APORIA

ST ANDREWS PHILOSOPHY SOCIETY JOURNAL

Issue 2: December 2008

Contributors:

Andreas Stokke

Michael Hicks

John Fluharty

Martin Beek

Mark Pexton

Also featuring an interview with Simon Prosser.

Edited by Joe Slater and Kyle Mitchell.

Welcome

Dear Philosophers,

Here we are sitting on a snowy Sunday morning, faced with a task that seems like too much of a challenge: “Write an introduction to a journal you haven’t even seen yet.”
In case you feel similarly overchallenged by the more difficult articles in here, why not join us in pondering some warm-up questions? Like the quote we found on one of the tea-cups which has been puzzling us for a while now: “Anything is good, if it’s made of chocolate.” (Jo Brand) Is that always true? (What sort of good – normative? Relative? What sort of thing – wholly chocolate? Or just partially? What sort of chocolate – dark? White? With nuts of without?)
Or how about this one: “Is it always better to take today off, knowing that having more energy will let you work harder tomorrow?”
Or, a last warm up: “Aporia means befuddlement – does that mean if I understand everything in here, I’ve missed the point?”

If you have answers to any of these, please find a pretty postcard and send them to us. Please feel free on that occasion to also join us in expressing great amazement, admiration and gratitude for the editors, contributors, sponsors, and, last but not least, readers of the second issue of the journal of the Philosophy Society of St Andrews.

Yours,

Jael Kriener
(President)

Duncan Reynolds
(Vice-President)

Credits

Special thanks to Simon Prosser for contributing the front cover photograph.
Design and Layout – Joe Slater.

University of St Andrews Philosophy Society Committee:

Jael Kriener – President
Duncan Reynolds – Vice President
Lukas Lohove – Treasurer
Benjamin Hofmann – Secretary
Fenner Tanswell – External Speakers Coordinator
Sarah Lohmann – Debates Coordinator
Joe Slater – Journal Editor
Kyle Mitchell – Journal Editor
Malcolm Collins – Publicity Officer
Hannah Arnold – Mundschenk

And thanks to all contributors and all who have supported Aporia.

Contents

5. Interpretation and Indexicality,
Andreas Stokke

15. Who Has My Thoughts?
Michael Hicks

23. Is it Reasonable for Art to have no Definition?
John Fluharty

31. Interview with Simon Prosser

34. Can Modal Agnosticism Save Constructive Empiricism?
Martin Beek

51. How We Each Maintain Our Personal Identity,
Mark Pexton

Interpretation and Indexicality

Andreas Stokke
Arché – Philosophical Research Centre
Department of Logic and Metaphysics
University of St Andrews
as572@st-andrews.ac.uk
February 2008

Prelude: What are Indexicals?

The term indexical is used in many ways in the literature. On its most inclusive use, the term seems to pick out any kind of linguistic expression the content of which depends on the situation, or context, in which it is used. On this broad construal all the following expressions are indexicals:

> *I, you, he, we, then, yesterday, now, here, over there, that, this, local (as in John went to a local bar), every bottle, all the students, strong (as in Steal isn't strong enough), the table, and many others.*

This list consists of expressions that are otherwise very different. *I, you, he, we* are pronouns, *here, now, yesterday* adverbs, *every bottle* a quantifier phrase etc. It is therefore highly likely that we will find that these expressions require quite different treatment. Hence, if we want the term *indexical* to have any substantive utility, we will need to constrain it considerably.

The most common way of doing so seems to be to construe indexicals as essentially *rule-governed* in the way they refer on an occasion of use. So, the category of indexicals, on this picture, may cut across different parts of speech such as pronouns, adverbs, quantifier expressions, predicates etc. For instance, one might think that a pronoun like *I* and a predicate like *strong* share the common feature that their content on an occasion of use – a person in the case of the pronoun, and a set in the case of the predicate – is determined by certain rules associated with the expressions. Both would therefore warrant the label *indexical*. Given this, whether or not some kind of expression is indexical or not depends on one's theory of how that kind of expression works in context.

Traditionally, however, theorists usually thought of indexicals as context-sensitive *referential* expressions. This narrows the field in that (at least) predicates like *strong*, adverbs like *local* and quantifier phrases such as *every bottle* (and perhaps incomplete definite descriptions such as *the table*) are ruled out. So on this picture, indexicals are first and foremost pronouns such as *I, you, he, that* and *this*, and adverbs such as *now*, *here* and *yesterday.*

The reason for this traditional way of carving the space seems precisely to be that most theorists would seem to agree that the way in which these expressions refer on a particular occasion of use is governed by rules, although the precise nature of these rules is highly contested, as is indeed the question of whether they should be regarded as rules of language, whatever that means.

The fundamental puzzle about indexicals, then, is *how* is this context-dependent reference achieved. Are there, for instance, linguistic *rules* which determine what an indexical refers to? Or are their reference perhaps more a matter of some kind of tacit negotiation between speaker and hearer? Or perhaps their reference just depends on what the speaker meant or intended?

All of these questions have deep consequences not only for the philosophy of language, but also for issues in philosophy of mind and epistemology. Indexicality is at the heart of the way we think and talk, and it is a well-documented fact that indexicality seems, moreover, to be *indispensable*. It has long been acknowledged that, given our cognitive limitations, it is impossible

to do away with indexicality. The phenomenon, therefore, cannot be relegated as a mere matter of a convenient way of facilitating communication; and by the same token, it requires more than just linguistic investigation. Philosophical problems arise and require philosophical solutions. This does not mean, however, that we should not recognize that indexicality is first and foremost a linguistic phenomenon, and the starting point of any kind of investigation of it, philosophical or otherwise, must therefore proceed from thorough examination of the role it plays in the way we use language to communicate with each other.

Abstract

The paper objects to the treatment of descriptive uses of indexicals proposed by Nunberg (1993). It is shown that the arguments Nunberg presents are ineffective against a pragmatic account of these uses such as that put forth by Recanati (1993). The conclusion is that there are good reasons to believe that the deferred interpretation in question are generated by non-semantic processes involving extralinguistic factors such as the speaker's communicative intentions.

1 Introduction

This paper is concerned with some problems surrounding the interpretation of utterances containing indexical expressions. (I will often refer to such utterances as `indexical utterances'.) In particular, the aim of the paper is to object to two arguments put forth by Nunberg (1993) regarding the interpretation of so-called descriptive uses of indexicals.

Nunberg presents a detailed and powerful theory of indexicality which he puts forth as an alternative to what he takes to be the inadequate standard view, namely the direct reference theory of indexicals. A version of the direct reference theory is endorsed by Recanati (1993) who directly engages with Nunberg's objections. In this paper, I will focus on Nunberg's objections, and I will not be concerned with the details of the direct reference theory. My aim is to show that Nunberg's two main arguments against the way the proponent of the direct reference theory handles descriptive uses of indexicals both fail.

Section 2 introduces the problem we are concerned with. Section 3 presents Nunberg's theory of indexicality. Section 4 objects to Nunberg's arguments. Section 5 attempts to draw some conclusions about the nature of utterance interpretation.

2 The Problem of Interpretation

One central problem of interpretation and indexicality is this: How do hearers arrive at the contents they do arrive at by interpreting utterances containing indexicals?

Let us take a somewhat naive perspective and think of things according to the following picture. An utterer produces a token of an English sentences containing an indexical expression. A hearer intercepts this token. What we are calling `interpretation' is the process, whatever it is, by which the hearer is lead from this interception to a state in which she entertains a certain content as a result of taking as input to the process the token produced by the utterer.

When the hearer has done this without erring along the way we are inclined to say that she has *understood* the utterance. In our everyday practices, we are often prepared to use locutions involving the verb *say* to describe something like the feature of utterances which is the target of hearers' interpretative processes; for instance, we have a practice of using notions such as *what is said*, *what the speaker says*, w*hat the speaker means to say* and so on.

As this suggests, we are prepared to distinguish between what is (literally) said by an

utterance and what a speaker means to say or communicate with an utterance. Correspondingly, we have at our disposal two intuitive notions of utterance understanding. That is, we can distinguish between understanding what is literally said and understanding what is meant by an utterance. To see this, we can note that it is possible to misunderstand what is meant while understanding what is said; and conversely it is possible to misunderstand what is said while understanding what is meant.

To illustrate, suppose Prof. X is the reader of the infamous student reference containing (1).

(1) This student is punctual and has excellent handwriting.

Suppose further that Prof. X were to interpret the utterance thereby made in such a way that the output content was simply that the student in question is punctual and has excellent handwriting. In this kind of case, normal practice does not hesitate in passing judgements like, “Sure, that was what the referee literally said in the letter, but Prof. X nevertheless still misunderstood; she didn't understand what the referee meant to say."

The reverse case is perhaps harder to construct. But it is perfectly possible to imagine that there could be a situation in which, for whatever reason, Prof. X interprets the utterance involving (1) such that on the one hand she takes *handwriting* to mean cooking skill; but on the other hand, she still takes the speaker to have intended to communicate the content that the student in question is not a good student. (Indeed, she might think that having excellent cooking skill is irrelevant for being a good student.)

This was Grice's (1989) point. At least at this level of analysis, we need a clear-cut distinction between the two types of content which we normally refer to by `what is said' and `what is meant', respectively. There is certainly no doubt that the way we talk about utterances outside our theorising makes such a distinction. It is natural to predict, then, that our theories will likewise need both notions to account for the phenomenon of utterance interpretation.

2.1 What kinds of Propositions can be meant by Utterances containing Indexicals?

A related question to the one we identified – how hearers get to understand indexical utterances – is the following: How comprehensive is the range of contents communicated by means of indexical utterances? In particular, availing ourselves of an established terminology of propositions, the question is: Are indexical utterances used to communicate only singular propositions, or are they also used to communicate general propositions? Roughly speaking, a singular proposition is a structured entity which contains an individual in the place corresponding to the subject of the sentence which is said to express it. By contrast, a general proposition is a structured entity which contains a property or a relation in the place corresponding to the subject of the sentence which is said to express it.

The answer to the present question seems at first hand straightforward. It seems to be an empirical datum that utterances containing indexicals are sometimes used to communicate general propositions. That is, there are cases in which, in order to count as understanding an indexical utterance, on one notion of understanding, it is required that the hearer arrive at a content which involves a property rather than an object. An example from Nunberg is (2).

(2) *Uttered by a condemned prisoner*: I am traditionally allowed to order whatever I want for my last meal.

Ordinary judgement does not hesitate in judging that a hearer who interprets the utterance in (2) such that the output of her interpretative process is the singular proposition involving the utterer misunderstands the utterance. Correct understanding, then, seems to require arriving at something

like a general proposition involving the property of being a condemned prisoner. Hence, it seems that an indexical utterance like the one in (2) can be used to communicate general propositions. (We return to the details of this case below.) In such a case, the indexical is said to be used *descriptively.*

The problem is whether this general content is arrived at by a process which, somewhere along the way, involves a literal content – in this case a singular content involving the utterer. Recanati thinks that is does; Nunberg thinks not.

Nunberg's paper presents an abundance of examples like (2) where it seems that an indexical utterance is used to communicate a general rather than a singular proposition. Recanati agrees that this conclusion is correct, but only for some of Nunberg's cases. Regarding these, Recanati holds a view according to which the general proposition is arrived at by a two-step interpretative process with a base level of literal meaning acting as input, the output being a level of non-literal, deferred content. As he says,

> It is true that *both indexicals and descriptions can be used either referentially or descriptively.* [...] Yet, at the basic level, indexicals must be given a *de re* interpretation, contrary to definite descriptions. (Recanati (1993, 314))

By contrast, for reasons we shall consider in detail, Nunberg argues that these descriptive uses of indexicals do not render themselves to the two-fold treatment.

We should mention a caveat about Recanati's framework before moving on. Recanati is explicit that he holds a view according to which

> the distinction between basic level interpretations and other, non-basic interpretations does not correspond to that between what is literally expressed (what is said) and what is merely `conveyed'. Non-basic interpretations such as those involved in Nunberg's examples of deferred reference themselves constitute `what is said' by the utterances which give rise to these interpretations. (Recanati (1993, 316))

So, why are we justified in following Nunberg in taking Recanati's position to be one where the basic level of interpretation – i.e. the level where *I* in (2) makes singular reference – is a level of literal meaning, the descriptive reading arising at the next level by a pragmatic process?

The reason has to do with the way Recanati uses the term `pragmatic'. The crucial point is that although the basic level is reached via pragmatic processes applied directly to the sentence meaning composed out of the linguistic meaning of the terms, the basic level is defined by Recanati as "the level of interpretation which is reached when no optional p-processes [i.e. pragmatic processes] occurs." (315) In other words, the basic level is reached solely by mandatory processes. By contrast, the next level – i.e. the level to which *I* in (2) contributes a property – is arrived at by optional pragmatic processes.

In other words, the difference between Recanati's view and the view we, along with Nunberg, are attributing to him which takes the general readings of descriptively used indexicals as arrived at by implicature is merely a terminological difference. The important point is that Recanati holds that the pragmatic process by which we arrive at the general content involving the property of being a condemned prisoner by interpreting (2) is optional, whereas the process by which we arrive at the basic level, which is required for the next level, is mandatory.

I therefore take Nunberg to be correct in the relevant respects when he summarises Recanati's view as follows:

> On Recanati's view, there is a coherent level at which indexicals like *I* can be given a "literal" interpretation where they refer directly, with their descriptive readings arising as a kind of implicature. (§4)

As we shall see, one of Nunberg's main arguments against this view is exactly that he thinks that such a putative basic level may involve contents which are incoherent.

3 Nunberg's Theory of Indexicality

In presenting Nunberg's theory I will focus on two aspects, which I take to be the most central. The first is a distinction Nunberg proposes to account for the meaning of indexicals. The second is his

view of the difference between referential terms such as indexicals and non-referential terms such as descriptions.

3.1 The Index-Referent Distinction

A central device of Nunberg's theory is a distinction between what he calls the *index* and the *referent* of an indexical. The former corresponds to what would traditionally be taken to be the contextually identified referent, i.e. roughly the value of the Kaplanian character associated with the term in question (where the context of utterance is the argument of the function). In other words, the index of *I* is the utterer, the index of *you* is the addressee and the the index of *tomorrow* is the day after the utterance. Nunberg reserves the term `referent' to what is contributed to the output of interpretation.

One central purpose of Nunberg's examples, we can now see, is to show that the thesis that the index and the referent of an indexical are always identical, which Nunberg attributes to the proponent of the direct reference theory, is false. For instance, with respect to the condemned prisoner case, Jesse James serves as the index whereas the property of being a condemned prisoner is the referent.

The immediate question now becomes: What is the relation between index and referent? A perhaps natural view is to hold that the referent is obtained from the index by means of some pragmatic process. Thus, one might hold that the prisoner case should be accounted for by appeal to a distinction of the kind we sketched above, namely between literal meaning, what is said, and a content which is arrived at by pragmatic processes involving, among other things, the speaker's communicative intentions. As we have seen, Recanati opts for a picture essentially like this.

However, Nunberg denies that the cases of descriptive readings of indexicals should be accounted for in terms of pragmatic transfer. For him, it is the lexical meaning of the indexical which takes us all the way to the referent, via the index. Thus, he takes the meanings of indexicals to be

> composite functions that take us from an element of the context [the index] to an element [the referent] of a contextually restricted domain, then drop away. (§2.5)

In other words, the meanings of indexicals are functions from indices to referents. This means that the process by which we are lead from Jesse James to the property of being a condemned prisoner is mainly a semantic process on Nunberg's view.

3.2 Asymmetry

Nunberg accepts that there is an asymmetry between referential terms such as indexicals and non-referential terms such as descriptions. Since descriptions can be used referentially just like, as Nunberg's examples show, indexicals can be used descriptively, the asymmetry cannot lie at the level of the range of interpretations.[1] Rather, what is at stake is exactly the question we pinpointed regarding how the descriptive interpretation of indexical utterances are generated. On Nunberg's view,

> indexicals can have roughly the same range of interpretations that descriptions can: the utterances that contain them can express singular or general propositions, as the case may be. What makes indexicals exceptional is the manner in which their interpretation arises. A description characterizes its interpretations; an indexical provides an object [i.e. the index] that corresponds to it. (§2.5)

For Nunberg, then, the interpretation of the prisoner's utterance, i.e. the general proposition involving the property of being a condemned prisoner, arises out of two factors. One is the meaning of the pronoun, which first gives us the index, i.e. Jesse James. This object is now the object that

1 [1]Donnellan (1966) famously showed that definite descriptions have both attributive and referential uses.

