself – involves arguments about impartiality and the veil of ignorance (cf. Harsanyi, 1953). Even though these arguments are highly interesting, relevant for this essay, and controversial (cf. especially Rawls, 1999), I will not discuss them here.

[6] This is Harsanyi's (1955) postulate c, which he calls 'individualism' and is also known as the Pareto principle.

[7] Diamond, 1967, p. 766.

[8] Broome, 1991, p. 26.

decide to whom s/he gives the kidney?[9] Two different alternatives resulting from coin tossing are proposed:

Person	Alt. (a)		Alt. (b)	
	Heads	Tails	Heads	Tails
P	lives	dies	dies	dies
Q	dies	lives	lives	lives

Following Harsanyi's theory both alternatives are equally good and the hospital practitioner should be indifferent between the two. According to Diamond's intuitionist objection, however, justice tells us that alternative (a) is strongly preferable since it gives P '*a fair shake*'[10] and such a biased process of choice as alternative (b) is unfair. Diamond's criticism highlights the relations of people to one another, which Harsanyi's theory consciously hides.

Harsanyi[11] is quite confident about his Bayesian rationality claims and flatly rejects Diamond's worries about fairness. He adds that it is implausible to ask for differing rationality conditions for individual and social choices. In the end one person dies and the other lives, no matter what the decision process was. To have a fair chance of living does not make any of the alternatives better than the other, as alternative (a) shows: one person is going to die even though both persons had a fair chance to receive the kidney.[12] Moreover, so Harsanyi thinks, '*the great lottery of life*'[13] gives everyone an equal prenatal chance to be in any position already; why should artificial lotteries be given more moral weight than nature's lottery? It seems that Harsanyi's modest desire for fairness has already been satisfied in the impartiality conditions which underlie his rationality requirements for social preferences.

[9] To make the case simpler, it is supposed that it will not affect anyone else, but the two, no matter what de decision will be and that both people want to live.

[10] Diamond, 1967, p. 766.

[11] Harsanyi, 1975.

[12] This argument is supported by Scanlon, 1998.

[13] Harsanyi, 1975, p. 317.

Harsanyi's rejection has a bitter aftertaste; did he reply satisfactorily to Diamond's worry? 'Fairness' is a complex and highly intuitive term, and it is possible to doubt that Harsanyi did it full justice in his clear cut rationality requirements or his trivializing life's lottery argument.

Here is where Broome[14] comes into the picture. He takes Diamond's intuitions about justice more seriously and analyses what is at stake when arguing over fairness. Brome elaborates a theory of fairness paying tribute to many intuitions and providing a rough idea of what is to be considered. His main suggestion is that if people have claims to goods, these claims should be satisfied in proportion to their strength. In case a good is indivisible, a lottery consisting of chances to win that are proportional to the strength of the claim is to be performed to provide a '*surrogate satisfaction*' for the losing claimants. Yet, Broome argues, the legitimacy of such a lottery depends on the importance of fairness in each case and the strength of each claim relative to other people's claims. Broome's ideas about fairness are certainly not easy to use in practice due to their intuitive and rather vague nature; but applied to the kidney dilemma they basically imply that each person is to be given an equal chance to receive the kidney.[15] And since alternative (b) does not allow for such equal chances, but alternative (a) does, (a) is to be preferred.

However, Broome[16] emphasises that to agree with Diamond and ask for equal chances in such cases as the kidney dilemma does not necessarily mean Harsanyi's enterprise is in danger. To demonstrate how Broome attempts to make Harsanyi's theory consistent with Diamond's demand for fairness it is now necessary to examine the precise detail in Harsanyi's rationality claims that entails indifference between both alternatives and is therefore Diamond's precise point of attack. As already mentioned, this is Harsanyi's second postulate, the rationality of social decision making. More precisely, it is that rationality for social choice is to conform to the sure-thing principle, a key axiom of EUT.

[14] Broome, 1990-91.

[15] This conclusion is supported by Kamm, 1985, and Kornhauser & Sager, 1988.

[16] Broome, 1991 (Unless otherwise stated, subsequent reference to Broome is to this book).

As Broome describes it, the sure-thing principle implies that outcomes of alternatives are independent of each other, that there are no properties that are not inherent in the outcomes themselves which would provide reasons to rationally prefer one alternative to another. The sure-thing principle thus says that if alternatives give equal probability to the same outcomes, then it is rationally necessary to ignore these outcomes and decide which alternative is preferable by considering only the other possible outcomes.[17] As can be seen in the present kidney dilemma, the outcome 'tails' in both alternatives are exactly the same and are given equal probability; and (given impartiality) the outcomes 'heads' are alike too. Since it has thus been established that there is no discernable difference between the two alternatives, it is rationally necessary not to prefer any of the alternatives to the other.

It is now possible to see where exactly Diamond disagrees with Harsanyi: the '*process of choice*' – i.e., fair choice versus biased choice – is not apparent when the sure-thing principle is applied. Thus Diamond rejects the sure-thing principle because, on the one hand he thinks intuition tells us that people are rationally justified to prefer one alternative to the other if the processes of choice differ, and on the other hand he is aware that the sure-thing principle conceals this information.

At this point Broome ventures to reconcile the intuitionist demand for consideration of the '*process of choice*' (i.e., a fair coin) and the sure-thing principle. Broome defends the sure-thing principle by including the alleged property of the '*process of choice'* (i.e., 'fair' or 'unfair') into the outcomes of alternatives themselves – this is what he calls the '*individuation*' of outcomes, or rather the '*dispersion*' of fairness to individual outcomes. 'Unfairness' will thus be added to the outcomes of alternative (b). By this move alternatives (a) and (b) are said to become different, and the sure-thing principle cannot be rejected in this example since it has no applicability. The added property 'unfair'

[17] Broome (1991) actually altered Harsanyi's theory (mainly the third postulate) and consequently does not talk about 'preferences' but about 'betterness'. Even though this is essential for Broome's development of Harsanyi's theory, I believe it can safely be ignored here. I think more clarity is given to the argument in this paper when I keep on using 'preference' instead of using Broome's 'betterness' terminology.

becomes a rational '*justifier*', as Broome calls it, to differentiate between the two alternatives and consequently prefer alternative (a) to (b).

Now, Broome claims that Diamond's demand to consider fairness has been acknowledged without endangering the validity of Harsanyi's rationality claims; the sure-thing principle, Broome is convinced, cannot be questioned anymore. And thus Diamond's intuition about fairness has been included in a utilitarian theory of social welfare. However, this conclusion might not be as solid as it looks.

Broome himself mentions the possible objection that by treating 'fair' and 'unfair' as justifiers to rationally distinguish both alternatives, one *wrongly* assumes that 'fair' and 'unfair' are properties of outcomes. If one were to agree to take Broomean fairness into account for social choices, and if 'fair' and 'unfair' were actually not properties of outcomes, Broome's dispersion of fairness (or unfairness) would fail, and the sure-thing principle could be rejected in the way Diamond rejects it. As Diamond seems to believe, 'fair' and 'unfair' are properties of the *process* of choice, not the *outcomes* of choice, since this feature arises only once all the outcomes of an alternative are considered relative to each other, rather than individually. Considering the outcome 'tails' in either alternative on its own, for instance, does not tell us whether or not the outcome is fair. It becomes apparent only if both, 'tails' *and* 'heads', are considered jointly.

Against this line of argument Broome asserts that 'fair' and 'unfair' are indeed properties of outcomes even though he acknowledges the interaction between outcomes. His reason for such a claim is that he takes the properties 'fair' and 'unfair' to be '*modal*'.[18] It depends on a '*counterfactual conditional*'[19] for whether an outcome has the property 'fair' or 'unfair'. That is, only if the outcome that does in fact not occur (for example, 'heads') were to occur, would the occurring outcomes ('tails') have the properties 'fair' or 'unfair'. Broome continues to explain that the properties 'fair' and 'unfair' supervene on the

[18] Broome, 1991, p.114. That is, it *might* be true of the outcomes to have the property 'fair' or 'unfair'.

[19] Broome 1991, p.114.

‘*nonmodal properties*’[20] of the outcome, and can therefore be regarded as genuine properties of the outcome. So Broome agrees with Diamond’s intuition that alternative (a) is fair and alternative (b) is unfair. But, while Diamond rejects the sure-thing principle, Broome does not.

I am sceptical of Broome’s claim that ‘fair’ and ‘unfair’ are real properties of outcomes. Broome gives the example of a wooden ship decaying at the bottom of the sea. He says this ship is inflammable even though it will not burn unless exposed to fire. This claim is highly questionable. It might very well seem absurd to assign the property ‘inflammable’ to a ship at the bottom of the sea. Such a ship, one might claim, has lost the property of being inflammable. It does not have the property of inflammability as long as it is at the bottom of the sea, or more precisely as long as it is not exposed to fire. And similarly, outcome ‘tails’ in alternative (b) can be said not to have the property of unfairness unless the relation to outcome ‘heads’ of the same alternative is considered too. But since the sure-thing principle hinders such consideration one might plausibly argue that the property cannot be assigned. Moreover, one of the arguments Broome gives in support of the sure-thing principle is that outcomes can be considered individually since one determines their value by judging what an outcome is like when it occurs, which implies that other possible outcomes did not occur and thus do not influence the value of the occurring outcome. With his claim that the properties ‘fair’ and ‘unfair’ are determined by counterfactual conditionals, however, Broome allows us to consider outcomes which will never occur jointly. This is inconsistent. Therefore, either the support for the sure-thing principle fails, or his argument that counterfactual conditionals determine the fairness property of outcomes does.

I think Broome has recognised the implications of this problem and deals with it in terms of what he calls the ‘*rectangular field assumption*.’[21] The rectangular field assumption, being another presupposition of EUT, states that any outcomes of all available alternatives must be

[20] Broome, 1991, p.114. That is, the necessary properties ‘heads’ or ‘tails’ of outcomes ‘heads’ or ‘tails’, respectively, are the bases to which properties such as fair or unfair might be assigned, once occurring.

[21] Broome, 1991, pp. 115-117.

possible outcomes of an alternative composed of arbitrarily assigned outcomes. By including fairness into the outcomes, Broome concedes, this assumption is at danger. Consider the following possible alternatives (where ‘DU’ stands for ‘dies unfairly’):

Person	Alt. (a)		Alt. (b’)	
	Heads	Tails	Heads	Tails
P	lives	dies	DU	DU
Q	dies	lives	lives	lives

Person	Alt. (c)		Alt. (d)	
	Heads	Tails	Heads	Tails
P	dies	lives	lives	lives
Q	lives	dies	DU	DU

According to Broome, the danger arises when the outcomes of alternatives (a)-(d) above are arbitrarily assigned to an alternative which might look like:

Person	Alt. (e)	
	Heads	Tails
P	DU	lives
Q	lives	DU

However, as Broome recognises, such an alternative is impossible since unfairness cannot in this case actually be a property of outcomes. Broome believes this problem is solved by accepting a possible rejection of the rectangular field assumption for the sake of saving the sure-thing principle and the dispersion of ‘fair’ and ‘unfair’.

Nevertheless, even if 'fair' and 'unfair' were true properties of outcomes, I would not be convinced by Broome's claim that this saves the sure-thing principle from Diamond's attack. Following EUT, to decide which alternative is to be the preferred one is to apply the sure-thing principle. By applying the sure-thing principle one is compelled to consider the outcomes individually. And when considering the outcomes individually, one does not realise that 'fair' and 'unfair' might be properties of the outcomes, since the other outcome, that reveals these properties to the decision maker, is to be ignored. In other words, even if 'fair' and 'unfair' were properties of outcomes, these properties would not be recognisable when considering individual outcomes, and can therefore not be part of rational decision making. So, if we want to include our strong intuitions about fairness into our social decision making, as Diamond and Broome demand, we must reject the sure-thing principle since it would otherwise conceal the property we seek to recognize.

Broome tried to reconcile Harsanyi's rational requirements for a utilitarian theorem with Diamond's rejection of the sure-thing principle due to a higher (intuitionist) demand for fairness. By accepting that fairness needs more consideration than Harsanyi included in his theory Broome accommodates Diamond's demand. By dispersing this valuation of fairness to individual outcomes Broome claims to have saved the sure-thing principle from Diamond's rejection. I have questioned the success of Broome's dispersal of fairness for reasons of implausibility and hidden inconsistencies. I would therefore suggest judging Broome's reconciliation attempt as failing and conclude by returning to the two original positions. One either has to accept that Harsanyi already included fairness in what he calls social preferences, or, if one thinks there is more to fairness than that, accept Diamonds rejection of the sure-thing principle and develop alternative rational requirements for social choice, which would, of course, present a counterexample to Harsanyi's utilitarian theorem.

Instead of trying to decide which of the two is more plausible, I highlight that the more intuitionist ideas are accepted and the more these ideas are incorporated into utilitarian theories, the more these theories might gain popularity, but at the same time they might loose their clarity and usefulness and become as obscure as common sense morality itself. As I hope to have shown, Broome's version of utilitarianism does not successfully overcome Diamond's objection to the sure-thing principle and includes too great an amount of intuitionism (about fairness) for it to be as clear and powerful as Harsanyi's. Broome's theory lessens in clarity due to the tension between fairness and the sure-thing principle and leans towards an admittedly intuitionist, but consequently rather obscure common sense morality.

Bibliography

Broome, J. (1990-91). 'Fairness', *Proceedings of the Aristotelian Society,* 91, pp. 87-102

Broome, J. (1991). *Weighing Goods,* Blackwell, Oxford

Diamond, P. (1967). 'Cardinal Utility, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment', *The Journal of Political Economy,* 75, pp. 765-766

Harsanyi, J. (1953). 'Cardinal Utility in Welfare Economics and the Theory of risk-Taking', *The Journal of Political Economy*, 61, pp. 434-435

Harsanyi, J. (1955). 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility', *Journal of Political Economy*, 63, pp. 309-321

Harsanyi, J. (1975). 'Nonlinear social welfare functions: do welfare economists have a special exemption from Bayesian rationality?', *Theory and Decision,* 6, pp. 311-32

Kamm, F. (1985). 'Equal treatment and equal chances', *Philosophy and Public Affairs*, 14, pp. 177-194

Kornhauser, L. and Sager, L. (1988). 'Just lotteries', *Social Science Information,* 27, pp. 483-516

Rawls, J. (1999). *A Theory of Justice*, 2nd edition, Oxford University Press, Oxford

Scanlon, T. (1998). *What We Owe To Each Other*, Harvard University Press, Cambridge, Mass

British Journal of Undergraduate Philosophy, 2(2), July 2007

How to be David Copperfield: a critique of Locke's personal identity model

Alexis Artaud de La Ferrière-Kohler
University of Sheffield
ega05ama@sheffield.ac.uk

In his *Essay Concerning Human Understanding*, John Locke writes that 'nothing but consciousness can unite remote existences into the same person.'[1] According to Locke, the self is one continuous thinking thing over time. What unites the present existence of that self with its past existences is consciousness. Thus, personal identity extends only as far as consciousness. This is what is striking about Locke's model – it encompasses personal identity within the disposition or act of consciousness (a mode), rather than within the body or soul (a substance). Indeed, Locke explicitly states that 'whatever Substance, made up of whether Spiritual, or material, Simple, or Compounded, it matters not' (II.27.17).

It would be beyond the scope of this essay to suggest an improvement on Locke's model. However, this essay will demonstrate why Locke is wrong to found his model solely on consciousness, which should at least indicate a direction to follow for an improvement. I presume he is correct to focus on the importance of consciousness. However, he is mistaken to posit consciousness as a sufficient criterion for personal identity. As I shall note, he fails to clearly state what he means by 'consciousness'. Though this is not devastating in itself – it merely unearths a far graver weakness. The real problem is that, regardless of how 'consciousness' is interpreted, the model is too lean to provide an adequate account of personal identity. It offers no coherent means of distinguishing between the remote existences that belong to the self and

[1] Peter Nidditch, ed. (Oxford: OUP, 1975), II.27.23. Further references to this edition are given within the text.

those that do not. In fact, when employed, the model produces an obviously defective account of personal identity.

Critics such as John Mackie have rightfully pointed out that Locke uses words loosely in his model.[2] Particularly, effort has been devoted to the task of disambiguating 'consciousness', which Locke never satisfactorily defines.

The text suggests that Locke sees consciousness as akin to memory: 'the forgetfulness Men often have of their past Actions, and the mind many times recovers the memory of a past consciousness' (II.27.23). Several times he insists on the memory of past *actions* and *thoughts* (II.17.20). Thus, personal identity is constituted by those past existences that the present person remembers, implying a mushrooming of persons within any particular man. Indeed, consider the example Locke uses of a man who becomes drunk. His drunken stupor would imply that he becomes a different person; or he might even cease to be a person, depending on how dedicated an imbiber he is. And if he wakes up not remembering the frolics of the previous night, he would again be a different person, though probably the same as the original sober person.

This raises questions concerning the psychological accountability of these persons. Certain memories must persist throughout, at least in a fragmented form. The drunken person can still speak, even if his speech is slurred, and the person with a hangover probably remembers a fraction of his drunken antics. Locke does not specifically address this issue. Nevertheless, I see no reason why his model would not allow for a distinction to be made between the mind and the self, given his neutrality on substance. Thus, we can imagine some psychological continuity within the man, while retaining different persons.

However, Locke also relates consciousness to a certain normative function within the man, noting the expressions, 'one *is not himself*, or is *besides himself*' (II.27.20). Certainly we usually expect to find the same person in the same man, most of the time. This suggests that the

[2] *Problems From Locke* (Oxford: OUP, 1976), p. 182. Further references to this edition are given within the text.

extra persons Locke talks about, such as the drunkard, are in fact deviants from the main occupant of the man. Where does this main person go while the others manifest themselves? How do we recognise him? Is the man some sort of time-share with the person in occupancy for the longest duration being the principal shareholder? That sounds rather queer. What if the man spends most of his life drunk? And where do the deviant persons go once the main occupier has returned? Psychological continuity does not account for these interpersonal relations within the man. We are left with more questions than answers. This kind of objection, however, misses the real weakness, which is to be found in the relationship between consciousness and remote existences – the model does not have a sufficient infrastructure to sustain that relationship.

Purging Locke's model of substance makes the establishment of personal identity a purely introspective project. The difficulty this creates is that it does not distinguish between self-knowledge and self-identity, both encompassed within consciousness. This dual charge is too weighty for consciousness to sustain without supporting apparatuses. As we have seen, the leanness of Locke's model does not permit such apparatuses, forcing him to reuse what he has. Thus, consciousness is not only that which unites remote existences into the self. What inevitably precipitates from Locke's model is that consciousness is also charged with discerning between those existences that have a place within personal identity and those that are alien to it. Consciousness is unable to do this properly because it is constituted by those very remote existences. It has no benchmark by which to compare the quality of those existences.[3]

Consider how personal identity is constituted in Locke by remote existences. What is a remote existence? I want to say that it is a temporally removed instance of my self, experiencing something. If I look back in my mind some ten years I can discern a child at a baseball game in Dodger Stadium. On the other hand, being President of the United States ten years ago is not a remote existence of mine. This,

[3] Bishop Butler develops an argument similar to this last point in his own critique of Locke.

because I was not President ten years ago; Bill Clinton was. How do I know the difference? Supposedly, consciousness will do that work for me. I can vouch for that boy being the same person as I because I am conscious of what his actions and thoughts were at that time. I am ignorant of the thoughts and actions of the President ten years ago. However, I see some problems here.

Say I am mistaken about being that boy. What if the memory of the boy was not an experience I myself had but one narrated to me by a friend, only I have forgotten that aspect of the story. My memory does fail me often, and promises to do so evermore as I age. As such, in this model I am clearly at risk of incorporating remote existences into my personal identity that are in fact alien to my person.

However, maybe we should be charitable towards Locke, as some critics have been, and elevate consciousness to some vague intuitive thing that is stronger than memory. Let us allow that even if at times this thing might be mistaken, that would be an exception to the rule – I can usually distinguish intuitively between events I have experienced and those I have only heard of.[4]

This is still not very convincing. A sceptic need only ask me what my thoughts and actions were ten years ago? I probably could not answer. Maybe if I thought about it, if I looked back in my notes, I could provide him with an answer. Yet, if I am legitimised in consulting exterior sources, then I could probably tell the sceptic with even more precision what the President of the United States was doing ten years ago.