“corresponds to" the thing which ultimately goes into the interpreted content. But how exactly does that process work?

As Nunberg notes, a natural response would perhaps be to say that the property contributed by the indexical in the prisoner's utterance comes from the index, i.e. Jesse James. So, we would take one of Jesse James salient properties, i.e. of being a condemned prisoner, and let that go into the interpreted content. However, Nunberg argues that this will not cover the broad range of non-referential uses of indexicals. We do not have to go into the details of this. The important thing is that Nunberg concludes that

> most of the work of specifying the interpretation is accomplished in the contextual background, rather than by the utterance, in a process mediated by speaker's intentions, the linguistic context, considerations of relevance and so on. Taken together, these factors define a domain of possible referents, along several dimensions. (§2.3)

With respect to our favourite example, it is easy to see why this appeal to broad contextual factors is needed. Even if we agree that the meaning of *I* can take us all the way to a referent which is distinct from the index, it is implausible that *for any case*, the lexical meaning will pick out the particular referent that an interpreter must hit upon to achieve understanding in that case. It is no part of the meaning of the pronoun which deters us from selecting as referent say the property of being a male individual whose first name starts with a *J*. It is because we share some vital information with the speaker, e.g. that he is a condemned prisoner.

This is the reason that, as we saw, Nunberg holds that the meanings of indexicals are functions from indices to contextually constrained domains. Yet, this also means that the process by which we arrive at the interpretation involving the property is not a *purely* semantic process, on Nunberg's view. This prompts the question of why, in that case, Nunberg still thinks that the Recanati type view is wrong. As indicated above, one reason is that he thinks that in cases like the prisoner case, the putative literal content would be incoherent. The central point of this is that if this is correct, then the process by which we arrive at the general interpretation is, contrary to Recanati's claim, not optional.

I will first examine in detail Nunberg's arguments for taking the process of deferred interpretation to be non-pragmatic. I shall then return, in the conclusion, to the point about optionality.

4 Are deferred Readings arrived at by Pragmatic Processes?

Nunberg has two main arguments for denying an account of the index-referent divergence involved in cases like the condemned prisoner's utterance (2) in terms of pragmatic transfer. The first argument attempts to establish that the putative literal level of interpretation might involve incoherent contents. The second argues that there is a significant difference between the processes of deferred interpretation in cases like (2) and ordinary processes of pragmatic transfer.

4.1 First Argument: Incoherence of Literal Content?

The first argument is expressed in the following passage:

> sentences containing descriptive uses of indexicals may be incoherent if the indexicals are interpreted as making singular reference. [...] In context, the adverbs *usually* and *always* [and *traditionally*] must be understood as involving quantification over instances, but these readings are not possible if the subjects of the sentences are interpreted as referring to individuals or particular times. So it is hard to see what coherent “literal" interpretations we could assign to these utterances. (§4)

We normally have no problem with using referential terms referentially in environments that are controlled by these adverbs. We can say things like “Nunberg usually makes good points" etc. Rather, the thought is that reading *I* in (2) as contributing Jesse James to the interpretation conflicts

with what is contributed by the adverb, which is analysed as a quantifier. On the contrary, I will argue that there is no such conflict.

Suppose we read *I* in the prisoner case as referring to Jesse James. That reading is represented by (3).

(3) Jesse James is traditionally allowed to order whatever he wants for his last meal.

One influential proposal for treating adverbs such as *usually* and *traditionally* as quantifiers is given by Lewis (1975). On the view Lewis opts for, the adverbs in question are regarded as quantifiers over what he calls `cases', where a case is a tuple of participants which provide values for the free variables in the sentence embedded under the quantifier. Lewis also presents persuasive arguments that the adverbs cannot be taken to be quantifiers over moments of time. So, charity compels us to not read Nunberg's "instances" in this way. Lewis' apparatus of cases therefore seems congenial to Nunberg's thought.

A case is an admissible assignment of values to the free variables that are used to represent the participants in the cases. As a device for restricting which assignments, or cases, are admissible, we can adopt Lewis' idea of using if-constructions. Following Lewis' recipe, then, we analyse (3) roughly as follows:

(4) Traditionally, if *x* is Jesse James and *x* is a condemned prisoner, *x* is allowed to order whatever *x* wants for *x*'s last meal.

We would then regard *traditionally* as an unselective quantifier, the resulting truth conditions requiring the embedded sentence in (4) to be true in most admissible cases/assignments, i.e. those that satisfy the if-clause.

This means that we are analysing *traditionally* like we would *usually*, i.e. as inducing truth conditions in terms of *most* cases. To be sure, it might be argued that *traditionally* has features that *usually* does not, but I ignore these complications, since they arise as artefacts of the case at hand. The example serves the same purpose if formulated with another adverb, as Nunberg's other cases clearly suggest. Similarly, if it is found that *traditionally* should work more like *always*, then the same point applies that this adverb is likewise one used by Nunberg to construct examples about which the claim about the incoherence of the singular reading is put forth.[2]

It might be questioned why we are allowed the specification in the if-clause that *x* be a condemned prisoner. The thought might be that, since this must be drawn from the extralinguistic context, we have abandoned Nunberg's semantic project. However, this problem is also a problem for Nunberg. As we saw above, he allows that the context does a lot of work in preparing the work for the lexical meaning to pick out the referent, via the index. So, we may take it that (4) is a plausible candidate analysis of the singular reading of the original utterance in (2).

In other words, we are now analysing (3) as meaning something like that in most cases where Jesse James is a condemned prisoner he is allowed to order whatever he wants for his last meal. Nunberg's claim about incoherence then seems to be motivated by the thought that one cannot order one's last meal more than once. I now want to make two related points regarding Nunberg's first argument.

1. *The singular reading does not lead to incoherence*. It is clear that if (4) is the right analysis of (3) and its truth conditions are as suggested above, then the reading of the original (2) on which *I* makes singular reference is certainly not incoherent. Indeed, (3) is not even intuitively incoherent. One way of bringing this out is to notice that by adjusting the setting, we can construct a context of utterance such that there is a strong intuition that (3) is true relative to that context.

2 Cf. for instance Nunberg's example, "Tomorrow is always the biggest party night of the year." (§4)

For instance, imagine a world in which there is no tradition that condemned prisoners have a privileged last meal. In this world Jesse James is a notorious criminal who has been sentenced to death a large number of times. However, each time he has been standing at the gallows with the noose around his neck, he has been pardoned at the last moment. Over the years, a tradition has evolved by which Jesse James is allowed to order whatever he wants for the meal he has on the night before he walks to the gallows. Our intuition is clear that, in such a world, an utterance of (3) is true. Indeed, it seems that, in such a world, the most natural interpretation of the original utterance in (2) is one on which *I* makes singular reference to Jesse James. (We return to this last point below.)

This shows that, whatever we might think of it, the singular reading of (2) is not incoherent, as Nunberg claims.

2. *Truth value intuitions about the singular readings vary depending on the context of utterance.* Parallel to the above, we are likewise able to construct situations relative to which we get the clear intuition that (3) is false. Just take a world where there is no tradition about the last meals of prisoners at all.

In other words, it seems that our intuitions about the truth value of (3) vary depending on which world we are evaluating at. This shows that (3) has intuitively clear truth conditions, and as I have suggested one proposal for analysing them is the one above using Lewis' treatment of adverbs of quantification.

This does not alter the fact, of course, that our intuitions regarding the status of (3) in our own world are clouded by the fact that, barring fantastic escapes etc., one does not get to order one's last meal more than once. Why is this?

I think the reason is very similar to the reason that our intuitions about case like (5) seem unclear.

(5) *Uttered about a room which contains one single book which is black*: Most books in the room are black.

In standard treatments of generalised quantifiers, *most* receives a clause of the following rough approximation (where F is the set of Fs and G is the set of Gs):[3]

(6) `Most Fs are Gs' is true iff $|F \cap G| > |F - G|$.

On this analysis, then, (5) is true iff the set of things in the room which are both books and black is larger than the set of things in the room which are books but are not black. Consequently, on this analysis, (5) comes out true. This analysis corresponds to the intuition that 'strictly speaking' *most* means more than half.

Nevertheless, ordinary speakers are likely react to the utterance in (5) with puzzlement. Yet, it is not unlikely that they could relatively quickly be brought to agree that the utterance is in fact true, although inappropriate. One way of seeing this is by comparing the utterance with the corresponding one using *all*, as in (7).

(7) *Uttered about a room which contains one single book which is black*: All books in the room are black.

We have no hesitation in judging the utterance in (7) true. Furthermore, 'All Fs are Gs' entails 'Most Fs are Gs'.[4] Realising this, along with the recognition of the truth of (7), should lead us to accept the

3 See for instance Larson & Segal (1995, ch. 8).

4 It might be observed that if, as is standard, `All Fs are Gs' expresses the subset relation between the Fs and the Gs, i.e. $F \subseteq G$, then for the

truth of (5).

The sense of infelicity we get from (5) is thus best explained as arising from the fact that it is, as we put it above, true but inappropriate; that is, it is hard to see what anyone could intend to communicate with such an utterance, although literally true.

This suggests that there is a persuasive case to be made that the utterance in (3) shares the same feature of being true but inappropriate, given that *traditionally* is analysed as a quantifier the truth conditions of which parallel those of *most*. In other words, in a world like our own where prisoners do not get to order their last meals more than once, the utterance in (3) is true but inappropriate. It is true because the number of cases in which Jesse James as a condemned prisoner did get to order whatever he wanted for his last meal, namely 1, is larger than the cases in which he did not, namely 0. But it is nevertheless infelicitous, since it is hard to see what anyone could want to communicate by uttering it.

To sum up, Nunberg's first argument against the view that the general reading of *I* in the prisoner's utterance (2) is generated by pragmatic transfer fails. The singular reading is not incoherent. As we saw, it has intuitively clear truth conditions although bringing them out requires some reflection. With respect to our own world, the utterance is true, although conversationally peculiar.

Furthermore, we might ask ourselves why, even if it *were* incoherent, that would show that the general reading cannot arise out of a process of implicature or a similar mechanism of pragmatic transfer. Indeed, it seems most likely that strong reasons could be given that such cases where what is (literally) said is incoherent – as for instance an utterance of `John is both a lawyer and not a lawyer' – generate implicatures in predictable ways.

4.2 Second Argument: Indifference of Expression Choice?

Nunberg's second argument for rejecting that the readings in question are generated from pragmatic transfer is expressed in the following passage:

> such a process would be expected to be indifferent as to whether the initial reference to the index was accomplished via indexical reference or the use of a proper name or referentially used description. These disparities could only be accounted for by postulating a semantic apparatus of some sort, which is to say that there must be a semantic provision for deferred interpretation. (§4)

This argument strikes me as stronger than the first one. Still, it will be fruitful to examine it in some detail before concluding anything from it.

Let us compare two scenarios. The first is the original one where Jesse James utters (2) using *I*. In the second scenario, a bystander utters the alternative (3). (Both repeated here.)

(2) *Uttered by a condemned prisoner*: I am traditionally allowed to order whatever I want for my last meal.
(3) Jesse James is traditionally allowed to order whatever he wants for his last meal.

Although the passage quoted above is somewhat condensed, I believe it would be most charitable to read it as presenting the following argument about the difference between these two utterances: If the process which takes us from *I* in (2) to the property of being a condemned prisoner is a pragmatic process, then it should be possible to move from *Jesse James* in (3) to the property of being a condemned prisoner *by the same kind of process*. Since this is not possible, the process involved in (2) is not pragmatic but semantic.

I believe that it is possible to use (3) to communicate the general content. As we did earlier,

entailment to be validated the clause in (6) must be amended so as to allow for the case in which $F = \emptyset$. So, the right-hand side of the biconditional in (6) should read '$|F \cap G| > |F - G|$ or $F = \emptyset$ '. I ignore this complication on the grounds that the entailment I am appealing to is merely an intuitive one. It is unquestionable that such an entailment holds in all the cases where $F \neq \emptyset$.

we can test our intuitions by varying the context of utterance.

Suppose we are in a world where there is only one condemned prisoner, namely Jesse James, and no one else has ever been condemned to death. Further, after James was given his sentence, capital punishment was abolished. So, James is the only prisoner that ever has been, is or will be condemned to death. Finally, James has been scheduled to receive his punishment several times, but each time the carrying out of the sentence has been held up by unexpected mishaps. Now imagine the following dialogue taking place in this world:

A: Is a condemned prisoner allowed to order whatever he wants for his last meal?

B: Jesse James is allowed to order whatever he wants for his last meal.

Given this stage-setting, we get a strong intuition that B's utterance communicates the content that a condemned prisoner is allowed to order whatever he wants for his last meal. What B is trading on is James' property of being a condemned prisoner. And she can do this in answer to A's question because, in this case, James is the only one who has that property.

Now, whether this proves Nunberg's claim about the difference between the two utterances wrong depends on to what extent we can assimilate the process of transfer involved in B's utterance to the one which is involved in the original case, i.e. (2).

Above (4.1) I argued that the singular reading of (2) is coherent and has firm truth conditions. It is attractive to think that the way in which the general content about condemned prisoners is generated is very similar to the way it is generated from B's utterance. What difference is there between the two cases, other than the fact that the *ease* by which we move to the deferred content depends on the contextual backdrop?

In other words, it seems that the process by which we arrive at the deferred content *is* indifferent to the choice between an indexical or a proper name, contrary to what Nunberg claims.

5 Concluding Remarks

The upshot of the above is that Nunberg has failed to establish that the deferred, descriptive interpretations of indexicals are generated by processes which are significantly different from familiar processes of pragmatic transfer, such as the one which takes the interpreter to the general proposition in the case of B's utterance of (3) using the proper name *Jesse James*. Thus, it seems that there is still room for a position like Recanati's.

As described earlier, the central point of Recanati's view is that the deferred interpretations are generated by pragmatic processes which are distinguished by their optionality. Since, as we saw, the singular reading of the indexical utterance in (2) is not incoherent, as Nunberg argued, it seems that the process by which interpreters arrive at the general content involving the property of being a condemned prisoner is indeed optional. That is, whether the process kicks in depends on the contextual background involving, among other things, the speaker's communicative intentions. The same point clearly applies to the general interpretation of (3) containing a proper name.

Recall that we said that intuitively there are two different notions of utterance understanding, i.e. of interpreting utterances successfully. On one of them, a hearer has interpreted the utterance correctly when she arrives at its literal content, what is said. On the other, understanding requires hitting upon the content the speaker intended to communicate with the utterance.

I believe that a lot of the appeal of a view like Nunberg's comes from our sense that a hearer who does not move to the deferred, general interpretation of (2) is intuitively guilty of having misunderstood the utterance in the second sense, whereas she might be said to have understood the utterance in the first sense. This is parallel to the case where Prof. X simply takes the student reference to mean that the student in question is punctual and has excellent handwriting. In these

scenarios, there is a sense in which the further process of deferred interpretation is *not* optional, i.e. it must be undertaken in order to achieve understanding in the sense of interpreting correctly what the speaker intended to communicate.

But it might be questioned why we need the notion of understanding the literal content at all. In cases where the literal content and the intended content diverge, there is a strong intuitive pull towards not attributing understanding unless the intended content is arrived at. In the cases where the literal content and the intended content coincide, understanding could then be explained by the same notion of understanding the speaker's intended content.

However, it seems that there are cases where we particularly need the notion of literal understanding. One way of seeing this is to imagine a scenario in which the hearer has no way of accessing what the speaker might have intended with the utterances. For instance, imagine that the hearer of (2) only possesses the information that the speaker is Jesse James, but knows nothing about James' doings or the traditions pertaining to condemned prisoners. In such a scenario, the singular, non-deferred, interpretation of the utterance is the most reasonable one for the hearer to opt for. And significantly, it would be wrong to say that in such a scenario the hearer has failed to understand the utterance; it is just that she did not have all the information required to realise that the speaker had a different communicative intention.

Consequently, it seems that the Gricean premonition that our theory of utterance interpretation will need both notions of what it is to interpret an utterance correctly is reinforced.

References

Donnellan, K. (1966). Reference and definite descriptions. *Philosophical Review*, *75*, 281–304.

Grice, H. (1989). *Studies in the way of words*. Cambridge, Mass. and London: Harvard University Press.

Larson, R. & Segal, G. (1995). *Knowledge of meaning: An introduction to semantic theory.* Cambridge, Mass.: The MIT Press.

Lewis, D. (1975). Adverbs of quantification. In P. Portner & B. Partee (Eds.), *Formal semantics - the essential readings* (p. 178-188). Oxford: Blackwell.

Nunberg, G. (1993). Indexicality and deixis. *Linguistics and Philosophy*, 16, 1–43.

Recanati, F. (1993). *Direct reference*. Oxford: Blackwell.

Who Has My Thoughts?

Michael Hicks, Virginia Commonwealth University

The too many minds problem can be adapted to attack nearly every account of personal identity. The problem can be phrased loosely as a question: why do certain things count as people and others not? For example, if this human organism is a person, why isn't this brain also a person? It seems to be thinking; but I (the organism) insist that I am the person, and the brain is just a part of me. The problem also arises as an objection to "perdurantist" theories that maintain that persons persist by

having temporal stages as proper parts; the perdurantist maintains that every person is made up of a number of different "temporal parts" at different times, the maximal combination of which is a person. However, there doesn't seem to be any strong reason to deny personhood to every person stage, and they seem even more likely to be persons than brains—they can look at themselves in the mirror, scratch themselves, pick their noses, and perform all sorts of other actions typically associated with persons that brains just cannot. While each version of the problem attacks a different view, they all rely on the same fundamental intuitions and can be presented by essentially the same argument.

In this paper I will canvass a few of the ways in which the argument has and can be applied and show how they can all be easily resisted by blocking a central premise.

Too Many Minds and Animals

The first use of the argument that I will examine is that of Eric Olson (Olson, 105). Olson utilizes the argument as an attack on psychological accounts of identity. He begins by pointing out that, according to the psychological approach, there are two coincident objects wherever there is a human animal: a psychological continuer and a human organism. At any time, these distinct objects have all and only the same parts. The difference in these two objects, according to Olson, lies entirely in their "modal or dispositional properties" (Olson, 105), namely, the human animal has the property of possibly continuing to exist without higher brain function, whereas the psychologically continuing person could not simultaneously exist and be brain dead. Olson claims that "[o]n the psychological approach, a rational, conscious being with the wrong persistence conditions would not be a person. But if you and I are not animals, there would seem to be plenty of beings with the right psychological features to be persons but the wrong persistence conditions, namely, human animals." (Olson, 106). Olson sees this as a *reductio ad absurdum* of the psychological approach. "There could not be non-people who are exactly like people but for their persistence conditions" (Olson, 108).

So, Olson's argument has two premises and is valid via modus tollens: (1) if the psychological approach is correct, then there are some things exactly like persons psychologically that aren't people, and (2)there are no things exactly like persons psychologically that aren't people. Thus, the psychological approach is not correct.

The second premise to Olson's argument can be seen as an appeal to a sufficiency condition for personhood: having complex psychological properties is sufficient for personhood. This is a strongly intuitive principle; it is difficult to see how something could be as psychologically complex as a person without being a person. While it is beyond the scope of this paper to give an analysis of just what constitutes a complex psychological property, we easily distinguish clear cases of such higher-order thinking from clear cases of *insufficiently* complex psychologies: the difference between a normal human and an eagle, for example. While this explication is vague, it is precise enough for the purposes of this paper. Hereafter, I will use 'thought' to mean the sort of complex psychological property sufficient for personhood.

Two Further Applications of Too Many Minds

Before I consider solutions to this version of the problem, I will give two other applications of it. The first is from Shoemaker (Shoemaker, 499-500). Shoemaker points out that Olson's own position is vulnerable to a similar attack. Olson claims that persons just are human animals. Shoemaker points out that even under Olson's view there is an object coincident with and indistinguishable from every person that has different persistence conditions: his "corpse to be." This is an object which is coincident with the human animal, apparently shares all of its physical properties, but will continue to exist after that animal dies. Apparently the only difference between a person and his

corpse-to-be is what Olson calls "modal or dispositional properties:" persistence conditions. Shoemaker claims that if Olson's reasoning is accurate, the corpse-to-be is exactly like a person psychologically, and so we can construct an argument logically indistinguishable from Olson's in which the first premise reads (1*) if the biological approach is correct, then there are some things exactly like persons psychologically that aren't people. According to Shoemaker, this argument is at least as strong as Olson's, despite being directly opposed to Olson's view. So, something must be wrong here. Shoemaker then offers a solution, which I will briefly discuss later in this paper.

The third application of the too many minds problem (so-called by Shoemaker) that I will consider is from Trenton Merrick's book *Objects and Persons*. Merricks uses the problem to motivate his position of ontological eliminativism. He sets up the problem explicitly as a paradox consisting of four statements:

(1) Within the region filled by atoms arranged (normal, healthy, awake) human organismwise, there is exactly one conscious entity.
(2) Any object with atoms arranged (normal, healthy, awake, human) brainwise among its proper parts is conscious.
(3) Within the region filled by atoms arranged human organismwise, there is a human organism that has atoms arranged brainwise among its proper parts.
(4) Within the region filled by atoms arranged human organismwise, there is a brain that has atoms arranged brainwise among its proper parts.

(Merricks p.49)

Merricks points out that any three of these statements are compatible, but the conjunction of all four is a logical contradiction. We must choose to deny at least one. Merricks then argues for three of these claims and against the fourth.

Merricks claims that we ought to accept (1). According to Merricks, denying (1) leads to an unacceptable skepticism: Merricks and Olson agree that if there are two things thinking my thoughts, it is impossible for me to determine which of them I am. As Olson has it, if only one of them is a person, then I cannot through introspection determine that I am a person (I could have the wrong persistence conditions). According to Merricks, we will also be led to unacceptable uncertainty about the truth of some of our statements, such as "'I am not a mere brain but instead a human organism'" (Merricks, 50).

It is not clear that Merricks' defense of (1) given here works, but only for reasons that expose more fundamental difficulties involved in denying (1). These difficulties arise from the fact that uses of self-referring terms are singular. If there are multiple thinkers of a given thought, then singular terms such as 'I' fail to refer. So, if 'I' does not refer to anything, then 'I could not know whether I was a mere brain rather than a person' would be a false sentence, as would *any* sentence that expresses an introspective report about the thinker (where 'the thinker' is a definite, and singular, description). It is also possible that, if (1) does not hold, then there are a plethora of thoughts corresponding to each token introspective report—one thought per thinker. Unfortunately, this is just as conceptually problematic. Either of these situations are epistemologically more deeply counterintuitive than the one explicated by Merricks and Olson. They get it right; we ought to accept (1).

Merricks also argues that we ought to accept (2): "[d]enying (2) might lead to unwarranted skepticism about who, or what, is conscious" (Merricks, 50). I will revisit this defense later in this paper. So, according to Merricks, we are left with (3) and (4). Merricks argues that, because of what sort of things we believe persons to be, we ought to accept (3)—that human organisms exist—but deny that brains exist (4), leaving us with a form of ontological eliminativism in which only simples and objects with non-redundant causal powers, such as persons, exist.

Note that there are a number of other ways to apply Merricks' version of the paradox: given any (non-dualist[1]) account of personal identity, simply find two distinct objects that have what

1 Some, but not all, dualist accounts of mind are immune to the problem. An account which allows for objects

Merricks calls 'atoms behaving brainwise' among their parts and you have the same conundrum. Considering the universality of the problem, no particular view of personal identity seems preferable as a response to it. In the next section I will examine some of these responses, and show that they fail.

Initial Solutions

The best options involve attacking the argument on grounds neutral to the various views of personal identity. We have three choices: we may accept that there can be multiple objects, only one of which is a person, all sharing the same thoughts; we may follow Merricks and reject the existence of any objects counting atoms behaving brainwise among their parts other than persons. Finally, we may reject the view that two physically identical objects must be psychologically identical.

Harold Noonan chooses to reject the second premise in Olson's argument, that there are no things exactly like persons psychologically that aren't people, and to deny Merricks' first statement, that "[w]ithin the region filled by atoms arranged (normal, healthy, awake) human organismwise, there is exactly one conscious entity" (Merricks, 50). Noonan first admits that there are in fact plenty of things that share the thoughts of any given person and then attempts to assuage the counterintuitive situation (Noonan, 209). According to Noonan, we don't realize that we constantly have company because of the way our language is formed. This also allows us to resist uncertainty regarding statements like "I am not a brain" or "I am a person." Noonan argues that the reference rule for 'I' is not that 'I' refers to whatever is using 'I', but instead to the *person* using 'I.' Noonan does not claim that only persons are able to use 'I;' instead he claims that whoever is using 'I' is talking about the person using 'I' (Noonan, 210). Recall that on the schema that Noonan endorses, there are a number of objects that all have thoughts, one of which must be a person. So, if anything is thinking a token thought, there is guaranteed to be a person thinking that same thought; Noonan believes that all of the thoughts reference the person. Noonan claims that this fixes the epistemological problem outlined by Olson and Merricks: "I can know that I am a person, since if I were the animal and not the person thinking the thought I am currently thinking in thinking I am the person, I would *still* be right" (Noonan, 211). This also allows singular terms to properly refer.

Noonan's response fails, partially because it is difficult to make sense out of Noonan's proposal. Note that Noonan himself comes dangerously close to breaking his own rule for the reference of 'I' when he claims that "if I were the animal […] I would *still* be right" (Noonan, 211). Depending on how this sentence is meant to be read, he may or may not actually break his rule here; if he doesn't, this statement is very odd: it is a subjunctive conditional in which the antecedent is necessarily false. The natural understanding of this sentence, in which the antecedent is true at some worlds (and 'I' refers to an animal), is not a conceptual possibility under Noonan's framework. This shows that Noonan is making an unjustified claim about the meaning of 'I', which he does not sufficiently defend; rather, he merely posits that we accept it to circumvent the paradox. In absence of further argumentation, Noonan's response to the argument seems *ad hoc*, especially considering that it does not confront the most fundamental issues of the problem.

Noonan's response fails to address either of the deeply counterintuitive situations of the too many minds paradox: firstly, if the paradox goes through, two distinct objects can share, not just qualitatively identical propositional attitudes, but numerically identical thoughts. Noonan's attempt to assuage our intuitions by explaining how all of these statements will always be true only addresses the surface of the problem. Recall that the heart of the dilemma was never about the truth of our statements; our inability to know whether our statements are true or false is merely a way of illustrating the absurdity of the consequent: *something else* is (also) thinking your thoughts.

---

composed of both mind and body is vulnerable to the attack, as both the mind by itself and the composite object made up of the mind and the body together seem to be thinking; however, dualist accounts that deny that the mind and the body together make up an object get around this difficulty.

Secondly, and more importantly, Noonan does not recognize any sort of psychological complexity as being sufficient for personhood. The intuition that psychological complexity is logically linked to personhood underlies Olson's second premise. It has strong intuitive backing and supports, not just Olson's biological view, but also most psychological accounts of identity across time. To these accounts, the person *just is* whatever is thinking the right thoughts. If Noonan denies this principle, he ought to provide some strong argumentation, which he does not.

So, we should accept that only one thing is thinking our thoughts, and that there are no things that have psychological properties indistinguishable from those of persons that fail to be persons. This rules Noonan out, and leaves us with a few options: first, we can follow Merricks' line and embrace ontological eliminativism, or we can deny that all physically identical objects are psychologically identical.

Here we should not commit ourselves to Merricks' stark ontology. The intuitive backing for the existence of non-person objects complex objects is clearly stronger than the support for the claim that (as Merricks puts it) "Any object with atoms arranged (normal, healthy, awake, human) brainwise among its proper parts is conscious" (Merricks, p. 49). Merricks presents his solution to the problem of too many minds as motivation for ontological eliminativism rather than as a knock-down argument for it. His main argument for his view comes later. So, Merricks denies that there are non-person objects for reasons independent of the too many minds problem. If we are going to follow him in ontological eliminativism, we should likewise do so for independent reasons, and be thankful that we do not have to worry about the problem of too many minds. If, however, we are not already eliminativists, we should not deny the existence of non-person animals for the sake of this paradox.

The Proposal: Brains Don't Think

This leaves us with one option: denying that all objects with functioning brains as parts have thoughts and psychological properties. In other words, we ought to deny Olson's claim that the human animal is "psychologically identical" to the person, despite being physically identical at a microscopic level. This at first seems *ad hoc*—aren't we just denying thoughts to animals and brains simply to circumvent Olson's argument?

Let us first examine the support that Olson and Merricks present to support the claim that all intrinsically physically identical objects are psychologically identical. Merricks supports his second claim (2) just by pointing out that "denying (2) might lead to unwarranted skepticism about who, or what, is conscious" (Merricks, 50) and challenging whomever denies (2) to "offer a replacement that would explain why atoms arranged brainwise, for some kinds of things (like me) is connected to consciousness, but not for other kinds of things (like my brain)" (Merricks, p. 51).

Olson supports his premise by giving a thought experiment according to which you are put into a replication machine. The machine does not destroy you; however, it does create an object physically identical to you a few feet away from you. Olson notes that, because the new object is physically identical to you, we believe that it is a person. He then notes that the only physical difference between this object and you is relational: it is a difference of a few feet. A few feet couldn't make prevent something from having psychological properties, so our intuitions stipulate that we can know that an object is conscious based only on its nonrelational microphysical features. So, he thinks, if this object is conscious, it looks like the human animal coincident with you is conscious--after all, it has the same microphysical features as your duplicate.

Here Olson's example shows that relational properties cannot make a difference in the psychology of two otherwise indistinguishable beings. This is, of course, a few steps away from showing that *only* microphysical features can make a psychological difference between two objects; presumably, you and your duplicate have the same modal and dispositional properties, even those that are not microphysical. But it does lend some intuitive support to the notion that really only the

microphysical is relevant.

Olson then examines various reasons to believe that two physically indistinguishable objects could be psychologically distinct. He considers the view that, because psychological properties are *emergent* properties, it should not be so surprising that two physically identical beings have different psychological properties. After all, the psychological properties are distinct from the physical properties. Olson rejects this view, saying, "things cannot have different emergent properties unless there is some underlying intrinsic difference between those things" (Olson, 101). Olson then uses the example of fragility, which is a property emergent from the microphysical features of the object, to support his claim that there must be some physical difference to correspond to the difference in emergent properties.

While Olson does not explicitly explicate the concept of emergence here, I will take it to be something along the lines of *supervenience* (for Olson, it actually has to be a bit more, as I'll show later). A-properties *supervene* on B-properties if and only if, for any change in A-properties there also is a change in B-properties. In this case, the A-properties are said to be *supervenient* and the B-properties said to be the *supervenience base*. To use Olson's example, fragility is a property that supervenes on microphysical structural properties; hence two microphysically identical objects are identical with respect to fragility.

There are a few distinct types of supervenience relevant to the relationship between psychological and physical properties (here I use 'mental' and 'psychological' interchangeably). A-properties can supervene *individually* on B-properties: that is, any two individuals (or objects) with distinct A-properties must have distinct B-properties. Secondly, A-properties can supervene *globally* on B-properties: that is, any two *situations* (or possible worlds) with distinct A properties must have distinct B properties. In order for Olson's argument to be cogent, psychological properties must be *individually* supervenient only on microphysical properties; macrophysical modal or dispositional properties, as he calls them, must not be part of the individual supervenience base (B-properties) (Chalmers, 33, and Kim, 158). Hereafter I'll use 'modal properties' to mean the sort of macrophysical modal properties disallowed by Olson's argument.

It may seem that his argument will be easily resisted by the property dualist, who believes that there is no logical connection between the physical and mental properties expressed, while there is a clear logical connection between microphysical properties and fragility. However, Olson's use of "emergence" is sufficiently vague to be compatible with the view that mental properties supervene only naturally, and not logically, on physical properties: in other words, that the mental properties are linked to physical properties by contingent laws rather than logical relations. Many property dualists accept that such connections exist: David Chalmers explicitly endorses this thesis (Chalmers, 124); even Donald Davidson admits that a *form* of supervenience is compatible with his view (Davidson, 250). Of course, for those that do not, there doesn't seem to be any reason to accept Olson's claim of dependence.

Under any materialist framework, the existence of psychological properties is logically implied by the existence of certain physical or functional properties: at the very least these psychological properties *logically supervene* on physical properties. Of course, many materialists claim that the relationship between psychological and physical properties is stronger: namely, that psychological properties are identical with physical properties of some stripe. As the explication I have given of supervenience implies that the supervenience relation is reflexive, these views are captured by this weaker principle; every property supervenes on itself.