So, how conscious am I of that child in Dodger Stadium? Even if I am sure, beyond a doubt that that child was I, is it really because I am conscious of his thoughts and actions as Locke claims? Hume identified this problem in his *Treatise of Human Nature*: 'Who can tell me, for instance, what were his thoughts and actions on the 1st of January 1715,

[4] Although, how I am to know whether this sort of error is isolated or represents a high proportion of my personal identity I cannot say. Locke does not suggest any empirical method for separating the tares from the wheat.

the 11th of March 1719, and 3rd of August 1733?'[5] Some things I am more or less conscious of. How much do I need to remember – half my thoughts and actions, or more than half? It seems impossible to expect anybody to remember everything about any previous point of their existence, even if it was yesterday. Attempting to quantify a level of consciousness that the thinking self has to satisfy in order to incorporate a particular remote existence into his person seems a mad endeavour.

Consider another scenario, one that is not based in error, but that presents the same problem. Imagine that I read a novel. To simplify matters, this is a first person narrative, focusing on the biography of the narrating protagonist, such as Charles Dickens' *David Copperfield.* I may read *David Copperfield* and in so reading I would become conscious of a great number of details of David's life. Indeed, I would be more conscious of David's early life than of my own.

Am I then to say that I am the same person as David Copperfield? This certainly sounds odd and the Lockean would probably not accept such a claim to be made on his behalf. He might protest that David Copperfield is a fictional character, and therefore I, a real person, cannot have the same identity as David. But what would the Lockean be saying in exposing David's fictitiousness? Surely, he is not objecting to the fact that the concept of David Copperfield has no material equivalence – Locke claims that consciousness need not be annexed to any substance, material or immaterial. He cannot claim that David is not a thinking thing, given that I am claiming to be David and I certainly am a thinking thing. In fact I think that I am the same person as David Copperfield. Locke's reasoning behind the demonstration that the Mayor of Quinsborough was not Socrates is that he was 'not conscious of any of Socrates' Actions or Thoughts,' regardless of whether or not they shared the same soul (II.27.14). How then should my claim to be the same person as David Copperfield run askew of Locke's own words? I make no reference to souls or bodies, only that I am conscious of David's actions and thoughts, and therefore the remote existences in that novel are encompassed by my consciousness.

[5] (New York: Prometheus Books, 1992), I.4.6.

I have chosen a fictional example deliberately to demonstrate that the Lockean model does not include the means of denying so outrageous a claim. This is because of the permissiveness of Locke's definition of personal identity, and the ambiguity of his use of consciousness. The point would hold if I used the example of reading a diary. Mackie dismisses this sort of attack on the basis that these 'causal links are of quite the wrong kind to constitute memory' and thus 'Locke's own theory [need not] be embarrassed by cases of this kind' (p. 184). I agree that these causal links should be of the wrong kind, but that is precisely why these sorts of cases are so embarrassing to Locke's theory. Locke never tells us what the right kind of link is for the construction of personal identity, and his model does not leave us with the means to discern it for ourselves.

Thus, Locke does not give a satisfactory account of personal identity. He offers an elegant attempt at dissolving the bond between substance and personhood through consciousness. He masterfully represents this process of consciousness while avoiding the fray over substance (though, as we have seen, only at the expense of raising a series of other uncomfortable questions). However, while consciousness alone does allow for the creation of homogenous persons, it cannot regulate the existences that constitute those persons. When we delve deeper into the constituents of consciousness, it becomes apparent that some other criterion will be required in order to form an acceptable personal identity model. What this criterion might be remains a contentious matter.

Bibliography

Butler, Joseph, *The Analogy of Religion* (London: Adamant Media Corporation, 2005)

Hume, David, *Treatise of Human Nature* (New York: Prometheus Books, 1992)

Locke, John, *An Essay Concerning Human Understanding*, ed. by Peter Nidditch (Oxford: OUP, 1975)

Mackie, John, *Problems From Locke* (Oxford: OUP, 1976)

British Journal of Undergraduate Philosophy, 2(2), July 2007

Does attention exist?

Keith Wilson
University of York
keith.wilson@mac.com

I. Introduction

In the introduction to the *Phenomenology of Perception*, Merleau-Ponty (2002: 34) states that 'Attention, […] as a general and formal activity, *does not exist*' (my italics). This paper examines the meaning and truth of this difficult and surprising statement, along with its implications for the account of perception given by theorists such as Fred Dretske (1988) and Christopher Peacocke (1983). In order to elucidate Merleau-Ponty's phenomenological account of human perception, I will present two alternative models[1] of how attention might be thought to operate. The first is derived from the works of the aforementioned theorists and is, I argue, based upon a largely inaccurate computational or mechanistic understanding of the mind. The second is drawn from the works of Merleau-Ponty and cognitive scientist and philosopher, Alva Noë, and takes into account recent neurological theories concerning the role of attention in human consciousness. On the basis of these models I will argue that attention is an *essential*, rather than *incidental*, characteristic of consciousness that is constitutive of both thought and perception, and which cannot be understood in terms of the independent faculty or 'general and unconditioned power' (*ibid.* 31) that Dretske *et al's* account requires. I will conclude by considering two potential counterexamples to my argument, and evaluating the threat that these pose to the phenomenological model.

[1] The term 'model' is intended to mean a simplified description or framework, and should not be taken to beg any important questions about the nature or basis of consciousness (for example, that it is reducible to a set of physical processes).

II. Two Models of Perception

Much of the recent literature in philosophy of mind and consciousness (for example: Dretske 1988, 2004; Peacocke 1983, 1998; Ayer 1973) adopts a particular account of the functioning of perception and attention. This account is directly descended from the views of Descartes, Hume and Locke, and to a certain extent reflects various widely held prejudices and opinions about the nature of the human body and the world in general; i.e. that they are fundamentally physical in nature. This view is also substantially influenced by the modern understanding of mechanism, and in particular the workings of mechanical devices such as the camera and audio or video recorders, as well as more recently – but perhaps even more significantly – the modern digital computer with its microprocessor 'brain'. Such devices employ a process by which initial inputs (light rays, sound waves, electrical impulses, etc.) are captured by some kind of sensory surface (a photographic plate, microphone diaphragm, magnetic tape, CCD sensor) and transformed into a covariant representation of the original signal that is stored for subsequent analysis or retrieval. Due to its relative simplicity and the obvious analogy between the workings of such devices and our own sensory apparatus – the eyes, ears, skin and so on – this model offers an attractive basis for understanding the corresponding processes of human perception. Indeed, many of these mechanical devices were substantially modeled upon or influenced by the workings of the human body – a fact which only serves to strengthen the analogy. I will call this the *snapshot model of perception* (cf. Noë 2002b: 2) due to its resemblance to the way in which a camera captures a complete image of a visual scene for later reproduction or viewing.[2]

Under this account, visual perception involves the formation of a 'picture' inside our head (the brain being at the centre of what is considered to be a primarily computational process) containing a more or less accurate representation of the external world. Although we take

[2] In the discussion that follows I will concentrate upon visual perception, but the same principles apply to other sensory modalities, such as touch, hearing and proprioception (inner-sense).

in or perceive the entire scene at once, our brains do not actively process all of this information simultaneously. Rather, we extract various salient features via the faculty or process of *attention*, which homes in on various aspects or details of the scene that our central nervous system represents to us. The conscious mind is then able to 'read off' information from this internal representation in much the same way as one might read off the information contained within a photograph, train timetable or visual display unit. Any redundant or irrelevant information is either discarded, or retained in memory for later recall and analysis. The key features of this model are that (i) the initial 'snapshot' phase creates an internal representation of the entire visual scene within the subject's brain prior to any further cognitive processing for the purpose of detecting objects, forming perceptual judgements, generating an appropriate reaction, and so on (Dretske *op. cit.* 162), and (ii) that attention is envisaged as a distinct faculty or power that extracts information from the previously captured 'sense data' (Peacocke 1998).

There are several problems with this view. As Merleau-Ponty points out, 'In order to relate [attention] to the life of consciousness, one would have to show how a perception awakens attention, and then how attention develops and enriches it' (*ibid.* 31). Since it is described in terms of objective physical processes and causal relations, the snapshot model can only explain the functioning of attention as a series of responses to stimuli, as opposed to a system that actively selects certain stimuli over others, as the model itself requires (*ibid.* 30). Secondly, it entails that we represent the world as an array of determinate and (in principle, at least) objectively verifiable data, whereas our actual experience of perception appears to contain a high degree of indeterminacy – around the fringes of the visual field, for example – and can even contain logical ambiguities and contradictions, as in the Müller-Lyer illusion, for example. Finally, by positing an internal representation of the entire visual field within the subject's brain, the snapshot model simply defers the problem of understanding attention and consciousness to this inner level in what Dennett (1991: 107) describes as the 'Cartesian Theatre'. Consciousness, in the form of attention, becomes an homunculus, or 'little man', that is 'looking out' at the sense data just as we are 'looking out' at the external world; an explanation which fails to resolve anything. To account for the apparently 'miraculous' (Merleau-Ponty *op. cit.* 30) powers of attention, the theory must either assert that the intelligible structure of the world is already contained within the perceived sense data, in which case the role of attention is reduced to mere symbol manipulation (*ibid.* 32), or that the world itself is already structured this way, in which case it is unclear why attention should be drawn towards one object rather than another (*ibid.* 31). Considerations such as these have led Merleau-Ponty and other philosophers to seek an alternative account of the nature of perception and attention.

In contrast to the snapshot model of perception, what I will call the *direct access model* denies that there is any internal representation of visual scenes prior to their entering consciousness. According to Merleau-Ponty (*ibid.* 43) and cognitive scientist and philosopher Alva Noë (2004: 420), the act of perception is itself a form of selective attention towards a world in which the observer is essentially embedded. Rather than being represented within the brain and then discarded, the unattended aspects of a perceived scene (e.g. the periphery of the visual field) are not actually *seen* by the subject at all, but are rather *sensed* as a vague and indeterminate presence on the horizon of consciousness (Merleau-Ponty *op. cit.* 78). As I sit at my desk looking at these words on a computer screen, for example, I do not *see* the wall behind the desk or the lamp and books to my left any more than I *see* the part of the room that lies behind the back of my head. Rather, I *sense* their presence as *objects that I could bring into perceptual focus should I choose to do so*.[3] This illustrates a key aspect of Merleau-Ponty's account, which is that all experience is structured as a series of 'figures' against a 'background' (*ibid.* 15). The dynamic tensions and oppositions between the foreground and background objects of experience is what forms the basis for both perception (*ibid.* 4) and attention, which Merleau-Ponty describes as 'a passage from indistinctness to clarity' (*ibid.* 32). However, rather than being a distinct process or mental faculty, attention forms an integral part of our system of perception and consciousness as a whole.

[3] This corresponds to what Noë (2004: 416) terms 'presence as absence', and is a distinctive feature of the phenomenological account of perception.

Under the direct access model, then, the function of attention is not to direct the conscious mind towards aspects of an already perceived scene, as if viewed on some kind of 'internal screen' (O'Regan 1992 in Thompson, Noë and Pessoa 1999: 167), but *to direct the process of perception itself*; that is, to orient the various organs of the body and senses towards those aspects of the environment that are relevant to our current thoughts and actions. We only *see* what we attend (or intend) to, nothing more (Noë 2002b: 5). Our impression of the world as a stable and persistent whole arises not from the integration or analysis of various sensory modalities as if this were something that occurred after the fact of seeing, hearing, and so on, but from our ability to gain direct sensory access to the world. Thus, it is not the case that I *see* the lamp, books on the desk, etc., and then discard these perceptions while concentrating upon something else. Rather, the mere possibility that I *could* direct my sensory faculties towards these objects is sufficient to give me a sense of their continued presence, even if they no longer form part of my visual field (as defined as the 'external horizon' of perceptual awareness (*ibid.* 78)).[4] Whether one calls this kind of awareness 'perception' or not is largely a matter of convention, but there is a clear contrast between this and the snapshot model in terms of what occurs at the perceptual level when we fail to attend to objects that are right in front of us.

Since perception and attention are already directly connected to (and indeed part of) the world, which functions as a kind of ultimate repository of perceptual information and awareness (O'Regan *op. cit.*), the direct access model does not require any kind of internal representation. This has the advantage that only objects within a subject's immediate field of interest need be represented by them, and only at a relatively high level of abstraction for the purposes of forming judgements, thoughts, and so on. However, since this view is perhaps less well grounded in pre-philosophical intuition than the familiar snapshot account, a more detailed reflection upon the phenomenological structure of perception and attention will be necessary in order to motivate and clarify it further.

[4] What Noë (2004: 422) refers to as 'presence as access'.

III. The Phenomenology of Attention

Imagine going for a walk beside a mountain stream on a hot summer's day. Looking around, you see water and trees below a clear blue sky, with birds circling overhead. The stream makes a pleasant gurgling sound as it trickles across the rocks, and you can hear birdsong as you walk along, enjoying the feeling of the warm sun on your back. Such a description might conjure up (or might *seem* to conjure up) something like a picture one might commonly see in a Rambler's magazine; i.e. a more or less photographic image of what you would see if you were actually there. This is the kind of image that would be captured by a camera, and is a faithful representation of what we know to be there, but to what extent does it represent how we actually *see* such a scene in practice? In reality, we do not apprehend such scenes in a single glance, but allow our eyes, ears and other senses to take it in piece by piece, much as you might have imaginatively reconstructed the scene described above as you read through it. For example, we might first notice the movement of the water, how it flows over the rocks, and its relation to the gurgling sound that we hear. Then we might notice the contrasting forms of the mountains, sky and rocks as our eyes saccade back and forth, taking in each detail. We might recognise the shape, colour and texture of the undulating masses of leaves on the trees – objects that we *know* to be there, but do not actually *see* until we examine them directly. Thus our experience of such a scene is comprised of a host of perceptual events spread out over a period of time. Far from taking in the scene in its entirety, as the snapshot model might suggest, perceptual experience has a distinctly temporal structure that is based on a series of figure-ground relations, resulting in what Merleau-Ponty (*op. cit.* 34) calls a 'perceptual field'.

On further reflection, we find that much of our sensory experience is fragmentary, indeterminate and incomplete (Noë 2002a: 191). By artificially fixing our gaze upon one spot, for example, we would be unable to pick out many of the surrounding features, which remain as vague and amorphous presences in the periphery of our vision. Although we might be able to guess their nature from the familiar context, in a more novel situation we would be at a loss to describe our surroundings in any detail, and could easily be mistaken. It is not until we turn our attention – and therefore our perception – towards these objects that we actually *see* what is there, and thus gain an overall sense of the scene before us (*ibid.* 10–11). However, we must be careful not to stretch the analogy too far. To say that we build up a *picture* of the scene in front of us would be to posit some form of internal representation over and above what is given in experience. Moreover, there is nothing in our experience to suggest that what we are seeing is some kind of image or representation within our own brains, as the rocks and trees appear to be *over there* rather than 'in the head' (Thompson, Noë and Pessoa *op. cit.* 187).

Similarly, the fact that, for the most part at least, we experience the world as a unified and integrated whole, and not as a series of fragmentary or incomplete perceptions, does not require us to represent the perceptual field to ourselves in order to perceive it. On the direct access model, the objects we see gain their sense of stability and persistence not from any internal picture, but from the characteristic ways in which their appearance changes in response to the movements of our eyes and body (*ibid.* 55; Noë 2004: 423), and from the knowledge that if our gaze were to return to them then they would still be there. In other words, the possibility of direct access to our surroundings via our bodily senses is sufficient to give us the sense of integration and embeddedness that we all take for granted, and to assure us that objects will not cease to exist when we turn away from them. No additional form of representation is necessary.

Another notable feature of perception is that we are not drawn as quickly, or as strongly, to every aspect of our environment. Rapidly changing or moving stimuli typically attract our attention more than static or slowly moving ones (Noë and O'Regan 2000); difference more than sameness; edges and textures more than flat surfaces; and so on (Thompson, Noë and Pessoa *op. cit.* 163–4). The characteristic 'grabbiness' (O'Regan, Myin & Noë 1991: 82) of objects also forms an important part of they way that our perceptual experience is structured. In *The Structure of Behaviour*, Merleau-Ponty (1983: 7) uses the example of a moving point of light in a darkened room to illustrate how an object may draw our attention to such an extent that it becomes almost impossible to ignore, and our behaviour in following it as 'appears as directed, as gifted with an intention and a meaning' (*ibid.*). This and similar cases illustrate the way in which our perceptual faculties are directed towards salient features of the environment by a set of instinctive or readily acquired motor skills and reflexes that keep us appraised of our immediate surroundings. Such principles are neither strict causal laws nor biologically predetermined. Rather, they can be acquired and shaped in light of the goals and experience of each individual subject. Professional sportsmen and women, for example, are trained to exclude all other factors and distractions – crowd noise, the weather, and so on – that are not directly relevant to their performance. Buddhist monks and nuns, on the other hand, are able to train their minds to become consciously aware of *all* perceptual phenomena, but without their attention becoming attached to or drawn in by any one thing, creating what could be described as a generalised non-specific state of awareness. Somewhere between these two extremes lies what is probably the normal state for most of us: a kind of restive flitting between one object of attention and another, allowing ourselves to 'latch onto' whatever most attracts our interest, whether it is directly relevant to our current activities or not. As a result, we are able to remain appraised of important changes in our immediate environment without having to attend to all of it all the time, instead relying upon our ability to notice change as and when it happens, and act upon it accordingly.

Experiments that involve the deliberate misdirection of a subject's attention, or extremely slow rates of change, demonstrate that when something escapes our attention (i.e. when we do not have occasion to notice it), we can remain completely unaware of surprisingly dramatics events, such as a gorilla walking across a basketball field (Simons and Chabris 1999) or a car mysteriously changing colour from red to green (Noë 2004: 420). Provided that our attention is being distracted by some ongoing task or event, or that the changes happen slowly enough, we simply fail to notice them. These effects are known to psychologists as *inattentional blindness* and *change blindness*, respectively, and strongly suggest that, contrary to the snapshot model, the brain does not represent or maintain a complete image of the visual field. If it did, then we would easily spot the difference between the changes in the 'external' world and our 'internal' representation of it (*ibid.*). In practice, however, we only take in and remember those aspects of the world to which we are currently attending, with everything else that remains unperceived also remaining outside of consciousness.

On the basis of the above evidence, the direct access model of perception is both compatible with the phenomenology of attention and capable of overcoming various problems associated with the snapshot theory; namely the need for internal representation, its inability to deal with indeterminate, ambiguous or contradictory data, and the fundamentally active nature of perceptual attention. However, the question of which model is correct is also partly empirical, and so we must also take into account the evidence of the physical sciences. As Merleau-Ponty (2002: 108–9) argues, this is problematic in that a purely objective description of the workings of the physical body and mental processes may be insufficient to explain the nature of subjective (or inter-subjective) phenomena like perception and consciousness. By omitting the very thing that it attempts to describe (i.e. subjectivity itself), and using concepts that are themselves derived from subjective experience, physical science may simply be unable to give a full or accurate account of the 'body-subject' (*ibid.* 105), or the nature of first-person experience. Nevertheless, the empirical evidence may still help to rule out certain hypotheses on the basis of their incompatibility with current scientific knowledge, and so the next question that I shall consider is whether it is, from a scientific standpoint, plausible to deny the existence of attention as a distinct cognitive process.

IV. The Neurological Evidence

In his book, *How Brains Think*, William Calvin (1988) proposes the existence of so-called 'Darwinian processes' (*ibid.* 136) within the physical brain by which thoughts and perceptions compete with one other for control of our limited cognitive resources. He goes on to suggest a plausible physical mechanism for these processes, involving the establishment of synchronised patterns of firing between neighbouring regions of the brain, with the winners of this internal power struggle going on to form part of our conscious mental state (*ibid.* 146). Edelman and Tononi (2000) arrive at a similar conclusion with their 'dynamic core' hypothesis, which correlates the contents of the conscious mind with a highly selective and constantly changing region of the subject's brain. However, rather than simply equating consciousness with physical brain processes, they describe the central nervous system as entering into concert with the subject's body and environment in order to elicit characteristic patterns of behaviour and thought (*ibid.* 50). This principle also extends to memory, which they describe as non-representational in that the act of remembering also modifies the structure of the subject's brain in a way that more closely resembles the practice of a skill or ability than a purely computational act of information recall (*ibid.* 99). Accordingly, 'every act of perception is, to some degree, an act of creation, and every act of memory is, to some degree, an act of imagination' (*ibid.* 101) – a sentiment that is highly reminiscent of Merleau-Ponty's (2002: 26) view that memory involves the 'reliving' of experience.