Under the property dualist framework, the existence of psychological properties is implied by the existence of physical or functional properties together with some additional psycho-physical laws: mental properties *naturally supervene* on physical properties. Olson's argument hangs on the assumption that this supervenience relation is individual supervenience on intrinsic properties. But even completely ordinary physical properties, if sufficiently complex, are *not* individually supervenient on microphysical properties! Consider the following example:

Imagine that every member of VCU's Philosophy Club was also a member of the VCU basketball team. Here we would have two objects that were microphysically identical, with the principal difference between them being their persistence conditions: the basketball team could survive a cut of funding to the philosophy club that the club could not survive, and the philosophy club could survive the dissolution of the team, whereas the team couldn't. Now, the property of scoring points is the poster child of a complex property that is dependent on microphysical properties: something physical must be different between two situations in which a team has different points values. Imagine that the basketball team had 82 points. While the philosophy club would be microphysically identical to the basketball team (by being made of exactly the same parts as the basketball team), the philosophy club would not have 82 points. So, while scoring points clearly *globally* supervenes on microphysical properties, and while the philosophy club is a system microphysically identical to the basketball team, the philosophy club does not instantiate the property of having scored 82 points, while the basketball team does. But the basketball team and the philosophy club are microphysically identical; the difference between them *in virtue of which* many of these ordinary properties apply or fail to apply is what Olson calls "modal or dispositional."

So, the modal and dispositional properties of microphysically identical objects *are* relevant to the instantiation of particularly complicated supervenient properties. In other words, complicated physical properties supervene individually on modal properties in addition to the non-modal microphysical properties. But of course, the relationship between psychological and underlying physical properties is considerably more complex than the relationship between points and the properties of basketball teams. So if Olson's claim is that psychological properties supervene individually on non-modal microphysical properties, it needs further support.

However, this does not show that psychological properties do not individually supervene only on non-modal microphysical properties, or that not everything that has the relevant physical properties has the relevant psychological properties. Right now, the defender of the two many minds argument is at an impasse; his argument is no longer entirely convincing, but neither is it decisively refuted. I will now show that there are good reasons to deny that this relationship holds between psychological and non-modal physical properties.

There are a number of ways to motivate the view that not all physically identical things are psychologically identical, many of which are specific to particular views in philosophy of mind. Shoemaker, for example, claims that a proper understanding of functionalism compels one to deny thoughts to certain objects. However, even if you do not agree with Shoemaker's functionalism, you ought to accept his conclusion.

Recall that one of the supports of the too many minds paradox was this: having complex thoughts is sufficient for personhood. This just means that everything that has thoughts is a person. But that is logically equivalent to the statement "anything that's not a person doesn't have thoughts." So, being a person is a necessary condition for having thoughts.

And if that's true, then those defending against the various incarnations of this argument have independent reasons to deny thoughts to person-stages, brains, or corpses-to-be: they aren't people. Of course, they will have to give some additional explanation of *why* they aren't people, but this explanation will surely be non-empty and vary with the account of personhood and personal identity: different views will give different reasons, but none of these reasons will be "brains don't have thoughts." Any view of personal identity will have some reason that I, and not my brain, is a person (unless, of course, according to that view I *am* my brain); even Merricks finds interesting reasons to count humans as persons and deny existence to brains, rather than vice versa. The point here is that that additional explanation, *whatever it is*, is enough reason to deny thoughts to person-stages, brains, or anything else, even if it involves modal and dispositional properties. If that explanation is lacking, it had better be lacking for reasons other than this.

Recall that Merricks defends his central second claim by challenging its denier to "offer a

replacement that would explain why atoms arranged brainwise, for some kinds of things (like me) is connected to consciousness, but not for other kinds of things (like my brain)" (Merricks, p. 51). This can easily be done by any account of personal identity: we just build in to the connection between the arrangement of the atoms and the possession of mental properties the criteria for being the unique experiencer of those properties. Establishing these criteria is closely related to the project of personal identity, and while stances on the problems in personal identity do not always imply particular criteria for personhood, they frequently narrow the candidate range. For example, perdurantist theories of personal identity rule out person stages as persons; psychological continuity theories rule out human animals[2].

The paradox of the too many minds is going after the problem of personhood and consciousness from the wrong direction. Given a set of numerically distinct objects, a set of token thoughts, in the case that all of the objects contain the systems from which the thoughts emerge, we ought not immediately assume that the thoughts belong to all of the objects. Instead, we should examine the differences between the objects and then decide which one is the person and has the psychological properties. In most cases, this can be done non-arbitrarily: recall that some of the objects can survive brain transplants whereas some can't, and some are able to look at themselves in the mirror while some of them aren't. Perhaps there are some cases in which multiple candidate thinkers are equally well suited to carry the thoughts; in these cases, it will be arbitrary which one thinks[3].

Four-Dimensionalism and Objections

Thus far I have assumed that this reply is compatible, and palatable, to all accounts of personal identity. This is not obvious: I claim that only persons, and not person-stages, have thoughts. This may seem anathematic to four-dimensionalist projects which seek to eliminate persons in favor of person stages. The idea that only persons (and not person-stages) have thoughts is no more of a threat to the reduction of persons to person-stages than the claim that only basketball teams (and not basketball players) have points is a threat to the reduction of basketball teams to basketball players. It is still clear that the team has its points *in virtue* of the properties of its players, even though the players themselves do not have any points. This view requires a modification of four-dimensionalist theories, but the modification is syntactical.

My argument does show that the relationship that holds between person-stages and makes some accumulation of stages a person, rather than some other, cannot be defined in terms of their thoughts, any more than the relationship between basketball players that bonds them into a team (let's call it the B-relation) can be defined in terms of the points of the members. Interestingly, the B-relation *could* be spelled out in terms of into which basket the players put the ball (with, of course, a few other details filled in). Similarly, the relationship between person-stages could easily be defined in terms of lower-level properties which do not have the disturbing implications of thoughts but retain the significant features.

So here's the idea: according to Olson, and according to many naïve notions of thoughts, the individual supervenience base for psychological properties does not include the modal or dispositional properties that make someone a person. I have argued that the base does include these properties, and that properly, thoughts only apply to people. I propose that the properties that link person-stages into a person supervene merely on those lower-level properties that don't imply personhood—the ones that Olson thinks make up thoughts all by themselves.

2 So, on a perdurantist view, four-dimensional maximality is a criterion for personhood, and therefore (on the current view) a criterion for having thoughts. Similarly, on a psychological view, not possibly existing without higher order brain function is a criterion for personhood, and so a criterion for having thoughts.

3 Like Noonan's response, this view solves the problem syntactically. However, this solution is superior for two reasons: firstly, it saves the intuitive definition of 'I', which Noonan ignores; and secondly, it preserves the logical implication from psychological complexity to personhood, and from personhood to psychological complexity.

This move to lower-level properties that have fewer logical connections should not be seen as a new or disturbing change in four-dimensionalist theories of persistence. Recall that person-stages cannot have all sorts of important psychological properties anyhow, such as remembering where they were three years ago, or correctly believing that they will be at work at such-and-such a time next week. Instead, many of these properties are already implicitly understood as lower level properties, even though memories are the only ones to have been explicitly redefined this way. The poster child of this kind of logically restricted property is the quasi-memory. 'Quasi-thoughts,' or properties that are explicitly defined to not imply personhood, should be no more threatening to a four-dimensionalist theory than quasi-memories, which explicitly do not imply identity. Although many refined definitions of quasi-thoughts would allow these views to bypass the too many minds paradox, the simplest would define quasi-thoughts as being psychological properties which supervene only on the non-modal properties in the supervenience base of thoughts.

We should accept this schema because it most accurately reflects our intuitions regarding personhood and consciousness, and because neither Olson nor Merricks have offered compelling argumentation to deny it. Shoemaker gives a similar claim, and motivates it via functionalism, but we can deny the problem of the too many minds even if we don't accept Shoemaker's functionalist account of the mind.

So, regardless of our beliefs regarding personal identity, survival, and personhood in general, we can and should resist the problem of too many minds by restricting the ownership of psychological properties to persons only.

Works Cited

Olson, Eric. *The Human Animal: Personal Identity Without Psychology,* Oxford (1997): Oxford University Press.

Merricks, Trenton. *Objects and Persons*, New York (2001): Oxford University Press.

Chalmers, David. *The Conscious Mind: In Search of a Fundamental Theory,* New York (1996): Oxford University Press.

Noonan, Harold. *Personal Identity,* Second Ed. New York (2003): Routledge.

Kim, Jaegwon, "Concepts of Supervenience" in *Philosophy and Phenomenological Research,* 45 #2, pp. 153-176.

Davidson, Donald, "Mental Events" in *The Nature of Mind,* Ed. by David Rosenthal New York (1991): Oxford University Press, pp. 247-256.

Is It Reasonable For 'Art' To Have No Definition?

John Fluharty

The concept 'art' has no definition. It is open to interpretation and change; what is constituted as art is based on a range of rational reasons, contextually different between individuals and situations. I will show that Wittgenstein's theories about aesthetics (different from family resemblances) were on the right path, and that Morris Weitz's open concept view was flawed, but can be revived by cluster accounts, such as Berys Gaut's. However, I will also show that Gaut was mistaken to give set criteria that constitute a highly disjunctive concept (definition) of art, but that rational reasons in an epistemic field can provide individual concepts of art, allowing for an explanation to the

vagueness in defining art. Thus, I shall begin with the anti-essentialists and move onto cluster accounts, finally finishing with a reason-based cluster account of art.

1. Anti-Essentialism and Objections

Anti-Essentialist—or most simply—the view that art cannot be defined, emerged as a response to the Essentialist Functionalists—the dominant view in the first half of the century—who hold that art is defined by most often one, but occasionally more, valuable functions that it fulfils[i]. Anti-Essentialism has reminded us that the valuable functions of art change over time. It has its roots in Wittgenstein's notion of family resemblances, and asserts that there are no necessary and jointly sufficient conditions that exclusively specify something as art. Weitz, in one of the most influential papers of the century in aesthetics[ii], claimed: 'If we actually look and see what it is that we call "art", we will find no common properties—only strands of similarities,'[1] because 'logic of the concept' precludes 'art' from having necessary and sufficient conditions.'[2] 'Aesthetic theory is a logically vain attempt to define what cannot be defined…to conceive the concept of art as closed when its very use demands its openness.'[3] Naturally, this opened up the floor to definitional responses with anti-essentialist elements such as: Institutionalists, Historical definitions, and hybrid theories.[iii] Before continuing with Weitz, though, I shall briefly look at Wittgenstein's views on aesthetics.

Wittgenstein holds that not only can aesthetic conception not be defined necessarily and sufficiently, but aesthetic appreciation cannot be described; this would demand complete description of aesthetic environment—myriad language games applicable to actual aesthetic situations. More simply: there is no reason to speak of the comprehensiveness of the language we use and further aesthetic appreciation is a sign of aesthetic understanding though the grammar of 'understanding' is spread across a intricate practice of aesthetic appreciation; no single aesthetic appreciation exhibits itself. There is no single means of aesthetic appreciation.[4] Philosophical aesthetics should be a form of grammatical investigation with our diverse interaction with art. This is threefold: '(i) drawing attention to the actual situation in which aesthetic judgements are being made (whether we agree or disagree, and about what, and why); (ii) proceeding by making comparisons between the occurrence of our aesthetic judgement and other language games as a means for drawing attention to the actual situation; (iii) trying to make one see the aspect, which is constitutive of the actual situation; that is, to notice that an interpretation is given as a description of an *experience* [my italics], and that assenting signifies, not the exchange of information, but […] finding one another (in language).'[5]

Weitz's view developed over three main essays: 'The Role of Theory in Aesthetics' (1956), 'Wittgenstein's Aesthetics' [1973], and *The Opening Mind* [1977].[6] Let me summarize his view that emerged over several years:

---

1 Pg 125, Weitz, Morris, 'Role of Theory in Aesthetics', Journal of Aesthetics and Art Criticism, vol. 15 (1956), pp. 27-35; reprinted in Philosophy Loots at the Arts: Contemporary Readings in Aesthetics, ed. Joseph Margolis, rev edn (Philadelphia: Temple U.P. 1978)

2 Pg 126, Ibid

3 Pg 122, Ibid

4 Pg 299, Gunter, Eran, 'Critical Study: An Inadvertent Nemesis—Wittgenstein and Contemporary Aesthetics' British Journal of Aesthetics Vol. 45 No. 3 July 2005

5 pg 300, Gunter, Eran, 'Critical Study: An Inadvertent Nemesis—Wittgenstein and Contemporary Aesthetics' British Journal of Aesthetics Vol. 45 No. 3 July 2005

6 Weitz, Morris, The Opening Mind (Chicago: University of Chicago Press, 1977), Weitz, Morris, 'Role of Theory in Aesthetics', Journal of Aesthetics and Art Criticism, vol. 15 (1956), pp. 27-35; reprinted in Philosophy Loots at the Arts: Contemporary Readings in Aesthetics, ed. Joseph Margolis, rev edn (Philadelphia: Temple U.P. 1978), and Weitz, Morris, 'Wittgenstein's Aesthetics' in Benjamin R. Tilghman (ed.), Language and Aesthetics, (Lawrence: The University of Kansas Press, 1973)

1) some empirically descriptive and normative concepts are open and some are closed
2) open concepts are open in a variety of logically distinguishable ways
3) open concepts enrich thought and discourse by serving conceptual functions that could not be so served if those concepts were closed
4) at least some, if not all, types of open concepts are always open
5) an open concept which is always open, such as a perennially flexible concept like 'art' or a perennially debatable concept like 'tragedy', is always open because of its function or use and its corresponding logic or logical grammar.[7]

The problem with Weitz's theory should be apparent from the start. Relying on a concept of family resemblances leads to a vacuous definition; everything does and can be made to resemble everything else in some shape or form. It follows then that simple resemblances should not be used to explain why a concept cannot be defined.[8] A further problem with Weitz's theory is that of art as 'open concept', which directly follows from having a family resemblance theory.[iv] Definitions, it is held by Weitz, apply only to closed objects; though, for example, 'family' can maintain as closed, though the members are changed , even by those unrelated by blood, i.e. adoption.

It has also been objected that Weitz misunderstands Wittgenstein, who has several different versions of family resemblance: e.g. 'language', 'chair', and 'number'. They all have an element of vagueness, but also have different employment in language. Fredrich Waismann coined the term 'open-textured concept' to refer to the recurrent dubitability appended to the conditions of application of certain concepts.[9] The similarity between Waismann and Wittgenstein is that anticipation of the application of certain concepts is impossible, and the conditions of application are uncertain. Wittgenstein, though, believes there are other ways for a concept to be open: (i) there are cases in which there are no rules of application, (ii) there are cases impossible to anticipate application, (iii) the concept of games, such as the one Weitz uses, , involves vague boundaries or blurred edges. To connect this with Weitz; Weitz does not distinguish the various kinds of open concepts. Art is open textured in all three senses, though Weitz compares art to 'games', which is open in only one sense.[10] Thus, an anti-essentialist definition should explicitly state which open-concept 'art' falls under, and why.

Furthermore points (4) and (5), above, give Weitz the most trouble, and indeed, he fails to answer questions such as 'What are concepts?' And 'What is it to have concepts?'.[11] Weitz uses the terms 'logic' or 'logical grammar', to imply a 'use-pattern'—or corresponding pattern of use—of open concepts. He is wrong to imply that use-patterns guarantee a concept is always open; functions and use patterns change over time.[12] Unfortunately, Weitz gives little support for this assumption. The ability of the concept 'art' to change over time needs to be explored, as well as the prospect of any unchanging features. Without, *some* sort of evidence of unity or a 'deep-seeded' structure in the concepts of artwork, open concept theories of art look vacuous and susceptible to collapse.

Following from this discussion, four clear objections to Weitz emerge: (i) family resemblances can lead to a vacuous account of art, (ii) closed definitions can take on new defining characteristics, (iii) Weitz is ambiguous as to which 'open concept' he supports, and (iv) he fails to explain what a concept is or what it means to have one.

---

7 Pg 38, Kamber, Richard, "Weitz Reconsidered: A Clearer View of Why Theories of Art Fail", British Journal of Aesthetics, Vol 38, No. 1. January 1998

8 This and other objections can be found in: Davies, Stephen, Definitions of Art (Ithica, N.Y.: Cornell University Press, 1991), Chapter 1.

9 Waismann, Fredrich, "Verifiability," The Aristotelian Society for the Study of Philosophy, Supp. Vol. 19 (July 1945): 119-50

10 pp 2-9 Scalafani, Richard, ''Art', Wittgenstein, and Open-Textured Concepts', The Journal of Aesthetics and Art Criticism, Vol. 29, No. 3 (Spring 1971)

11 pg, 39 ramber, Richard, "Weitz Reconsidered: A Clearer View of Why Theories of Art Fail", British Journal of Aesthetics, Vol 38, No. 1. January 1998

12 Pg 39, ibid

2. Reviving Anti-Essentialism and Cluster Accounts of Art

To maintain an anti-essentialist theory these objections must be answered. Richard Kamber asserts that it is correct to be sceptical about the possibility of a theory to reveal the continued unity of art over time; the mistake comes in assuming there is an essential openness or disunity in concepts of art. Weitz's scepticism is shared, but not his analysis. Kamber's analysis is threefold: first, art is an umbrella concept such as contemporary users cannot agree on a necessary condition for something being an artwork; second, if contemporaries did show a necessary conditions for something's being an artwork, it would need to show the property is a result of deep structure; finally, no theory has made this case yet. There is no deep structure connection.[13]

Kamber reinvents Weitz by throwing out the open concept solution and replacing it with an umbrella concept, which simply claims that the concept of art keeps encompassing new dimensions. This would seem to help objection on (ii), though not the others. Berys Gaut, however, has identified a necessary condition for art, and handles objections (i)-(iv) well, while still maintaining there is no definition of art.

Gaut argues that art is not a resemblance-to-paradigm construal (something is art by virtue of resembling paradigm art-works) such as Weitz's theory suggests, but rather a cluster construal to family resemblance that gives correct characterization of art. The argument rests on counterfactual cases of supposed art objects rather than the importance of originality in art. Wittgenstein indeed developed a cluster account, from family resemblances, of the meaning of proper names. Weitz's resemblance-to-paradigm model leads to his vacuous account; the cluster account avoids this by stating criteria.[14] This route, also would be supported by Wittgenstein's direct views on aesthetics, stated above—in which he holds that there is no single means of aesthetic appreciation, but aesthetics should form an investigation of interaction with art—as well as supported by Kamber's umbrella concept.

A cluster account 'is true of a concept just in case there are properties whose instantiation by an object counts as a matter of conceptual necessity toward its falling under the concept. The properties, called *criteria*, are simply the possession of a property which is a necessity of an object's being a concept.[15] Gaut gives several reasons why a criterion counts towards a concept:

1) if all of the properties that are criteria are instantiated, this suffices for an object to fall under the concept; and more strongly, if fewer than all of these properties are instantiated, this also suffices for the application of the concept. So there are jointly sufficient conditions for the application of the concept.
2) there are no properties that are individually necessary conditions for the object to fall under the concept (that is, there is no property that all objects falling under the concept must possess)
3) there are disjunctively necessary conditions for application of the concept. By the second point, it follows that if a concept's meaning is given by a cluster account, one cannot define that concept, in the sense of fixing *individually necessary* and jointly sufficient conditions for it.[16]

Indeed, there is a logical difference separating the resemblance-to-paradigm accounts from clusters; cluster accounts appeal to general properties to explain the relevant features, resemblance-to-paradigm accounts explain them by resemblance to particulars.

Gaut appeals to Wittgenstein to *define* the contents of his cluster account. "Don't think, but

13 pg 44-45 ibid

14 pg 275, Gaut, Berys, "The Cluster Account of Art Defended", British Journal of Aesthetics, 45, 2005, pp. 273-88.

15 pg 273, Gaut, Berys, " The Cluster Account of Art Defended", British Journal of Aesthetics, 45, 2005, pp. 273-88.

16 Pg 27, Gaut, Berys, "'Art' as a Cluster Concept", in Theories of Art, Noël Carroll (ed.), University of Wisconsin Press, 2000, pp. 25-44.

look!"[17] This is simply an appeal to look at how art is used in language, to which, Gaut concludes that art must challenge the intellect rather than being merely just for pleasure. Gaut then gives us a list to help *define* what he think should count towards an objects being art: '(i) possessing positive aesthetic qualities (I employ the notion of positive aesthetic qualities here in a narrow sense, comprising beauty and its subspecies); (ii) being expressive of emotion; (iii) being intellectually challenging; (iv) being formally complex and coherent; (v) having a capacity to convey complex meanings; (vi) exhibiting an individual point of view; (vii) being an exercise of creative imagination; (viii) being an artefact or performance that is the product of a high degree of skill; (ix) belonging to an established artistic form; and (x) being the product of an intention to make a work of art.'[18] Having a list, as such, yields a concept that is noncircular.

Furthermore, Gaut gives a necessary condition for something's being an artwork: action. Each one is a product of action, and importantly, selection (such as found art) is a form of action. Selection as a work of art adds to properties of something, or changes them, 'a piece of driftwood in nature cannot express despair,'[19] but selection can give it that quality[v]. This identifies that deep structure of art, to which Kamber asserts as important. I shall now turn to some of the objections that have been made by Thomas Adajian, and Gaut's responses, but concentrate more on Robert Stecker's claim that Gaut's cluster theory is a disjunctive definition of art, incognito. I shall then bring up some objections of my own.

Adajian objects that Gaut's cluster account is—because it appeals to Wittgensteinian positions of family resemblance—not any better than definitions of art. Moreover, by leaving his cluster account open to revision, Gaut appears to contradict himself by denouncing definitions, while holding a concept that is a disjunctive definition. In parallel, simply because a definition has not been pinned down, does not mean there is not one, however it can be responded that using a cluster theory sidesteps many of the problems for definitions. Gaut concedes that the epistemic version of the argument fails, though a heuristic one holds, based on reasonable search principles. Stecker argues that Gaut's cluster account is really just a disjunctive definition in disguise:

> If the concept of F is a cluster concept, then there are several different sufficient conditions for being an F, no conditions are individually necessary for being an F, that is, there is no condition that all the Fs must satisfy, and finally, there are disjunctively necessary conditions for being an F, that is, it must be true that if something is F, then it satisfies one or another of the sufficient conditions for being F.[20]

However, for an F to have these characteristics, it would be true of F that it is disjunctively defined. What makes Gaut's version distinct from other disjunctive definitions, is that, whether something is a sufficient condition to qualify as art, is indeterminate. Further, there are no set numbers of disjunctive conditions. One response is to affirm that his definition is not conjunctive, and draw a distinction between highly disjunctive and variegated definitions and simple disjunctive and conjunctive definitions, and we find that there is not a correct definition of art that is simple disjunctive. Thus simple disjunctive and conjunctive definitions are defective. Highly disjunctive theories have many disjuncts.[21] What must be explored now, is the candidate account.

Let us explore the ten criteria that Gaut proposes in the candidate account for an object to be art. Stating specified criteria that an object—which is considered to change in form over time—must fulfil to constitute it as that thing is flawed. Many things that are not art, with simple disjunctive definitions can fulfil Gaut's criteria: the games of chess, for example, would fulfil i-viii; that is, it is beautiful, graceful and elegant to watch (or play); it can express emotion for the players; is intellectually challenging; complex; each move or strategy has meaning; it exhibits an individual

17 Wittgenstein, Philosophical Investigations, part 1, 66

18 pg 274, Gaut, Berys, "The Cluster Account of Art Defended", British Journal of Aesthetics, 45, 2005, pp. 273-88.

19 Pg 29, Gaut, Berys, "'Art' as a Cluster Concept", in Theories of Art, Noël Carroll (ed.), University of Wisconsin Press, 2000, pp. 25-44.

20 Pg 48, Stecker, Robert, 'Is it Reasonable to Attempt to Define Art?', in Theories of Art, Noël Carroll (ed.), University of Wisconsin Press, 2000

21 Pg 286, Gaut, Berys, "The Cluster Account of Art Defended", British Journal of Aesthetics, 45, 2005

point of view; exercises creative imagination; and it is part of a performance. A sexual experience could fulfil most criteria as well, especially an early sexual experience which could prove intellectually challenging. The only two chess does not fulfil are (ix) and (x), specifically (ix) which begs the question: *What are the established art forms and who decided they are established?*[vi] Furthermore, (x) being a product of intention to make a work of art undermines a subcategory of Gaut's necessity of *action*: selection. Selecting something in nature as art, does not constitute the making of a work of art, and it certainly was not intentional for nature to make it as art. Further, *action*, seems unimportantly necessary; action denotes that *something* is being created (or selected) but does not give a defining characteristic. Moreover, note that (ii-vi) all are forms of communication, something I will come back to.

Gaut's theory allows for some of the criteria to be objected; not all criteria must be filled to constitute something as art. The criteria in the cluster are necessary disjunctive. However, the example of the chess game or sexual experience shows that lists can be flawed. Set criteria, strictly necessary, or loosely necessary, do not work for constituting an open concept. Moreover, is disjunctive necessity rationally possible? Necessity involves a condition an object *must* meet to be considered that object, though a set list of criteria that art must necessarily fulfil some (different ones for each art object) to be considered art, begs the question: *Are any of them really necessary?* We can formulate this discussion into a few clear objections to the cluster account: (I) More concepts than art fulfil a substantial number of criteria; (II) (ix) begs the question and (x) self-contradicts the necessity of *action*; (III) are highly disjunctive definitions based on a set of criteria rationally possible?

3. Reason-based Clusters

I suspect the cluster account can be saved, though, without a set list. Gaut states that certain qualities give reason to constitute an object as art. Can *reason* be exploited to give support to a cluster account? I think it can. Drawing on ideas about rational reasons in relation to a subject, or actor, a contextual and subjective concept of art can be built up to help prevent Gaut's cluster account from collapsing without a set list. To begin, we can distinguish three types of reasons: practical, epistemic, and evaluative reasons[vii].

- He has reason to feel proud of himself: despite all the pressure on him he won the match.
- That electrician has failed to turn up again! Yes, you have some reason to be annoyed with him.
- Freda has good reason to be resentful about the way she was treated[22]

I shall focus mainly on the epistemic reason, which can be applied to theories of art:

- I have reason to think Frank Gehry's Guggenheim is art: it is beautiful, graceful, and elegant.

Reasons can be seen as a fact in a 'non-committal, formal or nominal way, in which facts can be simply equated with true propositions, propositions being understood as information-contents, Fregean thoughts'[23] Actions, beliefs, and feelings are intentionally linked to their accountability as reasons. Moreover, reasons are *facts* that stand in relation to an actor, and reasons can be strong or weak based on the number of *facts*:

> Thus R, the reason relation we're considering, holds between a plurality of facts, an actor, an act-type, a degree of strength of the reason, and a time:
> The facts *pi* are at time *t* a reason of degree *d* for *x* to φ.
> R(pi, t, d, x, φ)[24]

22 pg 1 Skorupski, John, 'The Unity and Diversity of Reasons' Unpublished

23 pg 2 Skorupski, John, 'The Unity and Diversity of Reasons' Unpublished

24 pg 3 Skorupski, John, 'The Unity and Diversity of Reasons' Unpublished

Therefore:

R (The facts that the Guggenheim is beautiful, graceful, and elegant are at this time a reason of good degree for me to believe it is art.)

Before we move on, I must clarify the concept of degree of strength. Reasons can be of varying degrees of strength. This is certainly one of the problems that plagues concepts of art. Definitional theories struggle because of the vagueness of some items to be considered art. Having reason that is context dependent, yields an explanation to this vagueness. It explains, rather easily, how something is called art at one time by one person and not at another time by another person. This can be refined by pointing out that most if not *all* knowledge of what art is is *a posteriori*. It comes from experiencing how the term is used: "Look, don't think!". One further element of reasons I would like to explore are epistemic fields. Epistemic reasons are relative to a field which contain facts, the strength for the belief depends on other facts in the field. The field can be enlarged as more information becomes salient; likewise, facts can be overwritten as new ones become salient.[25] Importantly, we need a notion of rationality to prevent a vacuous theory. If we have a reason to believe something is art—that is not forwardly rational—than perhaps that can allow anyone to dub anything as art without good reason. Thus, we can introduce a concept of rationally self-determining actors, who can assess reasons to believe or feel by their own reflection. They decided whether they should do more investigation, or have sufficient reason to believe. This is called self-audit. There are warrantable reasons, open to self-audit obtained by reflection. These warrantable reasons are held within the epistemic field, called the epistemic state. 'The fact that *p* is in x's epistemic *state* at *t* if and only if[26]:

i) *x* could come to believe, at *t*, that *p*, simply by attention to the fact that *p* without any further action, and

ii) in an epistemic field of *x*'s that contains the fact that *p* that very fact is sufficient reason for *x* to believe that *p*, whatever other facts obtain in the field'[27]

Thus, we can see that experience coupled with rational thinking can lead to good epistemic reasons to believe some set of relations to an actor. I believe this can be applied to a cluster theory to yield an anti-essentialist definition of art able to side step all of the objections yet raised. Let me try to put this theory into a set of premises:

i) There are no necessary and sufficient conditions that specify something as art

However

ii) Art used as cluster account such that there are properties whose instantiation by an object counts as a matter of conceptual necessity (set by individual warranted rationality) towards its falling under the concept.

iii) There is no universal list of properties

iv) There are, instead, individual clusters set by epistemic reasons subjective to an individual

v) Reasons are factual propositions that stand in relation to an actor and vary in degrees of strength

vi) Factual propositions come through experience

vii) Epistemic reasons about what constitutes something as art are relative to a field of facts obtained through experience, which can expand or be self-audited by rational deliberation

Therefore,

The concept of art has no objective necessary and sufficient conditions, but rather, is made up of a cluster of individually necessary subjective properties set by *a posteriori* facts in

25 pg 6, Skorupski, John, 'The Unity and Diversity of Reasons' Unpublished

26 All these ideas on reason originate from Skorupski, John, 'The Unity and Diversity of Reasons' Unpublished

27 Pg 9, Skorupski, John, 'The Unity and Diversity of Reasons' Unpublished

relation to that subject, from which, rational deliberation give the subject warranted rational reason to constitute something as art.

Art, then, is instantiated as anything a rational agent believes it to be, in so far as he has rational reason to believe it. This proves a contextual definition, and is faithful to other anti-essentialist positions. It changes over time; holds true to the three types of open concept based on the reasons we use them in language; the epistemic field of reason yields a cluster of facts with which to constitute an object as art. It also explains indeterminacy of disjunction by showing different subjects have a different field of epistemic reasons to constitute an object as art; moreover it is not a highly disjunctive because there is no set list of instantiation. Furthermore, the objections raised against Weitz's original theory are satisfied by this definition where open concepts are based on experience in art's use in language, rationalized by warrant. Moreover, it satisfies the epistemic challenge against the cluster account, as well by avoiding a set list of criteria, and the disjuncts.

4. Conclusion

The majority of Gaut's cluster criteria were based on communication, as was Wittgenstein's account of aesthetics based on grammatical investigation *experience* of interaction with art. Basing rational reasons on the interaction of the construct of 'art' in language seems the *rational* course. It fully covers all the set objections brought against the anti-essentialists by taking them head on, or side stepping then. Thus, due to the vagaries of experience and use of 'art' we find there are no necessary and sufficient conditions to define it. 'Art' is open to take on new meaning. Therefore, there is no definition of art.