Significantly, neither of the above theories requires anything resembling the distinct faculty or power of attention that Dretske and Peacocke's account requires. Rather, attention is thought of as a characteristic of the process by which thoughts or perceptions gain prominence over one another, either by means of some kind of internal voting mechanism, as in the case of Calvin's Darwinian processes, or by entering into the dynamic structure of consciousness, as per Edelman and Tononi. According to these theories, consciousness is inherently attentional in nature. By engaging in a constantly shifting series of interactions with its environment, the conscious subject selects which aspects of the world are experienced and brought into conscious awareness, thus allowing it to shape and direct its future thoughts and actions. Such actions guide and refine the progression of consciousness, either by predisposing the subject to seek out further perceptual experiences that are appropriate to its current goals and stimuli, or by bringing about thoughts and actions that are directed towards achieving these goals. It is this process of selection and direction towards autonomously created goals and behaviours that corresponds to what we normally call 'attention'. Under this account, attention is both partially *constitutive* and an essential characteristic of consciousness that arises from the manner in which the conscious mind evolves and adapts in response to its environment. If this view is correct, then attention and consciousness cannot be separated *because they are both aspects of a single integrated system*, and not two distinct faculties, as Dretske and Peacocke's account supposes.

These views closely match those of Merleau-Ponty, who states that '[t]he first perception of colours properly speaking, then, is *a change in the structure of consciousness*' (*ibid.* 35; my italics). In other words, to perceive (or pay attention to) something is to *bring it into consciousness*, thus generating new or altered structures of consciousness. These structures are what the neurological theories mentioned above are attempting to describe (notwithstanding the methodological difficulties previously noted). In contrast to the snapshot model's passive 'reading off' of information from previously acquired sense data, the direct access model characterises attention as a fundamentally active process that is centred upon the goals and nature of the embodied subject, and an integral part of the cycle of action and interaction that constitutes conscious awareness.[5] This is the meaning of Merleau-Ponty's claim that attention 'does not exist' (*ibid.* 34), which is supported by his account of the fundamentally integrated and systemic nature of sense perception and consciousness.

To illustrate the point by way of a thought experiment, try to imagine a being that possesses the ability for conscious reflection but without any of the attentional processes described above. Instead of being drawn to those features of the environment that capture its interest, such a being would be equally and simultaneously aware of *all* of the elements in its visual, auditory and other sensory fields. Its mental processes would lack the interplay of mental and perceptual objects that arises as a result of the figure-ground structure, and would instead comprise of a simultaneous progression of its entire mental state in a manner that is more akin to computation or symbol manipulation (albeit of a massively parallel kind) than thought as we know it (*ibid.* 17). Indeed, it is difficult to imagine how such a creature could be anything more than a passive mirror to its environment, as without the figure-ground

[5] A comparison may be drawn with Wittgenstein (1967: §608), who denies that psychological phenomena can necessarily be 'read off' the physical state of the brain or body.

structure that is so essential to sense experience, the concept of consciousness itself begins to breaks down. Although this does not in itself prove that such a radically different form of consciousness from our own could or does not exist, it does demonstrate the closeness of the relationship between consciousness and attention, at least as far as our own thinking is concerned.

V. Two Possible Counterexamples

Two potential counterexamples to the direct access model of perception and attention described above are (i) the physical structure of the visual cortex, and (ii) the phenomenon of photographic memory. The first of these objections is motivated by the existence of highly regular and organised neurological structures for detecting movement, lines and edges of various orientations throughout the lower rear portion of the brain (e.g. Garey 2001). Although the existence of such structures might be thought to provide evidence of the kind of 'representational surface' that the snapshot model requires, current empirical evidence fails to resolve the issue either way. At best, these regions form the first rung in a series of complex neural mechanisms that undoubtedly participate in visual perception, but it is unclear how or at what point such 'signal processing' turns into what could properly be called perceptual awareness. To simply assume that such structures function in the way that the snapshot model requires would be to beg the question against the direct access model, and so cannot be taken to resolve the issue without further evidence and understanding of the precise neuro- and physiological processes involved.

The phenomenon of photographic memory is, however, more problematic. In such cases, subjects are apparently able to 'read off' details of a previously perceived scene – a page of a book, for example – whilst experiencing the phenomenological characteristics of precisely the sort of 'internal screen' that the direct access model denies. Such evidence could be claimed to support the snapshot theorist's notion of internal representation, with attention as a process that is common to both normal and so-called photographic perception. Indeed, Merleau-Ponty (*op. cit.* 118) and contemporary researchers, such as Ramachandran and Blakeslee (1999), often emphasise the importance of similar pathological cases in providing evidence for the normal functioning and structure of the human mind. However, in the present case it is unclear whether such extraordinary feats of memory can be described as a form of perception at all, since the subject cannot be said to *see* the additional detail either when they are first exposed to the scene, or when they are later able to recount previously unnoticed aspects of it. Rather, photographic memory is, as the name suggests, an unusually vivid form of *recall* that acts alongside ordinary perception, but in which the normal order of events is reversed, with memory playing the role that is usually associated with the senses. On this account, the existence of photographic memory is not necessarily indicative of normal perceptual processes, as the snapshot theorist would wish to claim, although the mere existence of such detailed memories of past sensory experiences could itself provide support for the kind of internal representation that the snapshot model requires. However, further empirical evidence would be required to support this hypothesis, and since both theories are able to provide an account of the phenomenon, this cannot be taken as a knockdown argument in favour of the snapshot model.

VI. Conclusion

I have argued that rather than far from being a distinct faculty, or 'phase' of consciousness, attention is an integral part of all perceptual and cognitive processes and, as such, is partially constitutive of them. The snapshot model of perception advocated by contemporary philosophers, such as Dretske and Peacocke, influenced by causal and physical notions of perception, and a computational view of the mind, fails to account for the empirical phenomenon of change blindness, and is at odds with the phenomenological structure of attention as we experience it. Furthermore, by internalising the perceptible world in the form of an internal representation or 'screen', the snapshot model is unable to explain the indeterminate and contradictory qualities of perceptual experience, its fundamentally active nature, or its role in consciousness in general. Such difficulties are simply deferred to the inner level, where they recur one step removed from the phenomena to which they relate. Conversely, by conceiving sense perception as inherently attentional and directed towards particular aspects of a world within which the subject is essentially embedded, the direct access model that arises out of Merleau-Ponty and Alva Noë's phenomenological account is able to explain the links between perception, attention and consciousness as aspects of a single integrated system which, when acting as a whole, yields the behaviour and conscious experience that we associate with living, sentient beings. Recent scientific theories, such as those developed by Calvin, Edelman and Tononi, show that the direct access model is both compatible with objective physical descriptions of the body whilst remaining sympathetic towards the irreducibly phenomenological approach that Merleau-Ponty and Noë espouse, despite the difficulty of attempting to describe the subjective realm of experience in purely objective terms.

In summary, Merleau-Ponty's denial of the existence of attention as something that exists over and above the phenomenon of perceptual awareness may be seen as a consequence of his views about the nature of perception and consciousness in general, and the primacy of the figure-ground structure in human perceptual awareness in particular. These views directly contradict the mechanistic accounts offered by Dretske, Peacocke, and other theorists who subscribe to a causal account of perception and attention, whilst successfully accounting for many otherwise mysterious aspects of perception, as well as recent developments in the rapidly expanding fields of cognitive and neurological science. Although the empirical evidence is currently inconclusive on this point, the direct access model's consistency with scientific explanation and explanatory power makes it highly plausible that attention as a distinct faculty or process does not in fact exist, but is rather just one aspect of the highly integrated and systemic nature of perception, thought and conscious awareness.

Bibliography

Ayer, A. J. 1973: *The Central Questions of Philosophy*, chs. 4–5, pp. 68–111. London: Weidenfeld

Calvin, William H. 1998: *How Brains Think: Evolving Intelligence, Then & Now*. London: Phoenix

Dennett, Daniel C. 1991: *Consciousness Explained*. Boston: Little, Brown & Company

Dretske, Fred 1988: 'Sensation and Perception'. In *Perceptual Knowledge*, J. Dancy (ed.). Oxford: Oxford University Press

Dretske, Fred 2004: 'Perception Without Awareness'. Paper presented at the Mind and Language Seminar, New York University

Edelman, Gerald M. & Giulio Tononi 2000: *Consciousness: How Matter Becomes Imagination*. London: Penguin Books

Garey, Laurence 2001: 'Cerebral Cortex'. In *The Oxford Companion to the Body,* C. Blakemore and S. Jennett (eds.). Oxford: Oxford University Press

Merleau-Ponty, Maurice 1983: *The Structure of Behaviour*. Pittsburgh: Duquesne University Press

Merleau-Ponty, Maurice 2002: *Phenomenology of Perception*. Oxon: Routledge

Noë, Alva 2000: 'Perception, Attention and the Grand Illusion' (with J. Kevin O'Regan). *Psyche*, 6 (15).

Noë, Alva 2002a: 'Is Perspectival Self-Consciousness Non-Conceptual?'. *The Philosophical Quarterly*, 52 (207), pp. 185–94

Noë, Alva 2002b: 'Is the Visual World a Grand Illusion?'. *Journal of Consciousness Studies*, 9, no. 5–6, pp. 1–12

Noë, Alva 2004: 'Experience Without the Head'. In *Perceptual Experience*, T. S. Gendler & J. Hawthorne (eds.). Oxford: Oxford University Press

O'Regan, J. Kevin 1992: 'Solving the Real Mysteries of Visual Perception: the World as an Outside Memory'. *Canadian Journal of Psychology*, 46, pp. 461–98

O'Regan, J. Kevin, Erik Myin & Alva Noë 2001: 'Towards an Analytic Phenomenology: The Concepts of Bodiliness and Grabbiness'. In *Seeing and Thinking. Reflections on Kanizsa's Studies in Visual Cognition*, A. Carsetti (ed.), Kluwer (in press)

Peacocke, Christopher 1983: *Sense and Content: Experience, Thought, and Their Relations*, ch. 1. Oxford: Oxford University Press

Peacocke, Christopher 1998: 'Conscious Attitudes, Attention and Self-Knowledge'. In *Knowing Our Own Minds*, C. Wright, B. C. Smith and C. Macdonald (eds.). Oxford: Oxford University Press

Ramachandran, V. S. & and Sandra Blakeslee 1999: *Phantoms in the Brain: Human Nature and the Architecture of the Mind*. London: Fourth Estate

Simons, Daniel J. & Christopher F. Chabris 1999: 'Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events'. *Perception*, 28, pp. 1059–74

Thompson, Evan, Alva Noë & Luiz Pessoa 1999: 'Perceptual Completion: A Case Study in Phenomenology and Cognitive Science'. In *Naturalizing Phenomenology*. J. Petitot, J-M Roy, B. Pachoud, & F. J. Varela (eds.). Stanford: Stanford University Press

Wittgenstein, Ludwig 1967: *Zettel*. G. E. M. Anscombe and G. H. von Wright (trans.). Oxford: Blackwell

British Journal of Undergraduate Philosophy, 2(2), July 2007

Lotteries, moles, and a belief-based account of assertion

Alex Rubner
Oriel College, Oxford
alexander.rubner@oriel.ox.ac.uk

This essay is on the topic of the norm of assertion. It centres on what I think is the most important counterexample to current theories other than the belief-based accounts. Generally I wish to defend a belief-based account of assertion which makes the rule of assertion 'Assert only if you have a sufficiently high degree of rational confidence'. So how high is sufficiently high? The answer is that you can assert when your degree of rational confidence is high enough for it to be reasonable for you to form the belief that you know the content of your assertion. So your evidence and your attitude to that evidence have to be strong enough that you can reasonably believe that you know, but it does not require that you actually hold that belief. Such a substantial issue, however, is far beyond the remit of this essay. The situation I will discuss is simply an important motivator for this general type of account, though it does not fully explain why the threshold for 'sufficiently high degree of rational confidence' is what I say it is, and I shall not be focussing on that aspect of the account (I only mention it here for the sake of completeness).

First I shall explain what precisely the issue is. The question is what makes an assertion proper. The answer is that a proper assertion is one that is warranted, so the issue is what warrants assertion. Warrant, as Williamson says, is a term of art and should not be taken to be synonymous with justification. A justified assertion might merely be one for which there is a reason, or one which has some measure of justification, but there can be assertions which are justified and not warranted. The best way to look at warrant is to identify it with criticism: it is a good rule of thumb that if your assertion is subject to criticism then it is not warranted, but if there is no relevant sense in which it can be criticised then it is warranted. I should point out here that we are talking about your act of assertion, and not just the content of your assertion. My account allows that the content of your assertion can be subject to criticism (if it is false, for example), but you can still be warranted in making the assertion itself.

So what sort of thing can warrant assertions? Things that might immediately spring to mind are things such as truth, reasons, belief or knowledge, and accounts based on all of these have been suggested. The prevailing view is that only knowledge can warrant assertion – you can assert only if you know what you are asserting. (Though you don't have to know that you know.) This view has the consequence that most of what we assert, we assert improperly; but that is not to say that we fail to assert. If I score a try while offside I am still playing rugby, just not properly; a poker player who has stacked the deck is cheating, but still playing poker. There are a number of motivators for the knowledge account, and I shall discuss the lottery case in a moment, but Williamson, the main proponent of the knowledge account, also looks to our everyday practices for confirmation. When you assert something, a common response might be 'How do you know that?' This implies that we expect people to know what they have asserted.

After explaining the lottery case, I shall go on to describe a case of my own which the knowledge account cannot deal with, but first a word on everyday practices. If 'How do you know?' is a common response, then surely as common, if not even more common, would be responses like 'Are you sure?', 'Do you really think so?', or just 'Really?'. The first two would lend confirmation to belief-based accounts, and the last a truth-based account. ('Why do you say that?' could be used to confirm a justification-based account.) But since every type of theory can look to challenges like these to lend support for their theory, we must not put too much stock in our everyday practices.

Something like the lottery case provides much more tangible evidence, and it is used by Williamson to argue against accounts based on truth. It runs as follows: Alice buys a lottery ticket and her friend Lola, after the draw has been held, but before the results have been announced, tells her that she has not won. Since Lola has no information about the result, her assertion 'Your ticket didn't win' is unwarranted. When Alice finds out that Lola has no inside information, and is basing her statement only on the high probability that Alice's ticket is not the winning one, Alice is liable to feel resentment towards Lola. That is to say, the assertion will be subject to criticism, and is therefore improper. (Williamson does note that there is a tone in which the assertion might not be criticised, like saying '(Come off it!) Your ticket didn't win'.) Bear in mind that it doesn't matter what the probability is, as you can make the lottery as big as you like and the assertion would still be unwarranted. Also, we should imagine that the lottery is more like a tombola than the National Lottery, so there is definitely a winning ticket.

The knowledge account, on the safe assumption that probability alone cannot yield knowledge, can easily explain what is wrong with Lola's assertion: if knowledge is what warrants assertion, probability alone cannot give one warrant to assert (unless the probability is 1 or 0).

The following example challenges this (that is, it doesn't challenge the lottery case, just the conclusion that probability alone can never warrant assertion). Imagine an elite military unit, whose mission is to keep slipping behind enemy lines in order to gather intelligence. The enemy soldiers patrol their borders randomly, so that for every time the unit crosses the border there is only a 1 in 50 chance that there will be a patrol. The last six times the unit attempted a crossing they were met by enemy patrols and barely escaped with their lives. The commander knows that, unless the enemy is being fed information, there is only a 1 in 50 chance of being caught on any one mission, and knows that only the unit has such information, and so concludes 'There is a mole in our midst' (or 'One of my men is a mole').

Pre-theoretically, it looks like this assertion is warranted – it is very difficult to see how the commander might be criticised for saying this, so it seems that he has asserted properly. In fact, he has almost been forced to make this assertion. If he doesn't make the assertion and sends his men back into enemy territory, he will be criticised for making such an obviously stupid decision. So are we to conclude that he is in a Morton's Fork situation, that he is damned if he does and damned if he doesn't? Surely it is easier (and more plausible) to suppose that the assertion is indeed warranted. But the knowledge account can't deal with this assertion's being proper, because he certainly doesn't know that there is a mole.

One might think that DeRose's distinction between primary and secondary propriety could be employed here: someone following the rule of assertion asserts properly in a primary sense, but someone who reasonably believes that he is following the rule of assertion asserts properly only in a secondary sense. This distinction can often be used to explain why it seems like someone has asserted properly even though he breaks the rule of assertion. But this approach cannot help the knowledge theorist because the commander's assertion is not secondarily proper on the knowledge account – when the commander asserts that there is a mole he does not believe that he knows this.

One possible line of response for the knowledge account proponent to take is to explain the permissibility of the assertion by appeal to other norms. The idea here is that while the norm of assertion might be 'Assert only if you know', in this situation another rule supersedes the primary norm to make the assertion permissible – something about risk or what's at stake. It still isn't a proper assertion, for it has broken the primary rule, but this does explain why we regard the assertion as permissible, and aren't willing to criticise the commander for making it – he has asserted according to the rules in some secondary sense. So let's say that there is a special norm for asserting when you are a commander of a unit whose mission it is to go behind enemy lines to gather intelligence when there is a 1 in 50 chance of being caught, and this norm outweighs the ordinary norm of assertion. Are we then to assume that there is also a special norm for the commander of a unit whose mission it is to go behind enemy lines to gather intelligence when there is a 1 in 49 chance of being caught? And another for the guy who has a 1 in 48 chance? And so on. There is nothing to conclusively refute this approach to salvaging the knowledge account; it merely has undesirable consequences. If we have one special norm we must admit others with little or no principled basis for doing so, and very soon the term assertion would only be properly applied to a minute class of expressions.

My belief-based account deals with this case without the baggage – the assertion is warranted because the commander has a sufficiently high degree of rational confidence. Simple as that. If you think that perhaps his belief that there is a mole is irrational, think how much more irrational it would be to believe that there is not a mole, or to suspend judgement completely.

But if both Lola and the commander have made assertions based entirely on probability, why does one have warrant and the other not? The answer is that Lola has a problem that the commander does not have – the reason that Lola said that Alice's ticket did not win can be extended to other tickets. If Lola can say that ticket #5 didn't win, she can say that ticket #5000 didn't win – her reasons would be the same; namely the overwhelming unlikelihood of the ticket's having won. But she knows that if she asserts of every ticket that it did not win, she will definitely have asserted a falsehood, since there is a winning ticket. In the elite unit case, the commander can say of each man that he is not a mole without knowing that he has asserted falsely. Granted he might believe that one of the assertions was false, but he would not know this (or claim to know it), for it is possible, though unlikely in the extreme, that there is no mole.

Nonetheless, the knowledge account proponent can still maintain that no assertion here is proper, so we must delve a little deeper and take a look at the men. The commander has served with these men for years, he has access to all of their files and they all have exemplary records – basically, he trusts them all absolutely. So he is very willing to assert 'Andy is not a mole', 'Brian is not a mole', 'Charlie is not a mole', and so on. But he is also willing to assert that there is a mole. Since warrant is based on belief, how can he consciously believe that not one of them is a mole, and also that one of them is a mole?

The answer is that he doesn't hold both of these beliefs. He holds the belief that there is a mole, and he holds individual beliefs about Andy, Brian, Charlie, and so on, but he doesn't hold the belief that none of the men is a mole. But is this possible? Can he believe that Andy is not a mole, and believe that Brian is not a mole, and so on, but not believe the conjunction of all of these beliefs? Basically, is warrant to assert (and therefore sufficiently high degree of belief) closed under conjunction? Can you have warrant to assert A, and to assert B, but not have warrant to assert C, where C is entailed by the conjunction of A and B? It very much seems that you can.

Let's imagine that all the men are as good and as trusted as each other, though in different ways. So when the commander examines the evidence, he comes to the same conclusion about them all individually (though for different reasons in each case) – he is all but certain that each of them is not a mole. Let's say that he's 95% sure that Andy is not a mole, 95% that Brian is not a mole, 95% that Charlie is not a mole, and so on. Let's also assume, purely for argument's sake, that 95% is the threshold for assertion. So he has warrant to assert of each one man that he is not a mole, but he can't assert even of two men that neither is a mole. By the time he gets to Charlie he is 86% sure, and if there's a Dave he goes down to 81%. But do bear in mind that the numbers don't matter – I'm just illustrating my point less abstractly. He has warrant for each statement, but not their conjunction, so warrant is not closed.