---

i. Modern Functionalist definitions include Beardsly, Monroe, "Redefining Art" The Aesthetic Point of View: Selected Essays, ed. Michael J. Wreen and Donald M. Callen (Ithica N.y.: Cornell University Press, 1982): Hanfling, Oswald, "Art Artifact and Function" Philosophical Investigations 18 (1995) 31-48: Rowe, M.W. "The Definition of 'Art,'" Philosophical Quarterly 41 (1991) 271-86.

ii. Weitz, Morris, 'Role of Theory in Aesthetics', Journal of Aesthetics and Art Criticism, vol. 15 (1956), pp. 27-35; reprinted in Philosophy Loots at the Arts: Contemporary Readings in Aesthetics, ed. Joseph Margolis, rev edn (Philadelphia: Temple U.P. 1978).

iii. Institutionalists who try to avoid anything functional, and define it by the way it (art) attains its art status include: Dickie, George, Art and the Aesthetic: An Institutional Analysis (Ithaca, N.Y.: Cornell University Press, 1974); and Diffey, T.J., "The Republic of Art" The Republic of Art and Other Essays (New York: Peter Lang, 1991). Historical definitions identify relevant similarities and trace them back to 'first art', these include: Levinson, Jerrold, "Defining Art historically" Music, Art, and Metaphysics (Ithaca, N.Y.: Cornell University Press, 1990); Carney, James , "The Style Theory of Art" Pacifica Philosophical Quarterly 72 (1991): 272-89. Hybrid definitions attempt to define art with some sort of function without pinning the functions down, these include: Stecker, Robert, Artworks: Definition, Meaning, Value (University Park: Pennsylvania State University Press, 1997).

iv. Many of these objections originally come from Maurice Mandelbaum: Mandelbaum, Maurice, 'Family Resemblances and Generalizations Concerning the Arts', American Philosophical Quarterly, vol, 2 (1965).

v. Gaut goes on to give methodological considerations, that are not immediately salient to this essay. An account of the concept should be:
(1) 'adequate to intuition: it must agree with out intuitions about what we would say about actual and counterfactual cases: if the account claims that some object satisfies the concept, but it intuitively doesn't (or vice versa), then that is one strike against the account.
(2) normatively adequate: some intuitions that do not fit the proposed account may be rejected: there will be a reflective equilibrium between the account and intuitions[…]it must include a theory of error: some account must be offered of why people have the mistaken intuitions they do.
and should have
(3) heuristic utility: it should be such as to figure in true or at least promising theories about the object to which the concept applies.'

These quotes were taken from:

Pg 30-31 Gaut, Berys, "'Art' as a Cluster Concept", in Theories of Art, Noël Carroll (ed.), University of Wisconsin Press, 2000, pp. 25-44.

vi. For example, Roger Scruton argues against Photography (which many would take to be an established art form) as an art form in: Roger Scruton, "Why Photography is Not Art," in Golblatt and Brown, op. cit., p. 90

vii. All explanations of reason in this section are drawn from: Skorupski, John, 'The Unity and Diversity of Reasons' Unpublished. They can applied to concepts of art to explain many of the problems with anti-essentialists.

Interview with Simon Prosser

(Conducted on 3rd June 2008. Interviewer: Joe Slater).

Joe Slater: It's common knowledge among your students that you were originally a physicist. At what point did you "see the light," as it were, and come to philosophy?

Simon Prosser: Not while I was still doing physics. I often get asked how I started off in physics and got into philosophy. It's not something uncommon, actually. There are quite a few philosophers who did either physics or mathematics at first. I think it's because in the case of people doing physics they're often interested in the big questions about space and time and the universe and so on. Certainly what happened in my case was just that I mis-located the things that I was interested in. I thought physicists were the people who addressed the really big questions about those topics, and it took a long time to realise that actually what physicists do is mainly mathematics; figuring out things that are specific to the actual world. So I did my physics degree feeling that something wasn't right, but I didn't know what. I even started a Masters degree and had a conditional offer of a PhD place before I finally had to accept that physics wasn't really the thing for me, but I didn't know what I did want to do at that time. I had done just one very short philosophy of science course during my physics degree, so I had just a little idea about what philosophy was. It seemed to come very naturally to me, but it took a couple of years after finishing with physics before I finally decided that I should come and study philosophy properly. Because things hadn't really worked out with the physics I was very cautious about it, so I did a one-year conversion course, a postgraduate diploma. I went into it thinking "well, just be very careful and see how this goes." But I very quickly realised that I was very, very keen.

J: You've been a member of staff here at St Andrews since 2002?

S: That's right.

J: What attracted you to St Andrews in the first place?

S: Well, to be honest, the job situation in academia generally and certainly in philosophy is such that you often don't have much of a choice, especially at first, so if I'm completely honest what brought me here initially was the fact that they offered me a job. Every time a job is advertised loads and loads of people apply. But I was very happy to get a job here because it's an exceptionally strong department. In surveys that people respect it's usually rated in the top two or three departments in the UK. And I always wanted to live by the sea. I'm very happy about that too.

J: Are you likely to still be here in a few years time, or are you planning to move on...?

S: No plans to move anywhere at the moment. I think I like the place more and more as time goes

on.

J: At the moment, you're working on Zeno objects, two dimensional content of consciousness and complex demonstratives, indexicals and immunity to error.

S: Well, those are some things I've written about in the past. My main research interest is in philosophy of mind and then after that metaphysics. After that some areas of philosophy of language insofar as they overlap with mind and metaphysics. What I'm going to be working on over the next period is a number of issues in philosophy of perception and some things that overlap with philosophy of perception and philosophy of time. I'm provisionally planning to write a book on time and perception, but before I do that I've got a number of articles that I want to finish, mostly on issues to do with perception but also some other issues.

J: You regularly make references to other current philosophers in your articles, such as in your article about the two-dimensional content of consciousness, you refer to Chalmers, Dretske and Tye. To what extent is there a dialogue among philosophers when writing such articles? Do you, for example, have David Chalmers' number in your phone book?

S: I don't think I have his phone number, but I've got his email address! For that particular article Chalmers did very kindly read a draft for me. It's a general practice that people send drafts of articles to people who might be interested or people who might have helpful things to say. Most of what gets published has already been read by quite a lot of different people before it even gets sent to the journal. It's very useful because people can point out errors that you don't see yourself. We also meet each other and discuss our work at seminars and conferences. You can't make anywhere near as much progress just working in complete isolation as you do with help from other people.

J: Are there any philosophers today who you particularly agree with?

S: That's a difficult one. Let's think...possibly the person I find myself agreeing with most often is Daniel Dennett. That doesn't mean that I agree with everything he says, but I tend to be sympathetic to his views.

J: On a similar note, who would you say has had the greatest philosophical influence on you?

S: That's a really difficult one. It's very hard to pick out one person. I suppose when I was writing my PhD, Gareth Evans was a big figure and then Dennett... There are probably lots of them. Well, Jerry Fodor, John Perry, David Kaplan, Fred Dretske... There are probably others that I'm not thinking of. More recently Robert Stalnaker's work made me rethink a few things.

J: Is there any advice you would give to any students considering or aiming towards a career in philosophy?

S: I'd say they should understand that it's extremely competitive. But, on the other hand, if you really feel that it's what you want to do, if you're passionate about it and if you're willing to make sacrifices, then be very stubborn, determined and persistent. You may have lots of setbacks, and it can take a lot of determination, but if you're persistent enough you may get there. Funding for graduate studies is very competitive, and there are nowhere near enough jobs to go around and a lot of very bright people with PhDs chasing them. And when you get a job you have to work hard and it can sometimes be stressful. But most of us in this profession really love what we do – I'd certainly never want to do anything else.

J: In the current philosophical climate, there appears to be an increasing amount of specialisation. No one today can be so successful across the board as Hume, or Kant, or even Russell, so do you

think in the future, it will reach the extent, where philosophers have to have very precise niches of specialisation or areas of expertise to make any contribution to philosophy?

S: It's hard to say. I hope it won't get much more specialised than it is now, because one of the nice things about philosophy at the moment is that even though people do specialise there is some scope for doing something quite different. For example, I've toyed with the idea of trying to write something in aesthetics. There are people who maybe work on metaphysics and philosophy of mind and then have managed to write something on ethics. It's certainly not possible to keep up with the literature on everything. There's just too much getting published. But maybe we won't be able to specialise too much more because of the fact that issues in philosophy do tend to be fairly interconnected. Sometimes you have to know a bit about one area to make progress in another area. But predicting the future is very difficult.

J: In lectures, you quite often refer to sci-fi series. *Star Trek* and *Red Dwarf* spring to mind. It does seem an interesting link between physics and philosophy. Do you think that the sci-fi genre particularly lends itself to being used in examples in philosophy?

S: I think it does. Yes, a lot of philosophers have used thought-experiments that have involved sci-fi, and in many different areas of philosophy. For example Derek Parfit has used sci-fi examples to do with teleportation in certain areas of moral philosophy. I think that sometimes the people who write sci-fi series must have been studying philosophy. I'm pretty sure that whoever wrote the scripts for *Star Trek* must know a bit of philosophy because you can sometimes identify specific philosophical issues and maybe issues from cognitive science as well. The older series sometimes seemed to be influenced by existentialism more than analytic philosophy, although there was one episode where they used the liar paradox to disarm a robot. But yes, sci-fi does lend itself to philosophical examples, I find. It's all about imagining far-off possibilities and thinking about what would happen.

J: What is your favourite sci-fi series?

S: At the moment I'd have to say *Battlestar Galactica.* Not the old one from the 1970s, but the new one.

J: Having experience in both physics and philosophy, what do you think should be the role of philosophy with regard to physics, or science in general?

S: Well, I suppose to clarify what it is that physicists and other scientists do and also to clarify the interpretation of their theories. There has tended to be this phenomenon with quantum mechanics in particular, that physicists have given interpretations of quantum mechanics that are really philosophical interpretations. What the physicists can do is figure out the theories and figure out the mathematical structure that gives the right predictions, but then, what that really tells us about the world is really a philosophical question. Philosophy also has other roles, such as in relation to ethical issues that arise from certain scientific investigations or discoveries.

J: I noticed on your webpage that you're a keen photographer.

S: That's right.

J: Do you think that helps out in any way with your philosophising? Does it inspire you?

S: Well, it's more something I do to give my brain a rest and try to do something more artistic. But I have to admit that I am feeling a certain temptation to try to write something on photographic aesthetics, because there is a literature of certain issues to do with photography. I haven't really read

very far into it yet, but it's possible I might try to do something like that at some point. Mostly photography satisfies a different need I think. There are some issues of overlap, because I work on philosophy of perception quite a lot and sometimes when you're doing photography it forces you to think about how the camera is representing the world. So perhaps it can help a little bit in thinking about how the mind is representing the world.

J: One last question. As was asked to Marcus Rossberg in *Aporia's* first issue, what is your favourite bar in St Andrews.?

S: Favourite bar? There are so many to choose from! (*Deliberates for a long time*). I suppose it's Drouthy Neebors probably, though I preferred it before they refurbished it, but there are several others that I like.

Can Modal Agnosticism Save Constructive Empiricism?

Martin Beek

Supervised by Dr. Katherine Hawley
14 May 2007

ABSTRACT

In this paper I argue that by adopting modal agnosticism, the constructive empiricist can overcome the scientific realist's main objection. After introducing Bas van Fraassen's constructive empiricism and showing how he can respond to three traditional objections to his view, I consider James Ladyman's recent critique. Ladyman correctly argues that the constructive empiricist needs to distinguish between the observable and unobservable in a non-arbitrary manner. To be able to do so, the constructive empiricist must recognise objective modality in nature, but doing so would be at odds with the position's principle motivation of doing away with inflationary metaphysics and objective modality. I next explain John Diver's modal agnosticism. I argue that the modal agnostic has the resources available for the constructive empiricist to be able to make the distinction Ladyman requires. Since modal agnosticism does not entail an inflationary metaphysics, I argue that it is compatible with, and can thus save, constructive empiricism from Ladyman's challenge.

1. INTRODUCTION

Scientific realists believe that our best, currently accepted scientific theories are approximately true. In his seminal work *The Scientific Image*, Bas van Fraassen criticises the realist for the 'inflationary metaphysics' the position entails and offers in its place a constructive empiricist account of science which aims to do without the latter. "To be an empiricist" argues van Fraassen, "is to withhold belief in anything that goes beyond the actual, observable phenomena, and to recognise no objective modality in nature" (1980, 202).

As I shall explain in Section 2, van Fraassen is able to respond to most of the traditional problems raised against constructive empiricism. In Section 3, I consider James Ladyman's recent critique (2000, 2004). Ladyman argues that because constructive empiricism recommends belief in the empirical adequacy of theories rather than in their truth, the constructive empiricist *must* recognise objective modality in nature in order to allow for a non-arbitrary distinction to be drawn between the observable and the unobservable.

Van Fraassen responds by maintaining that the constructive empiricist *can* circumscribe the observable in a principled manner without appeal to objective modality. Additionally, he argues that if constructive empiricism does in fact need objective modality, being a modal realist would not be incompatible with constructive empiricism. I argue that Van Fraassen's first response is insufficient to counter Ladyman's criticism, and that his second is completely at odds with the empiricist's motivation of doing away with 'inflationary metaphysics'. Thus, as Ladyman concludes, constructive empiricism seems to be an 'untenable' position (2000, 855).

The main aim of this paper is to suggest a way in which the constructive empiricist can avoid Ladyman's objection. After outlining the possible world debate in Section 4, I explain modal agnosticism as developed by John Divers (2004) in Section 5. The modal agnostic, who holds herself as having no warrant for believing in the existence of any possible worlds other than the actual world, aims to secure at least some of the benefits associated with David Lewis' modal realism while avoiding the costly ontology. The modal agnostic does however retain the ability to express, among other things, counterfactual claims.

In Section 6 I argue that by retaining licence to assert counterfactual conditionals, the modal agnostic is able to objectively evaluate observability counterfactuals. This is exactly what the constructive empiricist must be able do in order to draw a principled distinction between the observable and the unobservable. Unlike realism however, modal agnosticism is compatible with the empiricist programme in that it does not entail an inflationary possible world metaphysics. Consequently, after examining a couple of possible objections, I conclude by suggesting that if the constructive empiricist adopts an agnostic view of modality, she may be able to save herself from Ladyman's criticism.

1.1 AN IMPORTANT ASSUMPTION

Before proceeding, I need to clarify an assumption I will hold throughout this paper. A commonly held belief is that only modal realism can analyse modalities *objectively*[1]. My understanding is that the constructive empiricist's *primary* motivation is to do 'without *inflationary metaphysics*' (van Fraassen 1980, 73). Since modal realism entails concrete possible worlds, I believe van Fraassen considers modal realism to be metaphysically inflationary. I assume that this is why van Fraassen consequently makes it part of the empiricist project to 'recognise no objective modality'. As will hopefully become clear, the modal agnostic can analyse many modal notions objectively without entailing an inflationary metaphysics. Consequently, I do not take modal agnosticism to be incompatible with constructive empiricism on the grounds that it can analyse modalities objectively, in spite of what van Fraassen says.

1 See in particular the concluding sections in Ladyman (2000) and (2004), along with Monton and van Fraassen (2003).

2. CONSTRUCTIVE EMPIRICISM

In *The Scientific Image*, van Fraassen criticises scientific realism (SR), which he defines as the position holding that:

> (SR) *Science aims to give us, in its theories, a literally true story of what the world is like; and acceptance of a scientific theory involves the belief that it is true.* (1980, 8, his italics).

In its place, van Fraassen propounds his own antirealist view of the aim of science and of theory acceptance which he calls constructive empiricism (CE). Constructive empiricism is that view that:

> (CE) *Science aims to give us theories which are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate.* (1980, 12, his italics).

The key difference between scientific realism and van Fraassen's constructive empiricism is the difference between literal truth and empirical adequacy. According to van Fraassen's 'preliminary explication', a theory is empirically adequate if "what is says about the observable things and events in this world is true – exactly if it 'saves the phenomena' " (*Ibid*, 12). More precisely, the theory is empirically adequate if it has at least one model that *all* the actual phenomena fit inside. Here, 'all' for van Fraassen is not exhausted by the phenomena already observed, or even those observed at a certain time, past, present or future, but by those which are observ*able*.

2.1 THREE COMMON CRITICISMS OF CONSTRUCTIVE EMPIRICISM

Critics of constructive empiricism have concentrated their attacks on three main points. Following Ladyman (2000, 840), these can be summarised as follows:

i) *Constructive empiricism grants ontological significance to an arbitrary distinction* because the line demarcating the observable from the unobservable is vague, prone to change over time and is an artefact of accidents of human physiology.

ii) *Constructive empiricism is incoherent* because the constructive empiricist accepts that the observable world is described using terms that refer to unobservables and also accepts that all language is theory laden to some extent.

iii) *The constructive empiricist is an arbitrarily selective sceptic* because all present data underdetermines which theory is empirically adequate just as much as it underdetermines which theory is true. Therefore, constructive empiricism is just as vulnerable to scepticism as scientific realism is, and the underdetermination of theory by evidence does not entail support for constructive empiricism as van Fraassen argues.

Though these arguments initially seem problematic for van Fraassen's position, the constructive empiricist can respond to each.

2.1.1 OBSERVABLES AND UNOBSERVABLES

The criticism that the constructive empiricist grants *ontological* significance to the arbitrary distinction between the observable and unobservable 'misses its mark' once it is recognised that van Fraassen's claim is an epistemological rather than metaphysical one (Ladyman 2000, 840). Van Fraassen explicitly states that "[…] observability has nothing to do with existence […] [it] is, indeed, too anthropocentric for that […]" (1980, 19). Constructive empiricism never claims that unobservables *don't* exist.

Van Fraassen argues that it is legitimate to attribute an *epistemological* significance to the observable/unobservable distinction. He recognises that the question of where to draw the sharp line demarcating observables from unobservables cannot be defined in a non-arbitrary manner. What follows from this is that 'observable' is a vague predicate. Van Fraassen argues that "[…] predicates in natural language are almost all vague, and there is no problem in their use; only in formulating the logic that governs them" (1980, 16).

Though meant only as a 'rough guide' to avoid fallacies, van Fraassen characterises what counts as observable as: "X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it" (1980:16). What we can or cannot observe is a direct consequence of the fact that:

> The human organism is, from the point of view of physics, a certain kind of measuring apparatus. As such it has certain inherent limitations – which will be described in detail in the final physics and biology. It is these limitations to which the 'able' in 'observable' refers – our limitations, *qua* human beings (1980: 17)

For example, the moons of Jupiter which we observe when we look through a telescope are a clear case of observation since our best current scientific theories say that *were* someone to get close enough to them, then they *would* observe them. On the contrary, our current theories do not tell us that we can directly observe particles in cloud chambers. Van Fraassen draws an analogy with a jet's vapour trail (1980, 17). When we observe a vapour trail, we do not observe the jet directly but rather *detect* it. If our current theories are correct in saying that positrons exist for example, then we detect them by means of observing their tracks in a cloud chamber. However, since we can *never* directly experience subatomic particles (as we can jets), empirically equivalent but incompatible rival theories which deny the existence of positrons are always a possibility. So, concludes van Fraassen, it is legitimate to attribute an epistemological significance to the observable/unobservable distinction.

2.1.2 THEORETICAL LANGUAGE

The claim that the constructive empiricist is incoherent because he accepts that the observable world is described using terms that refer to unobservables and that all language is theory laden to some extent, is a criticism which van Fraassen pre-empts early on in *The Scientific Image*. Accepting a theory, he says, "clearly involves more than belief" in the theory's empirical adequacy (1980, 12). This is so because a scientist is never confronted with a complete theory. Consequently, in accepting an incomplete theory the scientist involves himself in a research programme – one that could have been very different if he had accepted another empirically adequate, but rival, theory. For non-scientists, acceptance still involves a certain commitment to "confront any future phenomena by means of the conceptual resources of [the] theory". "Thus", concludes van Fraassen,

"acceptance involves not only belief but a certain commitment" (1980, 12).

Consequently, van Fraassen concedes to the realist that the constructive empiricist often has to use the language of science understood literally. In fact, sometimes there is 'no other way' to describe the world and the objects in it, such as microwave ovens or VHF receivers. But, he says:

> From this it does not follow that I believe that the concept of very high frequency electromagnetic waves corresponds to an individually identifiable element of reality. Concepts involve theories and are inconceivable without them […] [b]ut immersion in the theoretical world-picture does not preclude 'bracketing' its ontological implications (*Ibid*, 81).

Immersing oneself in the world described by scientific theories is thus necessary, but, as Ladyman notes, van Fraassen maintains that this only ever provides *pragmatic* support for a theory's theoretical commitments. While the language of science should be understood literally, "there is no need to believe good theories to be true, nor to believe *ipso facto* that the entities they postulate are real" (van Fraassen 1980, 11-2). Using the theoretical language of a theory remains consistent with withholding belief in the truth of the theory. Thus, van Fraassen counters criticism (ii).

2.1.3 UNDERDETERMINATION AND SELECTIVE SCEPTICISM

Ladyman calls the third criticism of van Fraassen's scepticism about unobservables 'the most popular realist response' – it is also the most problematic for the constructive empiricist. Ladyman characterises the underdetermination problem as follows (2000, 842):

> […] all the facts about observable states of affairs will underdetermine theory-choice between T_0, a full realistically construed theory, and T_1, the claim that T_0 is empirically adequate. However, all the evidence we have available now will underdetermine the choice between T_1 and T_2, the claim that T_0 is empirically adequate before the year 2001. Furthermore all the facts about all actually observed states of affairs at all times will underdetermine the choice between T_1 and T_3, the claim that T_0 describes all actually observed events.

Thus, even the judgement that T_0 is empirically adequate is underdetermined by the available evidence. The realist at least may argue that *inference to the best explanation* (IBE) warrants belief in T_0 and breaks the underdetermination (*Op Cit*), but van Fraassen cannot since he rejects IBE by saying:

> [A person] becomes irrational […] if he adopts it as a rule to [use IBE], and even more so if he regards us as rationally compelled by it (1989, 142).

In place of IBE, van Fraassen advocates what he calls 'voluntarism' in epistemology, according to which ampliative inferences are not irrational so long as constraints of consistency (e.g. such as those imposed by probability theory) are not violated (1989).

The constructive empiricist needs an ampliative principle to support the move from the extreme sceptical hypothesis that the world looks *as if* it exists to the view that the world *does* exist. Michael Devitt (2005) argues however that the same principle could be used to accept abduction to the existence of unobservables. Since van Fraassen nevertheless refuses to believe in the existence of unobservables, his scepticism must be arbitrary and selective concludes the realist (Ladyman 2000, 845).

2.2 A POSITIVE ARGUMENT FOR CONSTRUCTIVE EMPIRICISM

Van Fraassen acknowledges that even in endorsing a simple perceptual judgement, and certainly in accepting a theory as empirically adequate, he is "sticking [his] neck out". But, he argues, "[t]here is no argument there for belief in the truth of the accepted theories, since it is not an epistemological principle that one might as well hang for a sheep as for a lamb" (1980, 72). That is to say, if belief in empirical adequacy is sufficient to account for the aims and practices of science, then, despite the fact that the ampliative inference used to move away from extreme scepticism could be used to support the realist's claim as well as the empiricist's, going further than empirical adequacy would amount to taking an unnecessary epistemic risk for no extra empirical gain.

On the other hand, realists such as Psillos claim that at least realism can offer *explanations* for the observable phenomena and claim that science has "push[ed] back the frontiers of ignorance" (Psillos 1996, 42). However, as Ladyman notes, van Fraassen is content to argue that *empiricists* should be constructive empiricists rather than scientific realists because, from an *empirical* point of view, "the extra strength of the realist position is illusory" (2000, 844).

So, van Fraassen rejects realism and advocates constructive empiricism, not because he thinks the former is irrational, but because the latter "makes better sense of science […] than realism does and does so *without inflationary metaphysics*" (1980, 73, my emphasis).

Despite this, Ladyman thinks that constructive empiricism has no normative force for a non-empiricist, and as such, a stalemate has been reached. In (2000), Ladyman presents the most troublesome argument yet against van Fraassen in an attempt to give positive grounds for rejecting constructive empiricism.

3. LADYMAN'S OBJECTION TO CONSTRUCTIVE EMPIRICISM

Ladyman begins by analysing van Fraassen's various (inconsistent) writings on modality[2], and concludes that there are several viable interpretations of van Fraassen's views. None of them however, particularly van Fraassen's modal nominalist position, involve the belief in objectively construed (theory-independent) modal statements, which Ladyman argues the constructive empiricist needs (in order to circumscribe the observable in a principled way), but explicitly rejects[3].

For van Fraassen, observable phenomena need not actually be observed. Rather, a phenomena is observable "if there are circumstances which are such that, if [the phenomena] is present to us under those circumstances, then we observe it" (1980, 16). For example, van Fraassen recommends that we should believe in dinosaurs and the moons of Jupiter because were circumstances to obtain such that they were present to us (e.g. if we were standing on the latter), then we would observe them. Ladyman identifies two consequent questions about the claim that entity X is observable (2000, 850):

a) Is X's observability a theory-independent fact?
b) If so how can we know such a fact?

Ladyman considers b) first. Recall that van Fraassen describes the 'able' in 'observable' as referring

2 See in particular (1979, 412), (1980, 197) and (1989, 213).
3 Recall what van Fraassen says: "To be an empiricist is to withhold belief in anything that goes beyond the actual, observable phenomena, and to recognize *no objective modality* in nature." (1980, 203, my emphasis).

to our limitations as measuring devices, described by the 'final physics and biology'. On this account, science determines what is or is not observable. However, scientific theories concern themselves not only with actual, but also with *possible* observations. But, argues Ladyman, "how can van Fraassen rely upon theoretical science, which he does not believe to be true, to determine the limits of his scepticism?" (2000, 850).

If observability was a theory-dependent matter, then whether certain phenomena were observable or not would depend on which theory was used to describe them. Were this the case though, then the observable/unobservable distinction would have "no epistemic significance and constructive empiricism could not be sustained" (Ladyman 2000, 850). Van Fraassen concedes this much:

> To find the limits of what is observable in the world described by theory *T* we must inquire into *T* itself […] This might produce a vicious circle if what is observable were itself not simply a fact disclosed by theory, but rather theory-relative or theory dependent. […] I regard what is observable as a theory-independent question. It is a function of facts about us *qua* organisms in the world (1980, 57-58).

Therefore, van Fraassen must answer a) with an affirmative: if X is observable then it *is* an objective fact that if it *were* present to us then we *would* observe it.

The problem arises when we recall that the circumstances necessary for the observation of certain observable phenomena never *actually* obtain: they are *counterfactual*. For van Fraassen to be able to demarcate these *as* observable, he must believe at least some counterfactuals implied by scientific theories such as 'if a dinosaur were presented to us in the appropriate circumstance, then we would observe it'. Furthermore, van Fraassen must take such modal facts to be objective. Otherwise, his epistemic attitude "will depend upon a distinction that is entirely arbitrary" (Ladyman 2000, 851).

It seems then that the constructive empiricist must engage in some objective modal metaphysics in order to sustain his epistemic attitude towards science. However, if van Fraassen were to recognise an objective modality (say, by becoming a modal realist) and consequently allow objective modal facts to determine the truth value of counterfactuals, this would be 'totally incompatible with constructive empiricism' argues Ladyman. Recognising an objective modality would be at odds with the empiricist's disdain for modality and would undercut van Fraassen's chief positive argument for the constructive empiricist position; namely that it can do away with metaphysics. Hence, "in the face of these problems" concludes Ladyman, "[…] constructive empiricism is untenable as a philosophy of science" (2000, 855).

3.1 VAN FRAASSEN AND MONTON'S REPLY TO LADYMAN

In 'Constructive Empiricism and Modal Nominalism' (2003), Bradley Monton and van Fraassen attempt to clarify various aspects of the constructive empiricist position in face of Ladyman's argument, by addressing whether the objectivity of the observable requires there to be objective modal truths. They argue that this is not the case and that counterfactuals relating to the observability of unobserved phenomena *can* be objectively true because observability is an objective, *non-modal* property. That is to say, the "status [of] observability […] is not different from that of […] 'made of brick' or '75 feet long' (2003, 413).

Monton and van Fraassen argue this point by means of an example: To determine whether gun flashes would be visible under certain conditions, a scientist determines the properties of these flashes such as their duration and intensity. If successful, he would conclude that a certain range of

values for certain measurable parameters pertaining to the flash would be 'jointly necessary and sufficient for visibility under certain conditions' (2003, 413). So, without involving any modality, the scientist is able to determine whether a gun's flash would be observable or not under certain conditions. Monton and van Fraassen propose that "what goes here for the visibility of gun flashes […] goes equally for observability in general of any sort of object, event or process" (2003, 413). Therefore, generalities about actual facts determine what is observable. No objective modalities are needed to account for the epistemic attitude of the constructive empiricist.

3.2 LADYMAN'S REBUTTAL

In response to Monton and van Fraassen's claim that what is observable will follow from certain generalities about actual facts, Ladyman correctly argues that unless we consider generalisations in the form of scientific laws to have correctly latched on to objective features of the world, generalisations about what actually happens to as-yet-unobserved phenomena will not be enough to determine anything about what would happen if such phenomena were present to someone (2004, 762). Since van Fraassen must reject the first premise (since accepting it would amount to becoming a scientific realist), Ladyman concludes that science could never be used to determine whether something was objectively unobservable or just as of yet unobserved.

Indeed, even if observability was an objective non-modal property, the connection between observability and counterfactuals could not be 'sundered', for as Ladyman rightly argues, "in the case of observable entities like unicorns and dragons, it is the truth of the counterfactual claim that if they had been present to us we would have observed them which assures us that there are no such things" (2004, 763).

3.3 CONCLUDING REMARKS

In the last section of their paper, van Fraassen and Monton seem to partly cave in to Ladyman's critique by suggesting that it may be easier for a modal *realist* to be a constructive empiricist than for a modal antirealist. They suggest that adopting modal realism would not 'vitiate' the arguments that make constructive empiricism plausible. While adopting modal realism *would* amount to embracing some inflationary metaphysics admits van Fraassen, he concludes that other motivating arguments still exist for constructive empiricism that 'do not depend on modal metaphysics' and allow constructive empiricism to make the best sense of science (2003, 421).

I think that van Fraassen's concluding remarks are inappropriate and unnecessary. Giving an account of the aims and practices of science without appeal to inflationary metaphysics has traditionally been one of constructive empiricism's greatest strengths and principle motivation, and it should not be given up so easily. In what follows, I argue that by adopting modal agnosticism, the constructive empiricist can overcome Ladyman's objection while remaining true to his programme's original rationale. To motivate my claim, I will first need to locate modal agnosticism in the wider possible world debate and explain the position in detail.

4. AN INTRODUCTION TO POSSIBLE WORLDS

Our everyday use of words such as 'possibly', 'might', 'could', 'necessarily', 'must' and so on, lays bare our intuition that some but not all things could have been otherwise. Questions about such matters are questions about *modality*. These modal notions should be distinguished from *epistemic* possibilities. Consider the statement "for all we know, there may or may not be a solution to the 'N

vs. NP' problem". Epistemically, either option is possible, but whatever the answer proves to be, it could not have been otherwise (it is necessary in a modal sense).

Modal notions are most commonly interpreted through talk of possible worlds. For such talk to be philosophically useful, we need to know what it means and what is gained by its application. This depends on the view one takes of possible worlds. The traditional debate is conducted by the realist, committed to the view that a plurality of possible worlds exists, and the antirealist who denies this.

4.1 GENUINE REALISM

The most infamous type of realism about possible worlds is David Lewis' genuine realism. The theory's salient ontological theses include (Divers 2002, 45-6):

(OC1) An infinite plurality of possible worlds exists.
(OC2) Possible worlds differ only in content, not in kind, to our own.
(OC3) 'Actuality' is indexical - from the standpoint of each world, that world is itself 'actual', and all other worlds non-actual.
(OC4) Possible worlds are spatiotemporally and causally isolated from every other world, except from itself, or part of a world.
(OC5) Individuals are world-bound in that they don't wholly exist in more than one world.

4.2 GENUINE REALISM'S INTERPRETATION OF POSSIBLE WORLDS

Genuine realism interprets possible world talk differently depending on whether it involves *de dicto* or *de re* modality. For the purposes of this paper, modalising *de re* will be understood as modalising about a specific object or thing, such as David Attenborough, while modalising *de dicto* is modalising about a proposition, such as 'blue swans exist'. Following John Divers (2002, 43), in a straightforward case of *de dicto* possibility, we start with a claim in English (DD1), move to a neutral possible world claim (DD2) and then to the genuine realist interpretation (GR1):

(DD1) There could have been blue swans.
(DD2) There is a possible world at which there are blue swans.
(GR1) $\exists x \exists y(Wx \,\&\, Pyx \,\&\, By \,\&\, Sy)$ [4]

As interpreted by genuine realism, a world ('Wx') is construed as possible by virtue of its unrestricted existence ('$\exists x ... (Wx ...)$'), and the existence of blue swans ('By & Sy') at the world is construed as the world having such things among its parts ('Pyx').

In the case of *de re* possibility, we start with a claim in English (DR1), move to neutral PW claim (DR2) and then to the genuine realism interpretation (GR2):

(DR1) David Attenborough could have worked for Channel 4.
(DR2) There is a possible world at which David Attenborough works for Channel 4.
(GR2) $\exists x \exists y(Wx \,\&\, Pyx \,\&\, Cyd \,\&\, Hy)$ [5]

(GR2) is interpreted as before, but additionally, for the world to represent David Attenborough as

4 Where Wx = x is a world, Pyx = y is a part of x, By = y is a blue and Sy = y is a swan
5 Where Cyd = y is a counterpart of d, d = David Attenborough and Hy = y works for H and H = Channel 4.

working for Channel 4, it needs to have as a part a *counterpart* of David Attenborough ('Cyd'), where that counterpart works for Channel 4 ('Hy').