Such considerations are why, on my account, you don't have to know or be totally sure to assert. If you did then warrant would be closed under conjunction, but the unit case suggests that it isn't. But I haven't quite proved it yet – the knowledge account theorist can still respond by denying that any of the assertions is warranted at all. Though this leads to far too many problems to discuss here, it still needs to be dealt with.

Go back to the original unit situation. They've almost been caught six times, so the commander believes that there is a mole. But let's say that there is no mole, and the six times were indeed pure coincidence. So when he examines the men and their backgrounds, he has warrant to assert of each of them that he is not a mole on any account. Specifically looking at the knowledge account, he actually does know that each of them is not a mole. Thus if warrant is closed, he should have warrant to assert that none of them is a mole. But if having warrant means knowing, he doesn't have warrant to assert that none of them is a mole

– he doesn't believe this, so he can't know it. Thus warrant is not closed under conjunction, and therefore the rule of assertion cannot be based on knowledge.

The focus has been on the case of the commander and his unit. First I used it to suggest that the threshold for warranted assertion is lower than knowledge, and then I moved on to a belief-based account. Williamson argues that probability alone cannot warrant assertion because it cannot yield knowledge, and bases this on lottery cases; I have based my argument, that probability alone can warrant assertion even if it can't yield knowledge, on similar cases, and explained in what ways they differ. The point has been to show that knowledge is not the norm of assertion – one can make warranted assertions without having knowledge. This is why the last paragraphs are the most important: the final variant of the commander-unit case shows that warrant to assert, unlike knowledge, is not closed under conjunction.

Bibliography

Kvanvig, Jonathan. (forthcoming) "Assertion, Knowledge, and Lotteries," in Patrick Greenough and Duncan Pritchard (eds.), *Williamson on Knowledge*. (Oxford: Oxford University Press)

Lackey, Jennifer. (1999) 'Testimonial knowledge and transmission.' *Philosophical Quarterly*, 49: 471–90

Lackey, Jennifer. (forthcoming) "Norms of assertion", available online

Weiner, Matthew. (2005) "Must We Know What We Say?" *The Philosophical Review* 114: 227-51

Williamson, Timothy. (1996) "Knowing and Asserting." *The Philosophical Review* 105: 489-523

Williamson, Timothy. (2000) *Knowledge and its Limits*. (Oxford: Oxford University Press)

British Journal of Undergraduate Philosophy, 2(2), July 2007

Is 'ontological security' possible?

Jessica Woolley
University of East Anglia
J.Woolley@uea.ac.uk

In *The Divided Self*, R.D. Laing coins the phrase 'ontological security' to refer to 'a centrally firm sense of [one's] own and other people's reality and identity'[1] which arises from the experience of one's 'presence in the world as a real, alive, whole, and… [temporally] continuous person.'[2] Laing's purpose in defining ontological security is to differentiate the normal, everyday being of 'the-man-in-the-street'[3] from the 'ontologically insecure' being of schizophrenia – a condition which he suggests is characterised by a desperate and alienating 'struggle to maintain a sense of […] being'[4] which is insufficiently supported, due to 'a precariously established personal unity.'[5] Thus, while ontological insecurity is explored by Laing as personally undermining and a source of suffering and psychosis, ontological security is presented as something to be desired and sought after, as the necessary foundation for a tranquil and fulfilled life.

But is ontological security possible?

This question can be apprehended as asking 1. whether Laing's formulation of ontological security is coherent, 2. whether ontological security as a concept is conceivable, and 3. whether it is achievable. I will attempt to address these concerns in order.

[1] Laing, R.D. *The Divided Self.* Harmondsworth. Penguin Books Ltd. 1973. p.39.

[2] Ibid.

[3] Laing, R.D. *Self and Others.* London. Tavistock Publications. 1969. p.35.

[4] Mullan, Bob. *Mad to be Normal: Conversations with R.D. Laing.* London. Free Association Books. 1995. p.6.

[5] Laing, R.D. *Self and Others.* p.36.

Is Laing's formulation of ontological security coherent?

In Laing's writing, a confusion seems to arise between *actual* and *experienced* ontology, as ontological security, despite being repeatedly defined as 'a *sense* of being'[6] – essentially an 'experience' or 'feeling' of one's existence – is also referred to as a 'basic existential position'[7] in which a person '*has* a firm core of ontological security.'[8] It is thus unclear whether Laing is positing ontological security and insecurity as actual states of being which cause people to experience their existence as respectively secure or insecure, or whether he is proposing ontological security and insecurity as modes of experiencing one's being, independent of its actual status.

A way of accounting for this confusion of actual being with experienced being is to suppose that for Laing the distinction either does not exist, or does not matter; that from the perspective of his enquiry, experienced being and actual being are for all intents and purposes indistinguishable. This account can be explored further through an analysis of Laing's method.

Laing presents his method as following the existential-phenomenological tradition in its attempt to describe, via the development of an internal understanding,[9] the transition from 'the sane… way of being-in-the-world to a psychotic way of being-in-the-world'[10]. Although Laing stresses that his method is 'not a direct application of any established existential philosophy,'[11] his study is nevertheless founded upon some important existential and phenomenological tenets. Thus it is not exempt from criticism regarding the adaptation and use of these philosophical elements.

[6] Ibid. p.4.

[7] Ibid. p.39.

[8] Ibid. p.42. My emphasis.

[9] An understanding which involves relating the actions of the patient 'to *his* way of experiencing' his existence. Laing, R.D. *The Divided Self.* p.34.

[10] Ibid. p.17.

[11] Ibid. *Preface to the Original Edition.*

Methodologically, Laing draws on the phenomenological tradition in several ways:

1) He rejects conventional theory and preconceptions regarding the phenomena of his investigation.
2) He 'focuses on the structure and qualities of objects and situations as they are experienced *by the subject*.'[12]
3) He treats the being of man as the origin of objectivity in terms of the relation between self, world and other.
4) He attempts, with his 'science of persons,'[13] to provide a description which 'fulfils rather than dehumanises the human world.'[14]

One crucial aspect of phenomenology is that it does not endorse the assumed distinction between appearance and being – and this would seem to justify Laing's conflation of experienced ontology with actual ontology, as 'phenomenology neither wishes to claim that all that exists can be simply reduced to appearings, nor to affirm an unknown and unknowable reality behind appearances.'[15] However, it seems to me that Laing's conflation of experience and actuality in the context of personal ontology is not exactly analogous or reducible to the phenomenological dissolution of the distinction between appearance and reality. Rather it is a specific symptom of Laing's adaptation of the phenomenological method, in that it arises from a particular failure to appropriately distinguish between two modes of observation: the view of self, and the view of other.

Phenomenology is essentially characterised by 'the attempt to get to the truth of matters, to describe *phenomena*, in the broadest sense as whatever appears in the manner in which it appears, that is as it manifests itself to consciousness, to the experiencer.'[16] Thus the phenomenological method derives its authority and authenticity from

[12] Moran, Dermot. *The Phenomenology Reader.* Edited by Dermot Moran and Timothy Mooney. London. Routledge. 2002. p.2.

[13] Laing, R.D. *The Divided Self.* p.34.

[14] Moran, Dermot. *The Phenomenology Reader.* p.3.

[15] Ibid. p.5.

[16] Moran, Dermot. *Introduction to Phenomenology.* London. Routledge. 2003. p.4.

the fact that its descriptions originate from the adherence of a *single* consciousness to the accurate disclosure of its experience as subject. From this methodological standpoint, the enquirer exists as the locus and originator of significance, as all that is described is described in terms of its relation and appearance to the investigating self. In the attitude of enquiry the self gives rise to a distinction between the one who enquires and that which is enquired into.[17] Thus the enquiring self is necessarily the only *subject* of the investigation, as that which is other to the self is apprehended by the self as an *object* for its enquiry.[18]

Laing describes his investigation as a 'study of human beings that begins from a relationship with the other as person and proceeds to an account of the other still as person.'[19] The Other here is conceived of as 'responsible, as capable of choice, in short, as a self-acting agent.'[20] In his account, Laing criticises the psychiatric reduction of the patient to a 'fictional 'thing''[21] and advocates the restoration of the patient's subjectivity by engaging in 'an attempt to reconstruct the patient's way of being himself in his world.'[22]

What Laing seems unaware of here is that in describing the patient – i.e. the Other – as subject, he effectively destabilises the structure of his investigation by bringing in a rival and incompatible locus of significance. Laing may be right in asserting that 'in existential phenomenology the existence in question may be one's own or that of the other.'[23] However, he misses the crucial point that in existential phenomenology, the existence of the Other is only in question in-so-far as it is an existence *for the self* – as the Other is apprehended by me, not

[17] 'Every question presupposes a being who questions and a being which is questioned.' Sartre, Jean-paul. *Being and Nothingness*. Translated by Hazel E. Barnes. London. Routledge. 2003. p.28.

[18] Not that this apprehension of the Other as object does not equate to a reduction of the Other to an in-itself 'thing', but rather establishes it as an objectivity to be transcended by the self's subjectivity.

[19] Laing, R.D. *The Divided Self.* Op.Cit. p.2.

[20] Ibid. p.22.

[21] Ibid. p.24.

[22] Ibid. p.25.

[23] Ibid.

as a being-for-itself, but as a being for which *I create* a being,[24] both in my being-for-Others and in the Other's being-for-me. I can describe the world of my experience as I experience it, and can describe others in terms of my apprehension of them in the context of my world. I can also assume, from my encounters with Others' descriptions of their existential experience, that they each have their respective worlds of experience in which they experience me in terms of their apprehension of me as Other. What I am not in a position to do though, is to describe *from their perspective* the world of their experience *as they experience it.*[25]

Thus in describing ontological security in terms of the experience of others-*as-selves* Laing not only removes himself as the validating locus of his enquiry, but transgresses the necessary relation that makes such an enquiry possible; when he attempts simultaneously to affirm himself as self and to apprehend the Other as self he places himself in the untenable position of describing from a singular point of observation two incompatible 'worlds' of experience. Laing absorbs the posited experience of the Other who is being investigated into that of the questioning self, and replaces himself as the subject of his enquiry with a being who is neither self nor Other, but a mixture of self speaking in the guise of the Other, and the Other presented as Other-self, speaking through the words of Laing. This results in a description without a coherent source, an investigation without a clear object, and an enquiry without a foundation or basis for its authenticity.

So, it seems that Laing's notion of 'ontological insecurity' cannot be properly made sense of in the context in which he presents it – i.e. as a phenomenologically described sense or state of being experienced by others; and the same obviously goes for 'ontological security'. However, this does not rule ontological security out as a possibility, but only

[24] 'objectivity is not the pure refraction of the Other across my consciousness; it comes through me to the Other as a real qualification: I make the Other be in the midst of the world.' Sartre, Jean-Paul. *Being and Nothingness.* p.316.

[25] 'The difference of principle between the Other-as-object and the Other-as-subject stems solely from this fact: that the Other-as-subject can in no way be known nor ever conceived as such.' Sartre, Jean-Paul. *Being and Nothingness*. p.317.

discounts it in terms of the methodological context in which Laing unveils it.

Is ontological security conceivable?

Laing describes ontological security as arising from the experience of one's being 'as a continuum in time; as having an inner consistency, substantiality, genuineness, and worth; as spatially coextensive with the body; and, usually, as having begun in or around birth and liable to extinction with death.'[26] Thus to be ontologically secure is to be secure in oneself[27] – to apprehend oneself as having a certain kind of essence, and to derive a sense of stability from this ontological self-identification.

But can such an experience of essential self-identification really be a source of ontological security?

When once asked in an interview about his sense of identity, Derrida proclaimed: 'Identification is a difference to oneself, a difference from-with oneself. Therefore *with*, *without* and *except with* oneself. The circle which brings one back to birth can only remain open, but all at once as an opportunity, a sign of life, and a wound. It would be death if it closed onto birth, onto a fulfilment of the statement, or of the knowledge which says 'I am born'.'[28]

To me this serves to show how, rather than providing a sense of personal unification, substantiality and identity, the attempt to be secure *in* oneself necessarily estranges one from one's being. This occurs in two interrelated ways: firstly, in the very act of self-identification, I fragment myself by distinguishing between that which identifies and that which is identified with; and secondly, in seeking to apprehend myself as a substantial and unified whole *in-itself*[29] (to use Satrean

[26] Laing, R.D. *The Divided Self.* p.41-42.

[27] Laing, R.D. *The Divided Self.* p.42.

[28] Derrida, Jaques. *A Certain 'Madness' Must Watch Over Thinking* (Interview by François Ewald). Educational Theory Vol.45, No. 3. 1995. p.273.

[29] 'Being-in-itself (*être-en-soi*). Non-conscious Being. It is the being of the phenomenon and overflows the knowledge we have of it. It is a plenitude, and strictly speaking we

terminology), I deny my status as a being *for-itself*.[30] Derrida suggests that it is the very estrangement and open-endedness of existence which allows one to live – without it, one would cease to exist as a being which freely projects itself towards possibilities, and would become instead a static, dead 'thing'.

This is something which Sartre explores in his description of the self's relationship to its past, writing that 'the past… is that which has consumed its possibilities. …In other words… it is an in-itself like the things in the world.'[31] Essence 'is all that human reality apprehends itself as *having been*', and 'it is here that anguish appears as an apprehension of self inasmuch as it exists in the perpetual mode of detachment from what it is.'[32] 'I can not enter the past… because the past *is*', and the 'only way by which I could be it is for me myself to become in-itself in order to lose myself in it in the form of identification' – something which 'by definition is denied me,'[33] 'because I am for-myself.'[34].

The only *authentic* state in which one's being could be identified as ontologically secure (i.e. unified, substantial, consistent and whole) is death, as 'by death the for-itself is changed forever into an in-itself in that it has slipped entirely into the past.' [35] On the other hand, my ability to live from moment to moment relies precisely upon the fact that *I am not what I am*[36] – i.e. it rests on the current of ontological

can say of it only that it is.' From Hazel Barnes' 'Key to Special Terminology' in Sartre, Jean-Paul. *Being and Nothingness*. p.650.

[30] 'Being-for-itself (*être-pour-soi*). The nihilation of Being-in-itself; consciousness conceived as a lack of Being, a desire for Being, a relation to Being. By bringing Nothingness into the world the For-itself can stand out from Being and judge other beings by knowing what it is not. Each For-itself is the nihilation of a particular being.' Ibid.

[31] Sartre, Jean-Paul. *Being and Nothingness*. p.139.

[32] Ibid. p.59.

[33] Ibid. p.142.

[34] Ibid.

[35] Ibid. p.138.

[36] 'whatever I can be said to be in the sense of being-in-itself with a full, compact density… is always *my past*. It is in the past that I am what I am. But on the other hand, that heavy plenitude of being is behind me; there is an absolute distance which cuts it

insecurity which carries me beyond my in-itself-ness and allows me to freely experience the world through the immediate transcendence of my 'factic'[37] being. Thus to attempt to be ontologically secure is to attempt to eliminate the conditions of one's existence as a conscious being. This is illustrated Sartre's account, with reference to Husserl and Heidegger, of the phenomenological idea of 'intentionality':

> To be is to fly out into the world, to spring from the nothingness of the world and of consciousness in order suddenly to burst out as consciousness-in-the-world. When consciousness tries to regroup itself, to coincide with itself once and for all, closeted off all warm and cosy, it destroys itself.[38]

In light of this it seems that ontological security as a personal experience or sense of one's being, far from being an authentic state of existential stability, must be conceived of as a form of bad faith,[39] as the self-affectation of one's being with a false attitude of solidity in the attempt to '*realize* value and flee the anguish which comes to it from the perpetual absence of the self.'[40] In the attitude of ontological security, one seeks to affirm oneself as a being which is grounded, fulfilled, and meaningful (i.e. in-itself) and yet simultaneously assert one's personal freedom, independence and potential for growth. This exemplifies Sartre's description of the art of bad faith, which lies in 'forming contradictory concepts which unite in themselves both an idea and the negation of that idea.'[41] In constituting oneself as what one *is*, one

from me and makes it fall out of my reach, without contact, without connections.' Sartre, Jean-Paul. *Being and Nothingness.* p.141.

[37] This is my term.

[38] Sartre, Jean-Paul. 'Intentionality: A Fundamental Idea of Husserl's Phenomenology', *in* Moran, Dermot. *The Phenomenology Reader.* p.383.

[39] 'bad faith is a lie to oneself' in which one 'is hiding a displeasing truth or presenting as truth a pleasing untruth.' Sartre, Jean-Paul. *Being and Nothingness.* p.71-72. Bad faith 'utilizes the double property of the human being, who is at once a *facticity* and a *transcendence*. These two aspects of human reality are and ought to be capable of a valid coordination. But bad faith does not wish either to coordinate them nor surmount them in a synthesis. Bad faith seeks to affirm their identity while preserving their differences.' Sartre, Jean-Paul. *Being and Nothingness.* p.79.

[40] Sartre, Jean-Paul. *Being and Nothingness.* p.143.

[41] Ibid. p.79.

escapes the responsibility of being an entity which is perpetually beyond itself, via a retreat into the objective structure of one's being – a structure which belongs to the being of the past and the in-itself.

Is ontological security achievable?

Despite having (and perhaps because of having) posited ontological security as a conduit of bad faith, I do believe that some form of ontological security *is* achievable. However, what I mean by ontological security here is not anything to do with actually being secure in one's being, but rather a sense of being at peace with the conditions of one's existence as one authentically experiences it – what might more appropriately be termed ontological quietude. To me, the way to best be at peace regarding one's existence is not to rely on 'self-affirming' values and facts regarding time, substantiality, identity etc., seeking to ground oneself in the manner of an object; but rather it is to try to accept as far as one can the responsibility of the freedom of one's being in all its aspects and apprehensions, until one might reach a point at which one is no longer troubled by the seeming insecurities, uncertainties, and tensions of one's existence, and can proceed with a sense of openness, sensitivity and practicality concerning one's possibilities and experience.

Bibliography

Derrida, Jacques. *A Certain 'Madness' Must Watch Over Thinking* (Interview by François Ewald). Educational Theory. Vol.45. No. 3. 1995

Laing, R.D. *The Divided Self.* Harmondsworth. Penguin Books Ltd. 1973

Laing, R.D. *Self and Others.* London. Tavistock Publications. 1969

Moran, Dermot. *The Phenomenology Reader.* Edited by Dermot Moran and Timothy Mooney. London. Routledge. 2002

Moran, Dermot. *Introduction to Phenomenology.* London. Routledge. 2003

Mullan, Bob. *Mad to be Normal: Conversations with R.D. Laing.* London. Free Association Books. 1995

Sartre, Jean-Paul. *Being and Nothingness.* Translated by Hazel E. Barnes. London. Routledge. 2003

British Journal of Undergraduate Philosophy, 2(2), July 2007

Plato's beard is not generally misdirected

Mirja Holst
University of Hamburg (visiting student, University of Sheffield)
mianho@gmx.de

There is something which Quine, in his paper 'On What There Is', calls the 'old Platonic riddle of nonbeing'. This puzzle led philosophers to countenance objects that intuitively do not exist – for example, unactualized possible horses. Quine argues that there is no need to accept them, because the puzzle rests on an assumption which is not only false but generally misdirected, namely 'Plato's Beard'. First I will introduce the puzzle and trace Quine's argument against Plato's Beard. Afterwards I will consider his further argumentation for the conclusion that Plato's Beard is generally misdirected and finally I will argue against it – for all Quine says, Plato's beard is not generally misdirected.

I. Plato's Beard

'Plato's Beard' is the named given by Quine to a well known paradox of reference. Quine describes an ontological dispute and the subsequent predicament of someone who wants to deny that there are certain objects – it appears that such a person cannot describe what is going on without admitting the objects they wants to deny. Quine formulates this 'old Platonic riddle of nonbeing' as follows:[1]

> [*Q*uine *1*]
> Nonbeing must in some sense be, otherwise what is it that there is not?

Q1 might be understood like this: an object which has in one sense of 'being' no being, must in at least one other sense of 'being' have being; and this is because we cannot say of an object that it has no being if there is no sense of 'being' at all in which it has being.

[1] Quine 1948: 1f.

The issue *seems* to arise when we *talk* about nonbeing: whenever we ascribe a quality, we ascribe it to an object; and we need to *refer* to it in order to ascribe the quality. If the object is not, we cannot refer to it, and thus we cannot ascribe nonbeing to it. Therefore we say something *senseless* when we try to deny the being of something which is not.[2] The assumption upon which the puzzle rests is:

> [*P*lato's *B*eard]
> We cannot meaningfully deny the existence of something which is not.

In *PB* I used 'existence' instead of 'being' because it seems to be a specifically philosophical custom to talk about *being*. Furthermore, I added 'which is not' because someone might perfectly well meaningfully deny the existence of something; namely by denying the existence of something which does exist. Although the assertion would be false, it would be meaningful.

What are the reasons to accept *PB*? Suppose someone wants to deny the existence of something. She may try and utter a sentence of the form '*e* does not exist'. Let *N* be a sentence of this form, where a singular term is inserted for '*e*'. Now we can formulate an argument:[3]

P1 If *N* is meaningful, the singular term inserted for '*e*' refers.
P2 If the singular term inserted for '*e*' refers, there is something to which it refers.
C If there is nothing to which it refers, *N* is meaningless.

C is more specific than *PB*, because it is about sentences of a certain form by which we deny the existence of something. One particular

[2] It *might* also be possible to hold that we would thereby say something *inconsistent* or *false*. Quine (1952: 220) argues that it is no good to take these statements as false, for if they were false, their negations would be true. But since 'Vulcan exists' is not true, 'Vulcan does not exist' cannot be false. Because Quine is concerned with the view that statements such as 'Vulcan does not exist' are 'nonsense', I shall talk about this view.

[3] The idea to formulate the premises like this came from Cartwright (1960: 630).

consequence of *PB* is that by uttering *N* it is only possible to say something false or meaningless, but it is not possible to express a truth.

II. Plato's Beard is false

A statement like 'The intra-mercurial planet does not exist' is intuitively *true*, which cannot be the case according to *PB*. Thus the argument does not seem to be sound. Quine attacks its first premise and asks: is it really the case that a singular term inserted for '*e*' has to refer in order for *N* to be meaningful? According to Quine, Russell's theory of definite descriptions[4] can be used to show that this is not the case.

How does it work? I take it as a common view that definite descriptions are singular terms –terms which purport to denote one and only one object.[5] In particular, they are singular terms of the form 'the *F*', such as 'the sister of Shakespeare'.[6] They purport to denote the unique object of which the predicate, represented by '*F*', is true. However, Russell analyses phrases of the form 'the *F*' systematically as fragments of sentences in which they occur. Here is an example:

S1 *The red planet* is big.

*S1** Something is a red planet and is big and nothing else is a red planet.

'the red planet' is not replaced by a unified expression. In the analysis there is thus no unique expression left which even purports to denote a unique object.

Quine holds that *S1** can be properly translated into a semi-formal language of classical logic: the English sentence *means the same as* the semi-formalised sentence. In particular he maintains that the existential

[4] Russell 1905.

[5] Although I guess that this is a common view, there *are* philosophers using 'singular term' in a different way and such that definite descriptions are not singular terms.

[6] Phrases which can, without loss or gain of meaning, be transformed (or translated) into phrases of the form 'the *F*', such as 'Shakespeare's sister' or 'Shakespeares Schwester', are also regarded as definite descriptions.

quantifier is synonymous with the corresponding phrases 'There is (an _x_ such that)' and 'There exists (an _x_ such that)'. This is the translation of _S1_*:

_S1_** ∃_x_ (_x_ is a red planet & □_y_ (_y_ is a red planet → _x_=_y_) & _x_ is big).

By the laws of the predicate-calculus it follows:

_S1_*** ∃_x_ (_x_ is a red planet).

Since _S1_*** means that there is something which is a red planet, we commit ourselves, by uttering _S1_, to there being something which is a red planet.[7] Now, consider a sentence of the form _N_, _S2_, which is analysed by _S2_**:

_S2_ The intra-mercurial planet does not exist.

_S2_* There is no intra-mercurial planet or there is more than one intra-mercurial planet.

_S2_** ~∃_x_ (_x_ is an intra-mercurial planet & □_y_ (_y_ is an intra-mercurial planet → _x_=_y_)).

_S2_**is a _negative_ existential sentence and does _not_ entail that there is an intra-mercurial planet. Since we do not refer to an intra-mercurial planet by uttering _S2_, and since _S2_ is meaningful, we _can_ meaningfully deny the existence of something which is not. Therefore, _PB_ is false.

Quine's argument rejects _PB_ and I think it is all that is really needed to reject _PB_. But Quine goes on to argue that we _generally_ utter a meaningful sentence by uttering a sentence of the form '_e_ does not exist'[8] – this is not only the case if we fill in definite descriptions for '_e_', but also if we substitute other terms, such as proper names.

[7] I will not go into Quine's reasons for that.

[8] At least if we exclude meaning_less_ singular terms.

III. Plato's Beard is generally misdirected

How can we say something meaningful by uttering the sentence 'Vulcan does not exist'? Quine claims that Russell's analysis can be applied to such sentences as well. Let us call the assumption he makes to argue for this claim 'Quine's Razor':[9]

> [*Q*uine's *R*azor]
> A singular term can always be expanded into a definite description.

Now, if the singular term inserted for '*e*' in *N* is a name, we only have to rephrase it as a definite description. We may rephrase 'Vulcan' as 'the intra-mercurial planet'. Now we can go on analysing 'Vulcan does not exist' by substituting the definite description for the name:

S3 Vulcan does not exist.

*S3** The intra-mercurial planet does not exist.

*S3*** $\sim\exists x$ (x is an intra-mercurial planet & □y (y is an intra-mercurial planet $\rightarrow x=y$)).

Quine admits that in some cases we might not be able to find a translation for a name, because some names correspond to especially 'obscure or basic' notions, for which we have no independently established phrases. For these cases a device is required in order that translations might be found systematically. Quine's proposal is:[10]

> [*Quine 2*]
> *…we could have appealed to the ex hypothesi unanalyzable, irreducible attribute of being Pegasus, adopting, for its expression, the verb 'is-Pegasus', or 'pegasizes'. The noun 'Pegasus' itself could then be treated as derivative, and identified after all with a description: 'the thing that is-Pegasus', 'the thing that pegasizes'.*

[9] Quine 1948: 8.

[10] Quine 1948: 8.

In *Q2* Quine proposes that *S3* could be analysed by *S3*** if 'Vulcan' corresponds to an obscure or basic notion:

*S3** The thing that vulcans (is-Vulcan) does not exist.

*S3*** $\sim\exists x$ (x vulcans (is-Vulcan) & $\Box y$ (y vulcans (is-Vulcan) $\rightarrow$ $x{=}y$)).

Quine believes that we can, with the help of this device, translate *every* singular term by a definite description, and thus that *QR* holds. He argues that *PB* is generally misdirected *because* *QR* holds. I will argue against *QR*.

IV. Plato's Beard might not be generally misdirected

IV.i. QR has artificial results

I take it that *PB* is meant to imply a claim about natural language: *in our use of natural language* we cannot meaningfully deny the existence of something which is not. To show that this is false, one has to explain how the existence of something can be meaningfully denied *in natural language*. Thus the proposal should not be artificial. And this is precisely what Quine's proposal seems to be.

It is a consequence of *QR* that *names* can be expanded into definite descriptions. But names are expressions which do not describe the objects they designate, whereas definite descriptions do. Thus it is artificial to translate names by definite descriptions.[11]

The device to translate names systematically has even more artificial results. According to this method, 'Vulcan' is translated by 'the thing that vulcans' or 'the thing that is-Vulcan', where 'is-Vulcan' and 'vulcans' are to be novel predicates. First of all, adopting this method, we have to accept *a lot* of new predicates. Second of all, it seems to be *ad hoc* to accept such predicates only because we need translations for

[11] The same argument applies with respect to pronouns and numerals because they do not describe the objects they designate.

names corresponding to obscure or basic notions. And in addition, these predicates are peculiar. Most likely, 'is-Vulcan' does not apply to an object under the same conditions by which a common predicate like 'is orange' does: while 'is-Vulcan' seems to apply to an object only if 'Vulcan' is a name given to that object, an object satisfying 'is orange' does not have to bear the name 'orange' or any other name at all.

IV.ii. QR is false under its best interpretation

It is not very clear what it means to say that a singular term can always be *expanded* into a definite description. The expansions, or translations, are supposed to help us giving *analyses*. The main feature of an analysis and a translation is to give new expressions that have the same meaning as the analysed or translated ones. If something lacks this feature, it does not seem to be an analysis or a translation at all. The best thing to do is thus to interpret Quine as stating that a singular term can always be expanded into a definite description *salva sensu*. The relation between names and their expansions is thus the relation of *synonymy*.[12]

But there are some well-known Kripkean objections against this claim. I will trace Kripke's modal argument only, although the others – the epistemic argument and the arguments from ignorance and error – apply as well.

In his argument, Kripke uses the notion of a *rigid designator*: a singular term is a *rigid designator* iff it designates the same object with respect to every possible situation in which it exists and never another one.[13] According to this definition 'Dublin' is rigid, whereas 'the capital of Ireland' is not, since it does not designate Dublin in a situation in which Kilkenny, for instance, is the capital.

[12] This might be another reason to interpret Quine along these lines: Quine's aim is to show that sentences of the form *N* are generally meaningful. If a definite description substituted for a singular term in a sentence *N* is *not* synonymous with the singular term, how can we show that the original sentence is meaningful by showing that the new sentence is?

[13] Kripke 1980: 48.

Suppose that *QR* is true under the given interpretation, and that 'Dublin' is translated by, and synonymous with, 'the capital of Ireland'. If this is the case, 'Dublin' is not rigid, since 'the capital of Ireland' is not. But 'Dublin' *is* rigid. Therefore, the expressions are not synonymous.[14] The argument applies with respect to other translations as well, because names are generally rigid while ordinary definite descriptions are not. What *QR* suggests, that names are synonymous with *ordinary* definite descriptions, is false.[15]

Does the argument apply as well to the claim that names are synonymous with definite descriptions which we obtain by application of Quine's device? To figure out if it does, we shall have a closer look on these expressions. According to the device, 'Dublin' is translated by 'the thing that dublins' or 'the thing that is-Dublin'. Quine holds that the predicate 'is-Dublin' is composed of the copula and a general term. But instead of building a usual predicate, 'is' and 'Dublin' are said to be 'indissoluble'.[16] How can we understand those predicates?

The first thing to note is that Quine cannot treat 'Dublin' as a singular term because it would need to be translated again. And if 'Dublin' is instead a general term, the 'is' cannot be the 'is' of identity, because the 'is' of identity connects singular terms. One interpretation might thus be that Quine treats 'Dublin' as an ordinary general term, the 'is' as the copula and the hyphen as something that stresses that 'Dublin' is not here a singular term. The problem with this interpretation is that we get ordinary definite descriptions again and Kripke's argument reapplies.

In order to avoid this objection Quine has to claim that 'the thing that is-Dublin' is a rigid designator, and thus explain why the predicate 'is-Dublin' is such that it applies to Dublin in every possible situation.

One possibility to do this is to claim that the 'is' is, although no proper identity sign, something very like it. This move might help because if the 'is' is something very like the 'is' of identity, it might be held that

[14] Kripke 1980: 57.

[15] To avoid this objection Quine might try to specify a *special* kind of definite descriptions to translate names.

[16] Quine 1960: 178f.

something which is-Dublin is-Dublin in every possible situation, and thus that 'the thing that is-Dublin' designates it in every possible situation. The problem with this move is that it is woefully obscure – what should this special relation between the 'is' and the general term be? No satisfactory answer to this question is forthcoming.

V. Plato's beard is not generally misdirected

Quine's argument that *PB* is false is convincing. However, his argument that *PB* is generally misdirected is not, since there are objections against the assumption on which it rests, namely *QR*. It is *artificial* to translate all singular terms by definite descriptions, especially by definite descriptions like 'the thing that is-Dublin'. Furthermore, Quine's device for enabling this translation so seems to be rather *ad hoc*, and as such it forces us to accept a lot of entirely novel (not to mention peculiar!) predicates. Finally, Quine seems to be committed to the claim that names are synonymous with definite descriptions, to which Kripke's modal argument effectively responds. Thus I conclude that for all Quine has said, Plato's Beard is not generally misdirected.

Bibliography

Cartwright, Richard, (1960): 'Negative Existentials', *The Journal of Philosophy 57*, pp. 629– 639

Kripke, Saul, (1980): *Naming and Necessity*, Cambridge, Massachusetts: Harvard University Press

Quine, Willard Van Orman, (1948): 'On What There Is', in: *From A Logical Point Of View*, Cambridge, Massachusetts and London: Harvard University Press (2003), pp. 1–19

Quine, Willard Van Orman, (1952): *Methods Of Logic*, London: Routledge & Kegan Paul

Quine, Willard Van Orman, (1960): *Word and Object*, Cambridge, Massachusetts: The MIT Press

Russell, Bertrand, (1905): 'On Denoting', *Mind 14*, pp. 479–493

British Journal of Undergraduate Philosophy, 2(2), July 2007

A primer on formal metaphysics

Andrew Bacon
Lady Margaret's Hall, Oxford
andrew.bacon@lmh.ox.ac.uk

Introduction

Ontology is the study of what there is. Often this is taken to include the project of categorizing entities into various kinds: individuals, properties, events, or what have you, and exposing the various dependencies which hold between these kinds. Some philosophers believe there is a 'primitive ontology', an ontology which has the property that all things are, or can be reduced to a more basic 'primitive' kind of thing. For example, monists hold that commitment to minds isn't a commitment to anything *new* over and above material stuff. Two examples from the philosophical literature which demonstrate a difference in primitive ontology are absolute and relational theories of space and time. Absolute theories of space-time will often attempt to reduce objects to space-time points and properties distributed over them, thus taking space-time points and properties as primitive. Relational theories will attempt to reduce space and time to objects, events and relations between them, thus taking objects, events and relations as primitive.

From such metaphysical theories arises the need for a certain degree of formal apparatus. Relationists can construct instants of time out of events and the simultaneity relation, by taking sets of simultaneous events (equivalence classes). It then turns out you can linearly order these instants using the 'earlier than' relation between events. Similarly absolute space-time theorists can take objects to be sets of space-time points. There are various formal frameworks which allow us to locate this brand of structure including set theory, category theory, topology and mereology. Here I shall talk briefly about set theory and topology but will concentrate mainly on mereology as it is the most distinctly philosophical and most nominalistically acceptable of the mentioned frameworks. As with all the aforementioned theories, it has a rich catalogue of philosophical applications.

Historical Background

Mereology is a collection of formal systems designed to capture the notion of 'parthood' – the relation of a part to its whole. I say a collection because there are variations in how mereology can be formulated. These depend on what primitives you choose,[1] which axioms they include (this usually depends on philosophical disposition) and whether you formulate your theory using first or second order logic.[2] The study of the parthood relation has presocratic roots but it was not until Brentano's work that mereology made its mark on philosophy. It was still later that mereology became the rigorous formal theory that we know it as today. In the hands of the famed Warsaw School of philosophical logicians it became a powerful tool in the study of formal metaphysics, foundational issues in mathematics, not to mention its use in the theoretical computer and information sciences. Particularly important names in the history of mereology are Leśniewski and Tarski from the Warsaw school, then Whitehead, Leonard and Goodman for bringing mereology into the mainstream analytic tradition. Finally, more recent writers on mereology include Lewis, Simons and Varzi. For further reading on these writers see the bibliography.

The relation of part to whole

If you have done any formal set theory you will be familiar with the technique of capturing the inferences involving a particular relation by giving a set of axioms which govern those inferences. In the case of set theory that relation is the membership relation, written as '$\in$' to be read

[1] Whether you take your system to be explaining the relation 'is a part of', as opposed to say 'overlaps with' or 'is disjoint from'. You can define each of these relations in terms of the other, so for logical purposes it does not matter which relation you take as primitive.

[2] In first order logic your quantifiers only range over objects in the domain. If your logic is second order the quantifiers can also range over subsets of the domain, so you can get the effect of quantifying over properties and functions.

as 'is a member of'. In mereology the relation is the parthood relation, written '≤' and read as 'is a part of'. To get a good grasp of what we mean by 'part', here are a few examples involving the parthood relation:

1. My hand is part of my body
2. The dustbin lid is part of the dustbin
3. That slice is part of the pizza
4. Wales is a part of Great Britain
5. The second movement was my favourite part of the symphony
6. 'The Empire Strikes Back' was the worse part of the trilogy
7. The whole numbers are only part of the rational numbers
8. Being pedantic is part of being a good logician

There is quite a diversity of examples here. (2) is an example of a part of something which needn't be spatially connected to the rest of it, (4) demonstrates a geographical part and a constitutional part, (5) is an example of a temporal part and (7) of parthood between abstracta. Finally in (8) it is controversial whether the parthood relation is being used at all – I put this in to make it clear that not *all* uses of the word 'part' should necessarily fall under the treatment we are considering here.

Barring example (8), this is the intuitive notion of parthood that we shall be trying to formalise here. The diversity of the examples here indicates that the parthood relation is *topic neutral*. Topic neutrality is a desirable property among theories contending for the title of 'pure logic'. Some have said topic neutrality is the sign of the logical – we can clearly see there seems to be no domain over which we cannot quantify and hence apply first order logic to, similarly there are not many domains from which we cannot form sets.[3] Some philosophers, for example David Lewis, have taken this one step further and claimed that mereology should count among the purely logical theories (for example he argues that identity, a logical notion, is merely a limiting case of overlap, a mereological notion meaning 'shares a part with').

[3] There are some exceptions. All the sets cannot be gathered into one set for example.

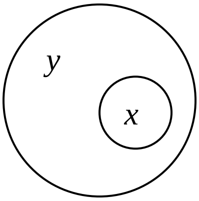
Finally there is a small caution to be noted about the way mereologists use the parthood relation. A mereologist will count the Eiffel Tower among the Eiffel Tower's parts, whereas in ordinary English we would only count strictly smaller parts of the Eiffel Tower among its parts. This is for convenience only – the mereologist could, if she wanted, take strict parthood as primitive, and define loose parthood (parthood which treats objects as parts of themselves) in terms of it by saying x is loosely part of y iff x is strictly part of y or x = y. Since they are interdefinable we shall always mean loose parthood when we talk of parthood short of an adjunct.

Some definitions

For sake of exposition, and for continuity with the literature, we shall take parthood as the primitive notion of mereology. As has been mentioned already, different relations can be used instead, for example 'overlap' or 'disjoint from'. Here I define in terms of the parthood relation some common terminology used among mereologists. The symbol for parthood is '≤' and, remember, it is to be read as 'is a part of'.

Proper part
x is a proper part of y, written 'x < y' iff x is a part of y and x is not the same as y

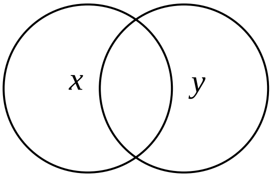
- $x < y \leftrightarrow [x \leq y \wedge \neg x = y]$



Overlap

x overlaps with y, written ‘x • y’ iff x and y have a common part

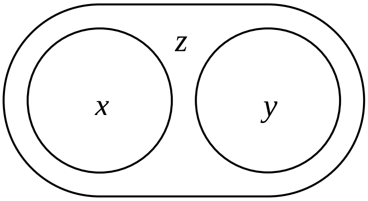
- $x \bullet y \leftrightarrow \exists z[z \leq x \wedge z \leq y]$



Underlap

x underlaps with y, written ‘x U y’ iff there is something of which x and y are both a part

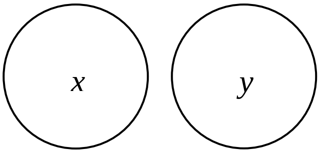
- $x \mathrm{U} y \leftrightarrow \exists z[x \leq z \wedge y \leq z]$



Disjoint

x is disjoint from y, written ‘x ⊥ y’ iff x and y do not have a common part

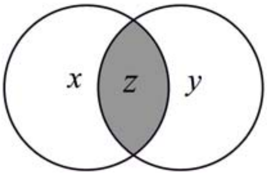
- $x \perp y \leftrightarrow \neg x \bullet y$



Product

If x and y overlap the product of x and y, written x×y is the object, z, whose parts are just the parts x and y have in common

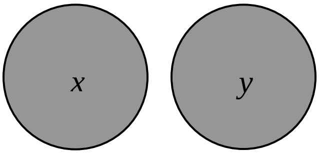
- $x{\times}y =_{\mathrm{df}} \iota z\, \forall w(w \leq z \leftrightarrow (w \leq x \wedge w \leq y))$



Sum

If x and y underlap the sum of x and y, written x+y is the object, z, such that something overlaps with z just in case it overlaps with x or it overlaps with y

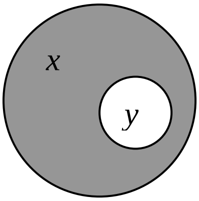
- $x{+}y =_{\mathrm{df}} \iota z\, \forall w(w \bullet z \leftrightarrow (w \bullet x \vee w \bullet y))$



Remainder

If x isn't a part of y the remainder of y from x, written x-y is the object whose parts are just those parts of x which are disjoint from y

- $x\text{-}y =_{\mathrm{df}} \iota z\, \forall w(w \leq z \leftrightarrow (w \leq x \wedge w \perp y))$



Axioms

It is now time to give some of the standard axioms of mereology. A few words on the need for explicit formal axioms are deserved. So far we have been talking about parthood intuitively, and whenever a formal definition of a term has been given it has always been accompanied by an equivalent in English. Why do we need to get formal at all? Part of the reason is that it allows us to get a handle on some infamously slippery notions, and apply logical techniques to test for the consistency and completeness of various theories based on them. Secondly it is all too easy to fall into paradox without due care. For example, some early versions of set theory allowed you to create sets by using a property to define its members. Similarly an obvious way to give an object would be to give its parts (I am my arms, legs, torso etc.). We might say that for any property F under which at least one things falls, there is an object x such that for any y, y is a part of x just in case y has the property F. So now let's consider the object whose parts are just those things which are not a part of themselves. (Here I'm using part in the layman's sense - mereologists say everything is a part of itself). If this object isn't a part of itself then it falls among the collection of things which constitutes its parts – it is a part of itself. If it is a part of itself then it has the property used to define it, namely that it is not a part of itself. This is a contradiction. Although this is less of a paradox than the analogue for naïve set theory, using a formal theory to make things explicit avoids linguistic confusions such as this.

So here are the axioms. Our language is first order and the only non-logical symbol is '≤'.

Reflexivity
Everything is a part of itself

- $\forall x\, x \leq x$

Anti-symmetry
If x and y are parts of each other, they are the same

- $\forall x \forall y[[x \leq y \wedge y \leq x] \rightarrow x = y]$

Transitivity

If x is a part of y and y a part of z, then x is a part of z

- $\forall x \forall y \forall z[[x \leq y \wedge y \leq z] \rightarrow x \leq z]$

Supplementation

If x isn't a part of y then there is an object whose parts are just those parts of x which are disjoint from y (for example take z = x-y)

- $\forall x \forall y[\neg y \leq x \rightarrow \exists z[z \leq y \wedge z \perp x]]$

Product

If x and y overlap then there is a unique object, z, whose parts are just the parts x and y have in common (i.e. z = x×y)

- $\forall x \forall y[x \bullet y \rightarrow \exists z \forall w[w \leq z \leftrightarrow [w \leq x \wedge w \leq y]]]$

Sum

If x and y underlap then there is a unique object, z, such that something overlaps with z just in case it overlaps with x or it overlaps with y (i.e. z = x+y)

- $\forall x \forall y[x \, U \, y \rightarrow \exists z \forall w[w \bullet z \leftrightarrow [w \bullet x \vee w \bullet y]]]$

The first three axioms simply say that parthood is a partial order. This is no surprise – the notion of a *part*ial order derives from the parthood relation. Reflexivity is a consequence of mereologists quirky terminology (see section 3). It should be noted that any mereology which has reflexivity and anti-symmetry as axioms will have the consequence that x and y are identical just in case they have the same parts. We call this extensionality.

Extensionality

- $\forall x \forall y[x = y \leftrightarrow \forall z[z \leq x \leftrightarrow z \leq y]]$

This is easy to prove. Suppose x = y, then something is a part of x just in case it is a part of y (this is an instance of the indiscernability of identicals). Conversely, suppose something is a part of x iff it is a part of y. Now x is a part of x by reflexivity, so x is a part of y from the assumption. By similar reasoning y is a part of x. So x is a part of y and y is a part of x. Now applying anti-symmetry it follows that x = y. That this is a theorem of standard mereology is sometimes said to pose problems for endurantism. For suppose object x undergoes the loss of one of its inessential parts y (for example if x lost a fingernail y). According to endurantism x will retain its identity despite this loss, so indexing our object with the times t and t` – before and after the loss – we get $x_t = (x\text{-}y)_{t`}$. However by supplementation $(x\text{-}y)_t$ exists and by extensionality it follows that $(x\text{-}y)_t = (x\text{-}y)_{t`}$ since they both have the same parts. Finally by transitivity of identity we get that $x_t = (x\text{-}y)_t$ which is a contradiction. Some endurantists have rejected extensionality on this basis and have developed intensional mereologies (cf. Simons, 1987, [11]). Similar problems can be formulated modally instead of temporally, and this fact can be construed as showing that these kinds of arguments are problematic for everyone, not just endurantists.

The last three axioms tell us that we can always subtract, take the product of or find the sum of entities. In particular, given any finite set of objects such that any pair of them overlaps (or underlaps) we can take the product (or sum) of them all simply by applying the respective axiom to the first and second object, then applying it again to this new product (or sum) and the third entity, and so on. Arbitrary products and sums are not permitted – it will take a stronger axiom schema to allow the product and sum of infinite collections of objects. There are some philosophical issues here involving the summation axiom. According to this axiom, for any two objects there is another object, their sum. This axiom can be applied whatever the objects are, leading to fusions of objects which needn't be spatially connected. Consequently philosophers have objected – it commits us to the existence of some very weird objects, for example the trout-turkey (an example from Lewis – the sum of a trout and a turkey). This is a matter of philosophical intuition, often the philosopher sympathetic with mereology will reply that what really exists out there has nothing to do with the way the human mind slices up experience to make it manageable. However, mereology claims to capture only the facts about the parthood relation. The objector might reply that, in contrast to the first five axioms, commitment to mereological sums does not seems to be a fact about *parts* at all.

Further Axioms

The following are various axioms that may be added to standard mereology to strengthen it. These principles are not assumed without being explicitly stated as they can often rest on controversial philosophical assumptions.

Unrestricted Fusion

Given any consistent property, there is at least one object, y, such that something overlaps with y just in case it overlaps with something having that property

First order version:

- $[\exists x\varphi \rightarrow \exists y\forall z[z \bullet y \leftrightarrow \exists x[\varphi \wedge x \bullet z]]]$

For any well formed formula φ with no free occurrences of y or z

Second order version:

- $\forall X[\exists xXx \rightarrow \exists y\forall z[z \bullet y \leftrightarrow \exists x[Xx \wedge x \bullet z]]]$

Unique Fusion

Given any consistent property, there is exactly one object, y, such that something overlaps with y just in case it overlaps with something having that property

First order version:

- $[\exists x\varphi \rightarrow \exists!y\forall z[z \bullet y \leftrightarrow \exists x[\varphi \wedge x \bullet z]]]$

For any well formed formula φ with no free occurrences of y or z

Second order version

- $\forall X[\exists xXx \rightarrow \exists!y\forall z[z \bullet y \leftrightarrow \exists x[Xx \wedge x \bullet z]]]$

Top

There is something of which everything is a part

- $\exists t\forall x[x \leq t]$

Bottom

There is something which is a part of everything

- $\exists b\forall x[b \leq x]$

Atoms

Everything has a part which has no proper parts

- $\forall x \exists y[y \leq x \wedge \neg \exists z[z < y]]$

Gunk

Everything has proper parts

- $\forall x \exists y[y < x]$

The Unrestricted and Unique Fusion axioms allow us to take arbitrary fusions of objects, whereas Sum only allowed us to take finite fusions. These axioms have been objected to on similar grounds as the Summation axiom. There is also a choice as to whether we use a first order or a second order logic. The pros of using a first order language are that they are supposedly ontologically innocent. Second order theories are said to commit us to sets and other abstract objects. The first order formulation is an axiom schema, and is thus actually infinitely many axioms (one for each choice of φ). The second order formulation, which allows quantification over properties, is only one axiom. On the down side, first order theories will always have unintended models. This is because, in a mereology with atoms, we expect the size of the universe to be 2^{κ} for some cardinal κ. If κ is finite so is the domain, and if κ is infinite, the domain is uncountable, so either way the domain is never countably infinite. If the mereology is gunky then the universe is always uncountable. But for first order languages there are always countable models if there are infinite models (due to the Löwenheim-Skolem theorem), so first order mereology will always have unintended models (cf. Bacon, [1]). Second order mereology avoids this problem. Also the Fusion axioms only quantify over monadic (one place) properties. Since we can interpret monadic second order logic in terms of plural quantifiers (Boolos [2]) we have a nominalistically acceptable way to formulate these axioms.

Top states the existence of the 'universe' – everything is a part of it. Bottom on the other hand is a widely rejected principle of mereology (except, perhaps, in universes containing only one thing). It states the existence of a 'null object', something which is a part of everything much like the way the empty set is a subset of every set. For obvious reasons the existence of a null object is philosophically spurious. However the existence of such an object makes mereology equivalent to a Boolean algebra, and assuming the existence of this object can simplify many proofs.

Call something an atom iff it has no proper parts, and call something gunky iff all of its parts have proper parts. Atoms then says that everything is made from atoms: the basic building blocks of the universe. Note that an atom is not to be thought of as a chemistry atom. What people take to be the atoms depends on their ontology. A popular choice might be space-time points although this isn't necessary (mereology doesn't choose you're ontology for you). Gunk on the other hand says that there are no atoms and everything is made up of gunk, which in turn is made up of more gunk and so on and so forth. Gunk is thus, in some sense, infinitely divisible. However gunkiness is a stronger property than that – a line (of real numbers say) is also infinitely divisible, you can keep cutting it in half, but a line isn't gunky since it is composed of points – it is a sequence of real numbers and each of these points has no proper parts. Gunkiness is a very bizarre property for something to have since it implies it has no basic parts. It would be very difficult to say what gunk was made of, since each part of it is made of more gunk, thus perpetually evading explanation.

Model Theory[4]

In this penultimate section we shall be concentrating on a particular collection of models for mereology. These are relevant to most of the metaphysical discussions involving mereology and should be enough to demonstrate the various dependencies between the axioms. If you understand this section you should be able to keep up with most of the philosophical literature on mereology.

The most important concept we shall need to grasp is Euclidean space. Euclidean space is a mathematical abstraction which is supposed to

[4] A model can by thought of as a mathematical structure which satisfies a given set of axioms.

model our intuitive idea of space (or space-time[5]). It can be thought of as a three dimensional graph each axis of which can be represented as a line of real numbers, written R. Three dimensional Euclidean space is then written as R^3 (or R^n for more generality). We now introduce the idea of a metric space. A metric space is:

- A non-empty set S
- A function, d, such that

$$d: S \times S \rightarrow R$$
$$d(x, y) \geq 0$$
$$d(x, y) = 0 \text{ iff } x = y$$
$$d(x, y) = d(y, x)$$
$$d(x, z) \leq d(x, y) + d(y, z)$$

Here S is to be thought of as a set of points. In our case we take S to be R^3. The function d is supposed to represent the distance between points in our set S. The first constraint on d says that d takes pairs of elements (two elements) from S and gives us a real number which is to be thought of as the shortest distance between those two points. The second and third constraint says that this distance is never negative and is zero between a point and itself but never between two distinct points. The fourth constraint says that the distance from x to y is the same as the distance from y to x. The last constraint says that for any three points x, y and z the distance between x and z is always more than the distance between x and y plus the distance between y and x (this can be seen intuitively by drawing a triangle of 3 points and noting that the combined length of any two of the sides will be greater than that of the remaining side). In the case of Euclidean space we define the function d as follows. Suppose x represents the three dimensional coordinate, (x_1, x_2, x_3) and y the coordinate (y_1, y_2, y_3) then:

- $d(x, y) =_{df} \surd(\ (x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2\ )$

This turns out to be a generalized version of Pythagoras's theorem. Don't worry if this doesn't make any sense to you – all you need to

[5] I shall talk only about space in three dimensions, but I will assume that this can be generalised to space-time and four dimensions if required (for example to discuss eternalism).

know is that d(x, y) represents the distance (as you would have intuitively thought of it) between x and y.

The next important concept we must tackle is the idea of a region of Euclidean space. We may think of a region of Euclidean space as a region of space as we would normally talk of it. However as all we have from the definition of Euclidean space is points and the notion of distance between points we must define a region of space to be a set of points. The region defined is simply to be thought of as the region of space which occupies just those points in the set. One important kind of region is the 'open ball'. An open ball should be thought of as a sphere minus its skin – a sphere without the spherical boundary surrounding it. Given a centre, a, and a radius, ε, we define the open ball around a of radius epsilon (the epsilon ball around a for short) as:

- $B_\varepsilon(a) =_{df} \{x \in R^3 \mid d(x, a) < \varepsilon\}$

This ball is open because it does not contain its skin (by the 'skin' of a region I shall always mean the two dimensional surface which surrounds that region). In general we shall define an open region, X, as follows. Remember X is a set of points from R^3.

- A region X is open iff for every $a \in X$, there is an $\varepsilon > 0$, $\varepsilon \in R$ such that $B_\varepsilon(a) \subseteq X$.

What this says intuitively is that, which ever point you take within the region (no matter how close to the edge of the region) you always have room to wiggle around in any direction and stay inside the region (this is expressed by saying that there is a ball small enough to fit inside X and contain your point). This is also equivalent to saying the region does not contain any of its skin, for if it did then there will be a point *on* the skin, and wiggling away from such a point will always force you to leave the region.

With this machinery in hand we should now be in a position to give some models for the various axioms given in section 5. Remember that mereology uses only one non-logical symbol, '≤', so to provide a model we simply must specify the domain and give an interpretation for '≤'. All this means is that we must specify a set of objects which the quantifiers of our theory range over, and a relation over our objects which is supposed to represent the parthood relation. For the six standard axioms with Unique Fusion, Top, Bottom and Atoms we shall take our domain to be regions of Euclidean space. Then to interpret ≤, we take the subset relation, ⊆, between regions. Remember that regions are sets of points and thus a subset of a set of points will correspond to a subregion of that region. So under this interpretation subregions are parts of regions. It is left as an exercise to the reader to show that the six standard axioms come out true on this interpretation (for example, Product and Sum are guaranteed by the fact that two sets always have an intersection and a union).

To see that Unique Fusion is true in this model consider the set of points, S, which satisfy the first order definable property φ (or the property X in the second order case). Given the interpretation, x and y are supposed to overlap iff the intersection of x and y is non empty. Something intersects with S non-trivially iff it contains a point of S – iff it contains (and hence overlaps with) a point having the property φ. Thus S has the characteristic that something overlaps with S iff it overlaps with something having the property φ. That this is true regardless of our choice of φ shows us that Unrestricted Fusion is true. Proving this fusion is unique is left to the advanced reader.

That Top and Bottom are true in this model is fairly easy to see. For Top we simply take the set of all points in Euclidean space. All regions of Euclidean space will be subregions of the whole of Euclidean space. Similarly for Bottom, take the empty set of points. The empty set is a subset of all sets and, in this model, is thus a part of all regions of Euclidean space (it is easy to see, in this case, why Bottom is so philosophically controversial). It is trivial to modify our model so that both ¬Bottom and ¬Top come out true. For ¬Bottom we simply take our domain to be *non-empty* subsets of Euclidean space, and for ¬Top we simply consider the *proper subsets* of Euclidean space.

To see that Atoms is true in this model we note that Bottom has no proper parts and is thus an atom, and similarly is also a part of everything. This is less helpful since most mereologists reject the existence of Bottom, so let us give a model for Atoms and ¬Bottom along with the standard axioms. Here we simply take the non-empty subsets of Euclidean space as for ¬Bottom. To see Atoms is true take the regions consisting of one point (the so-called singleton sets – a set containing exactly one point) to be counted as the atoms. Each of these regions contains only one element, and since Bottom (the empty set) is discounted, it will have no proper parts. Notice also that every region is a set of points so every set of points will have a singleton set containing a point as a subset. Thus in this model the points are the atoms.

Finally we shall give a model for Gunk. Gunk is the negation of Atoms and thus cannot be consistently added to a system already containing Atoms as an axiom. For this model we take the non-empty regular[6] *open* sets of points in Euclidean space as our domain. So here we have restricted ourselves even further by discounting all regions which contain part of their skin. So points are not in our domain since a point does not contain any open ball around itself (because open balls always have non-zero radius) so singleton sets are not open. All parts of an open set will have further proper parts since you can show it contains an open ball which is strictly smaller.

If you have been following so far, and you know your completeness theorem for first order logic, it should be clear that we can glean some independence results from the preceding remarks. An independence result simply says that a certain axiom is not already provable from some other axioms and hence isn't superfluous (which is a good thing). We have provided models for standard mereology + Unique Fusion + Top and for standard mereology + Unique Fusion + ¬Top. This means that, given standard mereology + Unique Fusion is consistent (which it is), Top cannot be proven from them. Similarly reasoning shows that Bottom is independent of standard mereology + Unique Fusion. Since

[6] In topological jargon, an open set is said to be regular if it equals the interior of its closure. The 'closure' of an open region is simply that region plus its skin, the 'interior' of a region is that region minus its skin (if it has any). Surprisingly removing and then replacing a regions skin can result in a different region!

Atoms and Gunk are mutually incompatible we can also show that these are independent from mereology because the model we provided for Gunk, non-empty regular open sets in Euclidean space, also satisfies standard mereology + Unique Fusion.

One very important model theoretic result about mereology, which just about trumps everything I've said so far, is the following:

Tarski's Theorem

- Any model of Standard Mereology + Unique Fusion + Top + Bottom is a model of a complete Boolean algebra and vice versa.

Similarly any Boolean algebra with the bottom element deleted is a model for Standard Mereology + Unique Fusion + Top + ¬Bottom (which is the formulation of mereology most people use). It is not important that we know what a Boolean algebra is, but if you happen to know then that is great. The result is here for completeness. A lot of results have been proved about Boolean algebras, so this result is very useful because it says all these results apply to mereology too!

Further Stuff

Mereology as I've discussed it above accounts for lots of facts about objects and their parts. However there has occasionally been the need to look at stronger systems which include some of the elements of mereology we have been discussing. What follows is a very brief overview.

The idea behind intensional mereologies was introduced by Peter Simons. The aim was to merge modal and temporal concerns with mereology. One of the problems with mereology as we have discussed it is that it has Extensionality as a theorem (recall that this meant x and y are identical iff they have the same parts). A consequence of this is that we seem to be committed to an ontology of 4D perduring objects (see section 5). Another consequence of this is mereological essentialism: an object necessarily has the parts it actually has. Intensional mereology mixes the parthood relation with elements of temporal and modal logic in an attempt to explain the problems surrounding extensionality. An important book to read on this is Simons, 1987, [11].

Mereology deals with the parthood relation. Various other concepts can be defined in terms of parthood, such as overlap, sums and so on. However the notion of connectedness[7] is not definable in terms of parthood – connectedness is a purely topological property. Mereotopology combines mereology with the idea of connection and has been used to solve various problems to do with boundaries, as well as being used extensively by computer scientists. Casati and Varzi, 1999, [3] cover mereotopology well. For the historical reading, Whitehead, 1929, [17] is probably its first appearance and see Clark, 1981, [4] for a rigorous version of Whiteheads system.

With the conclusion of these two very brief notes I feel you must finally be primed for formal metaphysics. We have barely scratched the surface of this fascinating topic, but I hope doing so has been useful and informative. And I hope even more that it has been fun.

[7] A region, X, is connected iff given any two points in X, you can draw a continuous line from one point to the other without leaving X. In topological terms we say a region is connected iff it cannot be represented as the union of two disjoint open regions, but it will take a little thought to see how these definitions are equivalent. For an explanation the term 'open' see section 7.

Bibliography

Bacon, A, 2006, 'First Order Mereology and Unintended Models', http://users.ox.ac.uk/~lady1900/writings.htm, unpublished

Boolos, G., 1984, 'To Be Is To Be the Value of a Variable (or To Be Some Values of Some Variables)', *Journal of Philosophy* 81: 430-449

Casati, R., and Varzi, A., 1999. *Parts and Places: the structures of spatial representation*. MIT Press

Clarke, Bowman, 1981, "A Calculus of Individuals Based on 'Connection'", *Notre Dame Journal of Formal Logic* 22:3, 204-217

Goodman, N, 1977 (1951). *The Structure of Appearance*. Kluwer.
Heller, Mark, 1984, "Temporal Parts of Four-Dimensional Objects", *Philosophical Studies*, 46: 323-334

Leonard, H.S., and Nelson Goodman, 1940, "The calculus of individuals and its uses," *Journal of Symbolic Logic 5*: 45-55

Leśniewski, Stanisław, 1992. *Collected Works*. Surma, S.J., Srzednicki, J.T., Barnett, D.I., and Rickey, F.V., eds. and trans. Kluwer

Lewis, D. K., 1991, *Parts of Classes*, Oxford: Blackwell

Needham P., 1981, 'Temporal Intervals and Temporal Order', *Logique et Analyse* 24: 49–64.

Simons, Peter, 1987, *Parts: A Study in Ontology*, Oxford: Clarendon

Simons, Peter, 2004, "Extended Simples: A Third Way Between Atoms and Gunk", *The Monist* 87: 371-84

Tarski, Alfred, 1984 (1956), "Foundations of the Geometry of Solids" in his *Logic, Semantics, Metamathematics: Papers 1923-38*. Woodger, J., and Corcoran, J., eds. and trans. Hackett

van Benthem, J, 1983, *The Logic of Time*, Dordrecht: Reidel (1991)

Varzi, A, 2003, "Mereology", *Stanford Encyclopedia Of Philosophy*

Varzi, A, 2004, "Boundary", *Stanford Encyclopedia Of Philosophy*

Whitehead A.N., 1929, *Process and Reality*, New York: Macmillan

Whitehead, A.N, 1919. *An Enquiry Concerning the Principles of Natural Knowledge*. Cambridge Uni. Press. 2nd ed., 1925

British Journal of Undergraduate Philosophy, 2(2), July 2007

Book reviews

Creation, evolution and meaning
Robin Attfield, *Ashgate*

Craig French
Heythrop College, University of London
craig.french@bups.org

Creation, Evolution and Meanihng is a book about God. Attfield's main thesis is a defence of divine creation. In writing about God Attfield exhibits his expertise across an extensive range of philosophical disciplines such as the philosophy of language, the philosophy of science, the philosophy of religion, ethics and the philosophy of value. Attfield's style is consistently and carefully argumentative. This is certainly to Attfield's favour since he engages with key philosophical figures such as David Hume, Michael Dummett, A.J. Ayer, Richard Rorty and D.Z. Phillips, as well as figures from outside of philosophy, such as the scientist Richard Dawkins, and the theologian Keith Ward. The book has three parts; my main focus will be on Part 1 'Meaning and Creation'. But first I'll briefly outline some of the key claims of Part's 2 and 3.

In Part 2, 'Creation and Evolution', Attfield discusses various arguments for the existence of God. His considered position comes in the form of a new version of the design argument, which he takes to be cogent. In the course of defending this argument, Attfield discusses and defends the existence of God against various criticisms, including those coming from versions of the Problem of Evil and those coming from Hume. A further feature of Part 2 is that Attfield argues (against, for instance the likes of Dawkins, and other neo-Darwinians) that the existence of a creator God is consistent with evolution and natural selection, as long as we understand that God *institutes* natural selection, and loves the intrinsic value which is the product of evolution. Attfield also offers insightful exegesis of Darwin's own views on the consistency of evolution and the existence of a creator God.

Part 3, 'Evolution and Meaning', offers reflections upon the meaning of action and life. Attfield argues that 'Life's meaningfulness turns out to involve an integrated sense of priorities, self-awareness and a sense of objective values which the people concerned can see themselves as safeguarding or honouring or promoting. In view of the value of the products of evolution, this can take the form of understanding oneself as a steward or trustee of such value' (p. 2). Furthermore, 'theistic stewardship (motivated by answerability to God the creator, regarded as the source of the world's value) has a greater coherence [than its atheistic counterpart], and much more directly makes life meaningful' (ibid).

Since Attfield wants to defend divine creation his first port of call is to ask what is *meant* by divine creation. In elucidating the concepts involved Attfield tells us that 'Creation… concerns not the Big Bang or some other earliest event, but the dependence of each and every physical entity on a divine creator, not situated in space or in time, possessed of the power and the knowledge to bring the world into being, and to select its natural laws' (p. 1). And regarding the concept of God, we are told that 'implicitly, to be God is to have the power, knowledge and wisdom to bring into being anything that can (without contradiction) be brought into being…' (p. 24). So, Attfield's concept of God is the concept of a non-spatiotemporal *creator* God.

In the course of clarifying the meaning of 'divine creation' Attfield offers a stern defence of a Realist understanding of religious language (he argues against forms of verificationism and the anti-Realism's of Dummett, Rorty, Cupitt, Phillips and Duhem/Quine). But even if we accept Attfield's arguments against anti-Realism, the question remains, as Attfield puts it, 'How [are we to] understand talk of what lies beyond experience when our language is perforce derived from everyday experience, concepts and purposes?' (ibid). Attfield recognizes that 'talk of God as author or as agent… cannot be taken in the ordinary sense of those terms, and the same applies to talk of God's will or purposes' (ibid). We need to know how it is possible to apply those predicates normally applied to people to God. And in finding out how this is possible, we must, Attfield thinks, steer a course between anthropocentrism and equivocation. Hence, Attfield endorses an analogical model of religious language (following Donald MacKinnon).

Ever since the time of Aquinas the analogical model of religious language has been proposed as a means by which we can understand the sense of religious statements such as those that predicate something of God. This, Attfield claims, works in two ways. First, we have the 'Analogy of Attribution', which 'suggests that God is [say] good in the sense of being the ultimate source or cause of goodness… [much like] fresh air… can be said to be healthy as being [a cause] of health' (p. 25). But thus far this is inadequate, as Attfield realises, since it 'fails to capture the meaning of 'God is good' (for much more is usually meant by this than that God is the cause of goodness)' (ibid). Hence, Attfield invokes a second aspect of the analogical model, the 'Analogy of Proportionality' according to which 'God's goodness and other attributes are held to be related to God's nature in the same manner or ratio as human goodness is to human nature' (ibid).

But Attfield seems to have got things the wrong way round in suggesting that the Analogy of Proportionality can pick up the slack left by the Analogy of Attribution. Being told that the ratio of God's goodness to God's nature is analogous to the ratio of human goodness to human nature presupposes that we understand what it is to attribute goodness to God in the first place, and we've already noted that the Analogy of Attribution is not suited to fully capture such understanding. So, at best, the analogical model is an incomplete model for understanding religious language. Therefore, it is not altogether clear how we are to understand religious language, and unfortunately Attfield doesn't have much else to say on the issue.

Attfield argues for the existence of a creator God, hence it is worth discussing how Attfield conceives of the God/World relation. Although this comes in Part 2 of Attfield's book, it can be linked in with the analogical understanding of religious language that comes in Part 1. We can begin by questioning even the partial meaning captured by the Analogy of Attribution. For example, we can understand the relation between fresh air and good health in physical (biological) terms, and we can understand it as a causal relation. But what about the relation between God and the good? Or, since we are discussing attribution to God *per se*, the relation between God and his creation (i.e. the physical world)? Since God is non-spatiotemporal it is not obvious that we can understand creation in *causal* terms. So how does the analogy help us? This is a difficult question. In a short section of Chapter 8 Attfield addresses the God/World (i.e. God/Creation) relation, and again aims to illuminate it through analogy.

Attfield remarks that, 'Belief in creation certainly means that creatures are dependent on God at all times, but it also means that God bestows on them their form; and this is done not in a single instant, but step by step in the course of evolution' (p. 166). Attfield views this as a 'timeless bestowing' (Chapter Four), but it is 'achieved though created temporal processes continually generating new forms and species… God creates through naturalistic processes, in which 'things make themselves'' (pp. 166-7). And here is where the issue of the relation between God and World comes into view, as Attfield says 'If creation operates in part through natural processes, God is to be seen not only as transcending the natural order but also as immanent in it… God will be present in the evolving world rather as a composer is present as his or her intentions are expressed during a performance of a music work such as a symphony' (ibid).

Attfield is offering a version of *panentheism*. Panentheism is, as Attfield puts it, quoting Arthur Peacocke, 'the belief that the Being of God includes and penetrates the whole universe, so that every part of it exists in Him, but (as against pantheism) that His Being is more than, and is not exhausted by, the universe'. Accordingly we get a mix of transcendence and immanence, Attfield sums up the position as follows

> *By actualizing and employing finite and temporal processes of creation, God does not cease to be infinite or eternal… much less become dependent on the created order, and by making creatures make themselves, and thus carry through the creative process, God does not cease to be changeless; nor need classical theism say otherwise. We are not tempted to say, even of a human composer*

> *such as Beethoven, that he grows or develops as a work of his is performed, even in cases where the performance occurs during his lifetime, simply on the basis that the delivery or execution of his intentions is spread across time… All the less should we be inclined to say that because God employs temporal processes of creation and is thus immanent in the world, the creator grows, develops or changes. (p. 170)*

Attfield has no direct argument for his panentheism. But this is OK; it serves as an *interpretation* of the God/World relation given what has already been argued in terms of creation and evolution. However, it simply isn't obvious that Attfield can have transcendence *and* immanence; he seems to want to have his cake and eat it. The matter is certainly open to interpretation since Attfield offers no explicit definitions of 'transcendence' or 'immanence' (perhaps because there are none). To return to a quote from earlier in the book, that God transcends His creation is given content in the claim that God is 'not situated in space or time' (and further flourishes on transcendence are that God is 'eternal and unchanging' and 'infinite'). But what is the significance of the term 'situated'? We can ask is God *situated* when He creates? (One might wonder, given what Attfield has to say, whether it even makes sense to talk of *when* God creates). Surely on any non-equivocal understanding of what it is to *create*, especially when we understand creation as motivated by the purposes and intentions of an *agent*, one (the agent) would have to be situated (somewhere) in order to create. And this seems to be what *immanence* captures here – that God is situated *within* his evolving creation. But then transcendence and immanence are completely at odds. God *cannot* be at once *not* in space and time, and *in* space and time. Attfield needs to reconcile creation *qua* 'timeless bestowing' with creation *qua* 'temporal processes of creation' (p. 170).

A further tension is that between Attfield's desire to avoid equivocation on the one hand and the notion of 'atemporal creation', or a 'timeless bestowing' on the other. What notion of atemporal creation uses a non-equivocal conception of creation? It's not clear at all what *atemporal* creation could even be, but whatever it might turn out to be, surely it can be *nothing like* creation as we understand it (i.e. temporal creation).

That is, we could not use locutions to characterise atemporal creation such as 'bringing into being' or 'making', without further specifying the *atemporal* sense of such locutions. But what is that sense?

Perhaps we can draw upon the analogy with the composer invoked by Attfield in order to clarify the God/World relation, and hence give some sense to the claim that God creates the world. But again, matters are not so straightforward. The claim is that God will be present in the evolving world as a composer is present as his or her intentions are expressed during a performance of a his or her work. But this suggests that we separate the creation (the composition) from the world (the performance). For example, Beethoven composes the 5th Symphony, and *then* it is performed and *then* Beethoven's intentions get *expression* in the performance of the Symphony. So, are we therefore to say that God creates the world *and then* God Himself gets expression in the world through evolution? If so then we have resorted to *temporal creation*, as in the Beethoven case. But again this is in tension with Attfield's insistence on *atemporal* creation. What's more, this understanding of the analogy makes creation sound suspiciously like a special event after all, like the special event of Beethoven composing the 5th Symphony. But then this is more like creation*ism*.

No doubt, one can understand the above analogy in different ways, and of course, understand 'transcendence' and 'immanence' in different ways. What I have tried to do, using the above quotations, is bring out some tensions and issues that require clarification. My suggestions are not motivated by any kind of anti-Realism, but are rather reactions to the positive things that Attfield himself says about the sense of religious statements. Nevertheless, Attfield's book is a vigorous attempt to defend the existence of a Creator God. The consistent attention to detail and strong argumentation is admirable, and hence *Creation, Evolution and Meaning* is a welcome addition to the philosophy of religion – and it would certainly be a happy supplement to the undergraduate's budding library.

Philosophical theology and Christian doctrine
Brian Hebblethwaite, *Blackwell*

Carl Baker
University of Leeds
carl.baker@bups.org

Brian Hebblethwaite's *Philosophical Theology and Christian Doctrine* is the third volume in Blackwell's 'Exploring the Philosophy of Religion' series, which aims to provide a middle ground between the abundance of introductory texts and in-depth monographs that are available in the discipline. This particular volume is an overview of how analytic philosophy has scrutinized the central tenets of the Christian faith, 'examin[ing] them for their meaning and plausibility' (p. x). Unlike much philosophy of religion, it does not (on the whole) spend its time trying to justify theism, but rather takes the core Christian doctrines as its starting point and subjects them to a philosophical examination.

Hebblethwaite's book provides an excellent starting point for anyone wishing to undertake study in this area, since the substantial notes provide an extensive bibliography of the discipline. The reader who wishes merely to gain an overview of the relevant positions will also gain much. Because the purpose of the book is to introduce the reader to a large number of debates in philosophical theology, it is inevitably breadth rather than depth that comes out on top – the book gives us a glance at *many* different questions, but rarely spends a large amount of time focusing on any one question. Each chapter contains around seven or eight sub-sections focusing on mostly distinct issues, which means that the book addresses something in the region of sixty deep problems in philosophical theology. This is not a regrettable feature of the book, however, since its aim is to provide an *overview* of these many problems. Those who wish to pursue the matters in further depth will make use of the comprehensive bibliographical references that are provided. While there is *a lot* to take in, Hebblethwaite does not bombard the reader with information in an unpalatable way. In the remainder of this review I will give a taste of the kind of questions that the author addresses in each chapter –though there is far more in the book than I have space to outline here.

Hebblethwaite's opening chapter provides a helpful recent history of the tension between theology and philosophy of religion. Some theologians, it seems, have been reluctant to accept that the results of philosophical study are relevant to their theological pursuits. Philosophy has been accused of an over-literal approach to the study of theology and of disregarding the analogical nature of talk about God. Indeed, some theologians influenced by Karl Barth have adopted a decidedly anti-rational approach, arguing that a complete theology will have a logic that is inaccessible to the human mind. Hebblethwaite convincingly argues that the questions theology is interested in cannot escape philosophical inquiry as regards truth and rationality. While philosophy *does* need to be sensitive to the context and tradition surrounding theological work, he argues, such demands cannot be used to immunise theology against critical analysis. And then before discussing the main creeds of Christianity, Hebblethwaite devotes his second chapter to a concept that underlies all of them: revelation. This covers the distinction between natural theology and revealed theology, and the intelligibility of claims that God 'speaks'.

The central chapters of the book deal with four key Christian doctrines – creation, incarnation, trinity and salvation. In addressing creation, Hebblethwaite explores which characteristics the first cause of the universe 'must have, if it is to fulfil its explanatory role and not be just part of what cries out for explanation' (p. 36). He argues for an understanding of God as *maximally great*, a conception which nullifies questions such as 'who made God?', since such a being is wholly self-explanatory. The second part of the chapter considers the created universe, including the alleged tension between science and religion. Hebblethwaite argues that the considerations of scientists such as Stephen Hawking have not demonstrated the redundancy of the explanatory role of the 'God hypothesis'.

One of the key doctrines that marks Christianity apart from other religions is that of the incarnation – the idea that Jesus Christ was the 'incarnate Son of God' (p. 57). Hebblethwaite's fourth chapter is devoted to examining what sense, if any, can be made of this claim. He outlines debates on this issue that have been prominent in contemporary philosophy of religion. There are at least three distinct opposing positions on the incarnation: that it is a *myth*, that it is a *metaphor*, and that it is a *truth*. The 'myth' position, typified by John Hick, argues that we could never have historical evidence sufficient to back up such an extravagant claim as the incarnation, and that it is simply a contradiction to believe that a man could be God. Likewise, if the latter charge of inconsistency sticks, then the incarnation is at best a metaphor. As Hebblethwaite points out, myths and metaphors are not useless – they can express 'truths hard to convey in straight prose' (p. 59). The creation myths found in the book of Genesis, for example, convey the utter dependence of the universe on God. In addressing the question of whether the incarnation *qua* truth is logically coherent, Hebblethwaite considers a number of positions, including the following: 'two-natures' Christology, according to which Christ alone possessed a human nature and a divine nature; the 'kenotic model', according to which Christ limited his divine attributes in becoming a human being; and Peter van Inwagen's use of the concept of relative identity to show that Jesus could have possessed properties incompatible with one another.

If the incarnation is not the most ridiculed of Christian doctrines, then the trinity is. The idea that God is one, and yet three, understandably draws puzzled looks. This doctrine is the subject of chapter five. Hebblethwaite considers two kinds of argument for a Trinitarian God – *a priori* arguments, and arguments from revelation. One *a priori* argument goes like this: love requires an object, so God could not possess the trait of perfect love unless there were an object of this love; therefore there must be more than one person in the Godhead. But why should we think that there are *three* persons in the Godhead? Hebblethwaite reports the idea that perfect love requires 'condilection' – that is, the mutual love of two beings for a third. He admits that, on its own, such an argument is 'tentative' (p. 81) – but suggests that a stronger case can be made when it is backed up with revealed theology from the scriptures. He argues that such an appeal to revelation is not a denial of rationality, because data acquired from revelation can still be subjected to analysis. His defence of the weight of the scriptures in *philosophical* debate about the trinity is unlikely to convince everyone.

Chapter six considers what sense it makes to say that Christ came 'for our salvation' (p. 91), and to reconcile man to God. This is what Christians call the 'atonement'. Hebblethwaite distinguishes two types of atonement theory: objective theories, which argue that our salvific state is actually *changed* by Christ's life, death and resurrection; and subjective theories, which suggest that Christ's life and death merely *inspire* us to change. The moral adequacy of varieties of these theories is considered, and sometimes rejected – for example in the case of the controversial 'penal substitution' theory, according to which the totality of God's wrath for mankind was taken out on his Son on the cross, thus paving the way for man to be reconciled to God in perfect justice. Hebblethwaite's discussion of salvation includes an analysis of concepts integral to the atonement, such as sin, justice and forgiveness.

The penultimate chapter addresses questions of death, eternity, heaven and hell. In tackling the difficult question of hell, Hebblethwaite outlines the dispute surrounding universalism, which is a popular position in philosophy of religion but unpopular in mainstream Christianity. Universalists believe that all of mankind will eventually be saved. They dispute the existence of an eternal hell, challenging it on moral grounds. Hebblethwaite (a universalist himself) reports Talbott's argument against hell: put simply, that it is inconsistent that a loving God should want to save everyone, have the ability to save everyone, and yet allow some to undergo eternal punishment. The author defends this universalist argument from a number of criticisms, including the claim that universalism seriously undermines human free will. The concern here is that a universalist God gives nobody a *choice* about whether they are reconciled to Him. This criticism forces the universalist to claim that all humans will eventually be won over to God *by their own free will* – and, because of this, he must admit that there are further opportunities for human salvation after death. The seriousness of this latter concession depends on one's other theological commitments. However, if it is necessary that all will eventually be saved by God, then the claim that humans could choose otherwise is undermined – how can my choice to do *X* be free if it is necessary that I will, eventually, do *X*? It seems that human freedom requires the *possibility* of resisting God indefinitely. If we have to admit this possibility, then it's not clear how we could be justified (at this stage in human history) in assenting to universalism.

In his final chapter, the main issue Hebblethwaite considers is that of special divine providence: the question of whether God intervenes directly in the course of human history. While this idea is central to Christian theology, it comes under attack from a scientific worldview and from the problem of evil (which together generate examples such as: if God intervened to make me catch the bus, why didn't he intervene to stop Auschwitz?). However, giving up the idea of special providence seems to put the idea of a personal God under threat. At the other end of the scale from a denial of special providence is the view that God's special providence is evident in absolutely *everything* that takes place on earth – but this latter view is again hard to reconcile with human freedom. We are also invited, in this final chapter, to consider whether special providence necessarily involves *miraculous* intervention, and how the above problems fit in with the idea of petitionary prayer.

This book will be accessible to those with an elementary grounding in philosophy or a basic knowledge of Christian theology. Those who lack such experience will find the book tougher, but still readable. While *Philosophical Theology and Christian Doctrine* is unlikely to convince those hostile to Christianity (or theism more broadly) that the positions it defends deserve assent, it may help to dispel some of the common perception that Christianity's central doctrines are obviously meaningless or incoherent. I heartily recommend it to anyone who wishes to learn more about this important division of the philosophy of religion.

Dawkins' God

Alister McGrath, *Blackwell*

The God delusion

Richard Dawkins, *Bantam*

The Dawkins delusion

Alister and Joanna McGrath, *SPCK*

Andrew Turner
University of Nottingham
andrew.turner@bups.org

Richard Dawkins' atheism has become infamous. It has been presented most recently and most fully in his provocative book, *The God Delusion*. The central idea that grounds his atheism is that a genuinely scientific attitude undermines belief in the existence of god. He supports this idea with arguments to the effect that belief in god is morally and scientifically superfluous as well as positively harmful.

Alister McGrath, on his own and with Joanna McGrath, tackles Dawkins' particular brand of atheism head-on. Their perspective is theological but their approach in this confrontation does not rest on theological assumptions. In fact, it is perhaps the best thing about the two McGrath books that the arguments therein ought to be acceptable to atheists, at least on principle – the criticism of Dawkins at no point rests on premises requiring belief in God. This approach has further merit insofar as it aims to expose Dawkins' ignorance of modern theological thinking. The general point in both *Dawkins' God* and *The Dawkins Delusion* is that Richard Dawkins builds something of a straw man out of theism by both misrepresenting it and incorrectly framing the debate between it and science. A fundamental criticism common to the two McGrath books is the charge that Dawkins' arguments are poor, and moreover that they are poor arguments as judged by the standards of scientific, evidence-based reasoning. The objection, then, is not only that Dawkins fails to understand the sophistication of theism, but also that the arguments Dawkins does advance fail to convince by Dawkins' own exalted (i.e. scientific) standards.

Here it is worth making a preliminary note regarding one common criticism of Dawkins – that his style of writing is somehow at fault. The claim here is often that when Dawkins discusses religion he is overly harsh, bullying, unsympathetic, and (or) just rude about theism. I agree that there is certainly an element of this in Dawkins' writing, and perhaps not all of it can be excused on account of the fact that his is a book that appeals to the popular rather than the academic markets. However, much to the McGraths' credit nothing substantial is made of this line of criticism – it is of course noted and lamented, but no essential use of it is made in their arguments. Nor *should* this line of criticism play any essential part in arguments against Dawkins. If Dawkins' arguments are sound and valid, then they stand regardless of how unpleasantly he might express them.

Putting these very general sketches aside, each book deserves a brief summary.

Explicitly the primary concern of *Dawkins' God* is to engage with the 'immensely problematic transition from *bio*logy to *theo*logy'[1] present in Dawkins' collective work up to and including *A Devil's Chaplain*. The book opens with a standard summary of Dawkins' popularised but sophisticated version of evolutionary theory. This section is concluded simply by noting that the potential scope of evolutionary theory is not limited to remaining an isolated explanatory hypothesis, existing solely within the realm of biology. Rather McGrath points out that the theory, as Dawkins presents it, represents a wider perspective on the world. For example, from an evolutionary perspective the concept of 'design' at the core of teleological thinking fails to be coherent in any cosmological sense.[2] Importantly and rightly McGrath distinguishes and distances Dawkins views from Social Darwinist views.

[1] *Dawkins' God*, p. 11.

[2] This is not new thinking on McGrath's part. Working out the scope and limits of evolutionary thinking is the subject of, for example, Daniel Dennett's *Darwin's*

McGrath goes on to explore how theism and atheism fit into an evolutionary perspective and then presents a discussion of the relationship between evidence and faith. The conclusions McGrath draws here are twofold. Firstly that theism and atheism are independent of evolutionary theory – that is, both are compatible with it – so Dawkins is just wrong to say that evolutionary thinking necessitates atheism. And secondly that Dawkins is massively ignorant of the important theological debates about the concept of faith and that concept's complexity, so Dawkins attacks the wrong target – and indeed a much weaker target – when he attacks what he takes to be definitive of 'faith'.

Next comes a discussion of 'memes', a concept invented by Dawkins as the cultural analogue of the gene and, among other things, put to use against religion. McGrath gives a sustained criticism of the notion of meme arguing, convincingly and I think devastatingly, that it simply fails as a worthwhile explanatory theory. Finally the book concludes by looking at the relationship between science and religion and puts forward a sketch for Christian theology that rejects the 'conflict model' of the relationship between science and religion. Again showing that Dawkins is wrong if he thinks that science and religion necessarily conflict.

The argument of *The God Delusion* goes as follows. First Dawkins argues that belief in the existence of a god is a belief that is truth-evaluable. According to Dawkins a universe that contained a god would be significantly different from a universe without one, so asserting the existence of a god amounts to an ontological claim that falls squarely within the domain of natural science. This Dawkins calls 'the god hypothesis'. It is worth noting that by 'god' Dawkins means something fairly particular; that is, a 'superhuman, supernatural intelligence who deliberately designed and created the universe and everything in it, including us.'[3] This stipulation spells out the content of the god

Dangerous Idea (1996), Penguin, London (Which is not to imply that Dennett makes a good job of it).

[3] *The God Delusion*, p. 31.

hypothesis. It is easy to see, then, that everything Dawkins says about the probability of the god hypothesis will fall wide of the mark if Dawkins' stipulation fails to capture what theists mean by god.

The second stage of Dawkins' argument is to show that the god hypothesis is wildly improbable and almost certainly false. Along the way he diffuses all the traditional arguments for a god's existence that are so often rehearsed in introductions to the philosophy of religion. He also offers at least one argument that is new (to me at least). It is I think worth a mention – the conclusion is that natural selection 'raises our consciousness' to the idea that science can explain how complexity emerges. The argument seems to be this. From natural selection we get the idea that complexity is necessarily the end result of a process of evolution. Complexity of any form can only occur after significant cosmological time has passed. A god is not simple. So the fundamental insight of natural selection, namely that complexity emerges slowly, finds an application to the effect that a non-simple god could not have been present at the beginning.[4]

Thus far Dawkins' argument has considered only the ontological aspect of theism. Now he moves to attack the moral aspect. He argues that both religion and morality have plausible naturalistic explanations. His conclusion here is not especially clear; perhaps it is that these naturalistic explanations undermine religion and its claim to moral authority. Though stating the conclusion like that makes it look highly suspicious. Alternatively Dawkins might not want to make so much of the naturalistic origins of religion – the conclusion might be more modest. This more modest conclusion would merely be that naturalistic explanations of religious belief make truth claims about the contents of those beliefs implausible rather than false.

Naturalistic explanations aside, Dawkins ends by making the case for the thinking that religious belief is pernicious. The three most important aspects singled out are: first that a 'religious attitude' tends towards fundamentalism; second that a religious upbringing is

[4] *The God Delusion*, pp. 114 – 120, and p. 31.

tantamount to child abuse; and third that the conciliatory function of religious belief offers the wrong kind of conciliation.

The Dawkins Delusion is, as the name suggests, a direct response to *The God Delusion*. It takes Dawkins on in four places:

First, the McGraths make a lot of the idea that *The God Delusion* builds a straw man out of religion. Dawkins' target, at the very least, just isn't representative of modern Christian thinking. Again, as in *Dawkins' God*, Dawkins' conception of faith is attacked for being misleading and simplistic. Particular criticism is levelled at Dawkins' discussion of the traditional arguments for the existence of a god. The McGraths' general point here can be summed up by saying that '[Dawkins is] clearly out of his depth, and achieves little by his brief and superficial engagement with these great perennial debates.'[5]

Second, another theme from *Dawkins' God* is taken up again. This is the claim, made by Dawkins, that science is incompatible with theism. The McGraths argue that this can't be right, and as evidence they cite the sociological fact that many scientists do believe in a personal god (though what this is meant to achieve I'm not sure). More significantly they argue that because Dawkins' arguments rest on the assumption that scientific criteria are the only good criteria by which to judge claims of religious belief, then Dawkins is incorrectly framing the debate and overlooking, or unfairly dismissing, other criteria on which to evaluate those claims. The point at its most general might be the claim that science does not have an unbounded scope to adjudicate all significant questions – though perhaps this is too crude since it makes Dawkins appear to be a naïve verificationist, something he certainly isn't.

Third, the McGraths criticise what Dawkins says about the origins of religion. This is surprising given what I think is an ambiguity, in *The God Delusion*, about what work this section is supposed to be doing. Again the method of this criticism is to show that Dawkins is building a straw man out of religion. For instance, there are important theological

[5] *The Dawkins Delusion*, p. 7.

debates about how to define 'religion'. These are debates which Dawkins makes no attempt to address. The implication here is that Dawkins is simply ignorant of them, or even worse that he is not interested in them. In addition there is a further criticism of Dawkins' use of the meme concept – a criticism that takes the original criticism from *Dawkins' God* further by linking it to Dawkins' purportedly 'pseudo-scientific' naturalistic explanation of religion.

Fourth and finally, the McGraths argue that the alleged harm and evil of religion is merely coincidental with religion rather then being characteristic of it. Their claim is that the kinds of evils Dawkins' thinks are peculiar to the religious attitude are in fact only accidental to it. Moreover they claim that belief in any ideal has a tendency towards fundamentalism whether religious or not – the failing that leads to that kind of evil is a human failing and not one that stems from any conceptual link to religion. To that end the McGraths end by giving an account of a 'Jesus ethic'[6] – which essentially amounts to liberal Anglicanism – that can provide a framework for theological criticism and, the hope is, avoid many of the evils that can admittedly result from Christian belief.

In the following I simply intend to highlight some significant areas of debate between Dawkins and the McGraths. What I want to do is pick out two interesting areas, which play a key role in the authors' overall arguments, and I hope to show that both Dawkins' and the McGraths' arguments fail to be convincing. The point that will be made repeatedly is that much more needs to be said on the subject than has been said by the three books reviewed.

Dawkins' claims can be distinguished as follows:

(a) The ontological claim: there is (almost certainly) no god.
(b) The moral claim: belief in god, and more generally religion itself, is harmful.

[6] My phrase.

This is, I think, a helpful way of dividing up the discussion. Note that Dawkins' atheism requires that both (a) and (b) be true, but that atheism at its most minimal requires only the truth of (a). All (b) is doing is supplying some extra punch and quite possibly motivating a move from atheism into anti-theism. The upshot of this is that Dawkins can be seen to be making a valid point against theism even if everything his says in arguing to for (b) turns out to be wrong (but obviously not if what he says against (a) is wrong).

Dawkins' argument for (a) aims to establish that the god hypothesis is truth-apt and that it is very improbable that it is true. An interesting point made by McGrath in attacking Dawkins' argument here is to ask why the theist can't deny that his claims about god are truth-apt? McGrath says; '[Dawkins] adopt[s] a very cognitive view of religion… Yet this is certainly not the sole aspect of religion; nor is it even necessarily the most fundamental. A more reliable description of religion would make reference to its many aspects, including knowledge, beliefs, experience, ritual practises, social affiliation, motivation and behavioural consequences.'[7] This is about as much as McGrath says on the mater, which is a shame. The point, however, is very worth mentioning because it can be made to do a lot of work against Dawkins if it proves viable. Assume that theism can be characterised non-cognitively, then Dawkins' claim that the god hypothesis falls under the remit of science must turn out false. To see this consider that on such an assumption god is no sort of 'hypothesis' at all – the framing of religious belief in the terms 'there exists…' and then taking it at face value is a misrepresentation of what the believer really means.

The obvious counterpoint here is that while a non-cognitive account of religion promises a lot, not enough has been said about it. Is it really clear that such an account could be given for instance? On the face of it, many religions do genuinely look as if they commit the believer to ontological claims that are straightforwardly truth-apt. If McGrath wants to defend his argument, then an explanation is owed of what the

[7] *The Dawkins Delusion*, p. 29.

proper meaning of these purported ontological claims is and why. To the detriment of the book, no such explanation is given.

Considering (b), neither Dawkins nor the McGraths are especially compelling with their arguments. Dawkins' arguments are attention-grabbing and deserve in depth treatment elsewhere (one such claim being that a religious education is tantamount to child abuse and that children have a special right not the have their minds 'addled by nonsense'[8]). There is room, however, for a general comment on the structure of the debate between Dawkins and the McGraths.

Dawkins thinks that religion necessitates certain kinds of evil. Specifically, he claims that the evils that arise out of religion can be characterised as evils that arise out of the attitude that religious belief encourages. For example the attitude that criticism ought to be suppressed and contrary evidence dismissed because truth is manifestly available from dogma. By far the biggest problem is that a lot of what Dawkins says on this subject doesn't do any work in establishing his conclusion. His conclusion, if I read him right, is that religious belief is pernicious *because* of some conceptual link between a religious attitude and the manifestation of various evils like child abuse (other polemical examples might include the oppression of women, suicide bombing, the perversion of good science and sexual repression). His argument for this conclusion doesn't amount to anything more than the citation of various examples of these evils. But merely citing examples of evil religious people and the evils they've done doesn't help to show that there is a conceptual link – examples of the evil that religion has done isn't the right sort of evidence for making his conceptual point. What we need to know is why a religious attitude inevitably has these results (if, of course it, does).

At this point there are two ways to go: if you sympathise with Dawkins you will want the conclusion to stand and so look for a better argument. Alternatively if you don't sympathise with Dawkins, then you will want the conclusion to fall. In which case it is enough to note

[8] In *The God Delusion*, p. 326. Note that the words are not Dawkins' own, but are merely quoted by him (with approval).

that what Dawkins supplies by way of argument is insufficient for his conclusion.

The McGraths think that the conclusion falls, but their position is stronger in that I think they would want to say positively that the conclusion is false, instead of merely pointing out negatively that Dawkins doesn't establish it. The problem is that it is hard to find an argument, in either book, to this effect. They accept that religion and evil often go together, and that religion and good frequently go together as well. The McGraths cite as an example Robert Pape's 'definitive account'[9] of suicide bombing and his conclusion that the motivation is characteristically political and not religious – religion is merely an instrument that makes the task of convincing people to kill themselves easier. McGrath implies that were religion not playing this role then it would be something else. Which is to say that religion is not what is causing the phenomenon, consequently getting rid of religion would not stop suicide bombings from happening. Put more simply the claim is that religion plays neither a necessary nor a sufficient role in the occurrence of this particular evil – and the point can perhaps be generalized cautiously.

The reason the McGraths' argument fails to be convincing is that they are attacking Dawkins at the wrong place. Remember that Dawkins' conclusion is a conceptual point linking a religious attitude to certain kinds of evil, which may be manifested by suicide bombings for example. My criticism of Dawkins is that he really only provides sociological evidence to the effect that religious people sometimes do evil things. Since this is the wrong kind of evidence for his conclusion he can't sustain the conclusion. Now all the McGraths seem to be doing is disputing what the sociological evidence really shows. But that is irrelevant to the conclusion I think Dawkins is trying to establish.

The McGraths are prepared to admit that 'Religion's in there, along with myriad other factors [in the ultimate causes of social division and exclusion]' but they say 'it also has the capacity to transform, creating a deep sense of personal identity and value, and bringing social

[9] *The Dawkins Delusion*, p. 50.

cohesion[10]'. This much can I think be admitted by Dawkins too, indeed it should be admitted because it's obviously true. The key question, however, is *why* religion is in the list of factors that cause social division – a conceptual question. For all their arguing the three authors don't get to grips with this, instead they only offer different accounts of how harmful religion is – a sociological question.

[10] *The Dawkins Delusion*, p. 53.

The British Journal of Undergraduate Philosophy is proud to announce…

The Winners of the BJUP 2007 Essay Contest

First prize:
What can Putnam and Burge tell us about belief?
By David Birch
St Andrews

Second Prize:
Evan's account of existential statements
Jack Farchy
Magdalen College, Oxford

Third prize:
Moral values and internalism about reasons for action
By Reema Patel
Girton College, Cambridge

This was all made possible by The Royal Institute of Philosophy, The Philosopher's Magazine, and Edinburgh University Press

www.royalinstitutephilosophy.org
www.philosophersnet.com
www.eup.ed.ac.uk

Upcoming BUPS and BJUP events

Philosophy is of course much, much better if you're with people who are passionate about the subject and know what they're talking about. BUPS and the BJUP exist to bring together undergrads who love philosophy. Our events offer opportunities to give or discuss really great papers, to meet and mix with other undergrads who think worrying about ethics or the fundamental structure of mind and world is kinda cool. To build an understanding of how philosophy is done across the country. To meet other students who like this stuff as much as you do, have done their reading and want to talk. BUPS also organises the UK's only big, annual national undergraduate philosophy conference, and the BJUP is Britain's only national undergraduate philosophy journal.

Interested?

Good, then you should be at the BUPS and BJUP events. You can get to information about all of these from our website – **www.bups.org**. Here you can also see a typical programme, browse past conference info, and even download a sample copy of an issue of the BJUP. If you're not already on the BUPS-L mailing list for announcements, or the BUPS-Dis forum for discussion, you can subscribe through the site. Don't worry – BUPS membership is free and our conferences are all tailored to fit a student budget. Submit a paper or come along when you can – we'd love to meet you!

Latest details of all our activities, profiles of the committee and a continually updated list of upcoming events are always available at: www.bups.org

Any enquiries can be addressed to: info@bups.org

Subscribing and submitting papers to the BJUP

BJUP Subscriptions

The BJUP is the Britain's only national undergraduate philosophy journal. We publish the best papers from BUPS' conferences, but also accept high-quality essays by direct submission.

Our non-profit status keeps the cost of subscription to our print version down, and all BUPS members receive the electronic version of the journal for free. New issues go out quarterly. We offer three levels of subscription:

BUPS Member Subscription (Electronic)
Becoming a member of BUPS is really, really easy – all you need to do is join the BUPS-L mailing list. The electronic version of the journal is distributed to all BUPS members. We hope you enjoy it!

Individual Subscription (Print)
An annual subscription to the print version of the journal costs £40 in the UK, and a little more for international postage. Printed in A5 size on 80gsm paper with a 250gsm card cover.

Institutional Subscription (Print + Electronic)
Institutions (libraries, schools, universities) wishing to subscribe to the journal receive both a print copy and a personalised electronic copy licensed for unlimited distribution to, and printing by, current students of the institution. This package costs £60 per year for UK delivery, slightly more for overseas postage.

Subscriptions run for a single academic year, a current subscription covering the print version of issues 2(1)–2(4). Full details of how to subscribe, and methods of payment we accept are available at the journal's webpage:

www.bups.org/BJUP

Submitting a paper to the BJUP

Most papers we publish will be 2,000 – 2,500 words in length. However we will consider papers of any length. We would suggest that you limit your submission to a maximum of 5,000 words, though, since papers longer than this are often better dealt with as a series of shorter, tighter, more focused essays.

What we're looking for in papers that we publish is actually quite simple. We like work that is:

- carefully structured
- argumentative rather than merely descriptive
- clearly written
- knowledgeable about a given subject area
- offering a new argument or point of view
- not just written for area specialists

As a general tip, don't write with 'This is for a journal, I must be technical, formal and use lots of jargon to show I know my subject...' running through your mind. Explanation to others who may not have read the same authors as you, clear laying out of thoughts and a good, well-worked-out and -offered argument that says something a bit different and interesting – these are the key characteristics of the best papers we've received. Don't be afraid to tackle difficult or technical subjects – we're all keen philosophers here – but do so as carefully and clearly as possible and you have a much better chance of being published.

Most of our papers are analytic, but we are delighted to accept and publish good papers in both the analytic and continental traditions.

We accept papers electronically as Microsoft Word .DOC. If you have problems sending in this format, please contact us and we will try to find another mutually acceptable file format.

Papers should be submitted via email to **bjup@bups.org** and should be prepared for blind review with a separate cover sheet giving name, affiliation, contact details and paper title.

Don't worry about following the journal's house style before submission. The only requirement we have in advance is that you follow English spelling conventions. Any other requirements will be made clear if your paper is accepted for publication.

Please do not submit papers for a BUPS conference and the journal at the same time. We'll make suggestions for rewriting or restructuring papers we think could be publishable with a bit of work. Please do not re-submit a particular paper if it has been rejected for a BUPS conference or the BJUP and has not been reworked.

Reviewing papers fairly is a difficult and time-consuming job – please give us a month or so and do not submit your paper elsewhere in the meantime.

We run the journal on the minimum copyright requirements possible. By submitting work you license BUPS and the BJUP to publish your work in the print and electronic versions of our journal, and agree to credit the journal as the original point of publication if the paper is later published as part of a collection or book. That's all – you are not giving us copyright over your work, or granting a licence to reprint your work in the future. We're budding philosophers not lawyers, so we hope that's pretty clear and fair.

Issue 2(2)

Plus: Book reviews, editorial, and event information

BUPS

British Undergraduate Philosophy Society

ISSN 1748-9393

2 2

9 771748 939006