As shown, Lewis' system offers a "system of analyses of the family of modal concepts in which no modal concept is taken as primitive, and which underwrite the practise of conducting our modal reasoning in the medium of ordinary first-order quantificational logic" (Divers 2004, 660). While Lewis believed that his realism was credible on a cost-benefit analysis (1986), the theory has struck many as simply too ontologically extravagant and epistemologically indefensible.

4.3 ERSATZ MODAL REALISM

Before proceeding, I wish to briefly distinguish Lewis' genuine realism from what he calls 'ersatz modal realism'. Ersatzists think a plurality of possible worlds exist, but that these are abstract rather than concrete entities. Only one of these ersatz worlds represents the concrete world correctly: it is the actualised ersatz world. All of the other ersatz worlds remain unactualised. Like the modal antirealist, the ersatz realist tries to derive the benefits of genuine realism from a more 'safe and sane' ontology. Unfortunately, it is beyond the scope of this essay to consider ersatz realism in further detail. I point the interested reader to Lewis (1986) or Divers (2002), both who argue against ersatzism, for a fuller treatment.

4.4 MODAL ANTIREALISM

Modal antirealists wish to benefit from talking *as if* possible worlds existed without being committed to the genuine realist's ontology. They must adopt an interpretational stance of possible world speak that avoids commitment to the following three-part conjunction: (a) declarative sentences involving possible worlds are truth apt, (b) some are true and (c) some have a semantic structure that validly entails the existence of a non-actual world (Divers 2002, 22).

In the following sections, as space is limited, I will only briefly introduce modal expressivism and modal fictionalism so as to be able to develop the error-theoretic and agnostic modal positions in greater detail.

4.4.1 DENYING TRUTH APTNESS: MODAL EXPRESSIVISM

There are three different strategies the antirealist could adopt. The most radical is to deny the first conjunct (a) and hold that possible world sentences are *never* truth-apt. A parallel can be drawn with Blackburn's Expressivism about moral judgements. Expressivists maintain that when we make moral judgements, we do not refer to a moral fact. Rather, we say something that is reflective of an attitude we hold. For example, when someone utters the sentence "murder is wrong", the expressivist account interprets them as saying "Boo! to murder" rather than "the statement 'murder is wrong' is true". While space does not permit me to do so, Blackburn has indicated how and why a modal antirealist might appeal to such a strategy (1984, 213-6).

4.4.2 DENYING APPROPRIATE SEMANTIC STRUCTURE: MODAL FICTIONALISM

Alternatively, the antirealist could deny conjunct (c) and adopt a 'structure-based antirealism' (Divers 2002, 23), refusing to accept that possible world sentences have the necessary semantic structure which permits valid inference to the existence of non-actual worlds. The most discussed

version of such a strategy is known as modal fictionalism.

The typical modal fictionalist claims that possible worlds are merely fictional entities: there is no actualised possible world in which blue swans literally exist. Rather the literal truth is that *according to the fiction of Genuine Realism*, there is a possible world in which blue swans exist. On this account, possible world talk such as "there is a possible world in which blue swans exist" should be understood in the same way as talk about paradigmatically fictional objects, such as "there is a brilliant detective at 221b Baker Street" (Rosen 1990). Unfortunately, formal proofs have recently been developed showing that modal fictionalism is self-refuting. In particular, I point the reader to Stuart Brock's paper (1993).

4.4.3 DENYING TRUTH: MODAL ERROR THEORY

Lastly, the antirealist could grant that declarative possible world sentences are truth-apt and that they mean what they appear to mean, but deny conjunct (b) by refusing to assent to the truth of any possible world sentence. Divers calls such strategies 'Factualist Antirealism' and identifies two strands: error-theoretic and agnostic.

The characteristic claim of the stronger error-theoretic position is that all sentences of PW discourse are false. Again, a parallel can be drawn with moral theory, where J. L. Mackie has proposed an error theory about moral properties. Mackie holds that moral claims ascribe moral properties (such as wrongness) to items (such as murder). Moral claims are true when they actually have the properties ascribed to them and false otherwise. However, Mackie argues that because moral properties would be queer if they existed, moral claims are *always* false because moral properties do not exist.

The ontology of the error-theoretic possible world theorist is that no world apart from the actual world exists. Subsequently, any sentence of possible world discourse which entails the existence of a non-actual world is false. As Divers notes, this position has a direct impact on our modal commitments, for "error theory about possible-worlds, when allied with Lewisian analyses, generates a collapse of the *de dicto* modalities"[6] (Divers 2004, 667). To see this, consider the following Lewisian interpretations of possibility, impossibility, necessity, and contingency:

(LP) It is possible that X iff there is some world at which X.
(LI) It is impossible that X iff there is no world at which X.
(LN) It is necessary that X iff at all worlds, X.
(LC) It is contingent whether X iff there is a world at which X and a world at which not-X.

Since the actual world is the only world that exists according to the error-theorist's ontology, all and only that which is true *de dicto* of the actual world is true of some and all worlds. Hence, all that is true *de dicto* of the actual world is *necessarily* true and all that is not true is *impossible*; the existence of donkeys is necessary, and the existence of blue swans impossible. Furthermore, since contingency requires the truth of X at one world and the falsity of X at a *different* world, error theory rules out what contingency requires (namely, the existence of a plurality of worlds). It seems then that error theory causes the collapse of more modalities than is desirable.

6 The error-theorist is entitled to assert certain de re modalities without contradiction. As this is also true of the agnostic position which we shall cover later (see section 5.3) an error-theoretic account of de re modalities will not be considered here.

Let us now turn to the weaker of the two factualist antirealist strands, modal agnosticism, which promises to partly prevent such a collapse.

5. MODAL AGNOSTICISM

In "Agnosticism About Other Worlds: A New Antirealist Programme in Modality" (2004), Divers introduces modal agnosticism. Like van Fraassen who wants the benefits of micro-physical theory without commitment to unobservables, the modal agnostic wishes to secure at least some of the benefits associated with genuine realism without committing to the genuine realist's ontology. The characteristic commitment of the modal agnostic is that she holds herself as having no warrant for believing in the existence of any possible world other than the actual world. The modal agnostic should not assert any sentence which entails the existence of a world beyond that of the actual world, even though for all they know, *some may be true* (2004, 668).

5.1 RADICAL OR MODERATE?

Two types of agnosticism are worth differentiating: moderate and radical. Both the moderate and the radical agnostic are agnostic about the existence of any possible world other than the actual one, but where the moderate agnostic's agnostic beliefs may give way to *dis*belief in light of characterisations which would make the world an impossible one (where the world instantiates Q and not-Q simultaneously for example), the radical agnostic would remain agnostic about the existence of such a world (Divers 2004, 669).

The distinction between moderate and radical agnosticism is significant, for just as the error-theorist cannot prevent the collapse of many modalities, if the agnostic adopts a radical stance, she will be unable to prevent becoming comprehensively agnostic about modality. Such an approach would be highly undesirable as none of the benefits associated with genuine realism could be salvaged. On a Lewisian cost-benefit analysis, it is unlikely that radical agnosticism would fare well.

5.2 MODERATE AGNOSTICISM

The moderate agnostic on the other hand can 'forestall retreat' as she does *not* need to be comprehensively agnostic about modality (Divers 2004, 669). The moderate agnostic (hereafter simply 'agnostic') can thereby retain the expressive power afforded by genuine realism to many modalities, including claims of necessity, impossibility, *de re* possibilities and counterfactuals.

5.2.1 NECESSITIES AND IMPOSSIBILITIES

The agnostic can warrantedly claim to know things[7] as long as they do not posit or require belief in a possible world beyond the actual world. With this in mind, recall the Lewisian interpretation of impossibility (LI) and necessity (LN):

(LI) It is impossible that X iff there is no world at which X.

This can be expressed as:

(LI*) $\neg \Diamond X \leftrightarrow \neg \exists y(Wy \,\&\, P \in y)$[8]

(LN1) It is necessary that X iff at all worlds, X.

7 At least in as much as the genuine realist has a warrant to claim to know the matters in question.

8 Where Wy = y is a world.

By principles of first order logic, (LN1) can equivalently be interpreted as:

(LN2) It is necessary that X iff there is no world at which not-X.

And this can be expressed as:

(LN2*) $\Box X \leftrightarrow \neg \exists y(Wy \,\&\, P \notin y)$

Unrestricted negative existential claims do not posit the existence of a world ('$...\neg \exists y(Wy...)$' as in (LI*) and (LN2*) above) beyond the actual, and hence, the agnostic retains licence to assert them[9]. Therefore, the agnostic can express claims of necessity and impossibility such (LI) and (LN1).

5.2.2 COUNTERFACTUAL CONDITIONALS

In addition to necessity and impossibility claims, Divers (2004) argues that the modal agnostic also has grounds for claiming modal knowledge of counterfactual conditionals of the form:

(CF) $A \,\Box\!\!\rightarrow C$ iff there is no selected world at which (A-and-not-C)[10]

According to Divers, the realist has to earn the right to assert a counterfactual by "making the case that the satisfaction of the selection condition […] presents grounds for believing that there is no world which is an A-and-not-C world *and a selected world*" (2004, 672). The three conditions Divers outlines which are needed for this to happen are:

i) Contextual factors and factual considerations about the actual world do enough to constrain an appropriate selection relation to fill out the truth condition of the counterfactual
ii) The pragmatic, linguistic and factual knowledge of the speaker combine to determine the value of the truth-condition
iii) The realist speaker has justification for believing that the truth-condition is satisfied.

The crucial point, concludes Divers, is that at no point is the modal agnostic deprived of this story, and so can legitimately claim grounds for asserting counterfactual conditionals.

The fact that the agnostic is not deprived of such modal knowledge can be further elucidated by expressing (CF) formally:

(CF*) $(A \,\Box\!\!\rightarrow C) \leftrightarrow \neg \exists y(Wy \,\&\, S_A y \,\&\, C \notin y)$

where the world selection condition $S_A y$ is such that world y is a selected world only if A holds at y:

(CFS*) $S_A y \rightarrow A \in y$

As with (LI*) and (LN2*), at no point does (CF*) formally entail the existence of a possible world

9 In fact, the agnostic can assert unrestricted negative existential claims whenever her logic presents them as logical truths, such as 'it is impossible that there is something that is Q and not-Q' (Divers 2004, 670).

10 This is the counterfactual form preferred by Lewis (1986, 20-2). As Divers notes (2004, 671), counterfactuals expressed in a positive existential form would render any counterfactual with an impossible antecedent such as (Q & ¬Q) $\Box\!\!\rightarrow$ R as false rather than true, as no selected world would exist at which Q & ¬ Q and R, (since such a world would be impossible).

(‘$\exists x...(Wx...)$’). Therefore, “in so far as the realist is in a position to assert the intuitively true counterfactuals, so is the worldly agnostic” (Divers 2004, 673). As I shall argue in section 6, this ability is key in overcoming one of the major objections facing van Fraassen’s constructive empiricism.

5.3 THE AGNOSTIC’S DEFICIT

As we have just seen, the modal agnostic need not be comprehensively agnostic about modality. However, despite the agnostic’s entitlement to certain claims of necessity and impossibility, she must remain agnostic about certain modal claims of possibility and contingency.

To illustrate the deficiency, recall the Lewisian interpretations of possibility and contingency:

(LP) It is possible that X iff there is some world at which X.
(LC) It is contingent whether X iff there is a world at which X and a world at which not-X.

The agnostic must remain agnostic about the right hand sides of the biconditionals in (LP) and (LC), entailing an agnosticism about the left hand side of the biconditionals as well. To understand why, consider an arbitrary biconditional, P iff Q. If one does not want to be agnostic about P, then one must hold P to be true or false. If the biconditional is to be true however, P must be true, and hence on pain of irrationality, Q must be held to be true also. However, holding Q to have a truth value is contrary to being agnostic about Q. Hence, agnosticism about the right side, Q, of a biconditional entails agnosticism about the left side, P, and vice versa.
For example, reconsider the familiar statement:

(DD1) ‘There could have been blue swans’.

As we have shown previously, the statement on a Lewisian interpretation becomes thus:

(DD2) There is a possible world at which there are blue swans.

Reformulated as a biconditional, (DD2) becomes:

(DD3) It is possible that there are blue swans iff there is some world at which there are blue swans.

Here then, the agnostic wants to remain agnostic about the right side of the biconditional ‘there is some world’, forcing the agnostic to remain agnostic about the left side: the possibility of there being blue swans.

5.3.1 LIMITING THE DEFICIT

This example of *de dicto* possibility may be misleading however, as the modal possibility and contingency deficit facing the agnostic need not be as comprehensive as the analysis initially seems to suggest.

In section 4.4 of *On the Plurality of Worlds*, Lewis discusses his theory of representation of *de re* possibility, stating that “[p]ossibilities are not always possible worlds” (1986, 230). To illustrate what he means, Lewis imagines that he himself could have been someone else, namely Fred, who

exists in the actual world. Lewis argues:

> [Fred] is even a possible way for me to be. He is my counterpart under an extraordinarily generous counterpart relation, one which demands nothing more of counterparts than that they be things of the same kind. […]. The possibility in question is a possibility for me, not for the world. It is not some other world, differing haecceitistically from ours, which represents *de re* of me that I am Fred; it is Fred himself, situated as he is within our world (Lewis 1986, 232).

The modal agnostic may take great solace in Lewis' theory of *de re* representation, for armed with it, she no longer faces an assertibility deficit of *de re* claims of the type 'possibly X', when it is false that X but an appropriate this-worldly counterpart is X in the actual world. So, the agnostic can assert as confidently as the realist that "It is possible that Attenborough could have worked for Channel 4", because a counterpart of Attenborough exists in the actual world who *does* work for Channel 4. This account appropriately captures the contingency of Attenborough's employment with the BBC, rather than making it necessary.

Therefore, as Divers summarises, the agnostic faces assertibility deficits only over claims of the type 'possibly X' where i) she has no warrant to assert that there is no world at which X and ii) no warrant to assert that at the actual world X (2004, 674).

6. SAVING CONSTRUCTIVE EMPIRICISM: MODAL AGNOSTICISM

Recall the constructive empiricist's dilemma: if constructive empiricism is to be a coherent position, observability needs to be accounted for in a principled, non-arbitrary manner. Since van Fraassen has not done enough to show that observability can be construed objectively yet non-modally in all circumstances, the distinction between observables and unobservables can't be drawn without recognising an objective modality in nature by accepting modal realism. Yet in spite of what van Fraassen says, doing so would amount to undermining the main motivation for being a constructive empiricist.

Ladyman's objection need not spell the end of constructive empiricism as a tenable position however. I think that if the constructive empiricist were to become a modal agnostic, she *could* circumscribe the observable/unobservable distinction in a principled manner without entailing an inflationary metaphysics, thereby avoiding Ladyman's objection altogether.
Recall van Fraassen's 'rough guide' to what counts as observable:

> (O1) "X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it" (1980, 16).

If we consider van Fraassen's popular example about the moons of Jupiter, and interpret (O1) in terms of possible worlds, we get:

> (O2) There is no world that is physically possible relative to this world in which the moons of Jupiter are present to us in the right kind of circumstances and we fail to observe them.

Subsequently, (O2) can equivalently be interpreted as a counterfactual conditional:

> (O3) If the moons of Jupiter are present to us in the right kind of circumstances then we observe them (they are observable) iff there is no selected world at which

they are present in the right circumstances and we fail to observe them.

According to Ladyman, this is the type of counterfactual conditional which the constructive empiricist must be able to evaluate objectively in order to sustain a non-arbitrary distinction, but can't without adopting modal realism since the latter is the only modal position which allows counterfactuals like (O3) to be evaluated objectively.

However, if 'the moons of Jupiter are present' = A, 'we observe them' = C and a world y is a selected world (selected by the selection condition $S_A y$) only if A holds at y[11], then (O3) can formally be expressed as:

(O3*) $(A \mathrel{\square\!\!\rightarrow} C) \leftrightarrow \neg \exists y(Wy \,\&\, S_A y \,\&\, C \notin y)$

Notice that (O3*) has exactly the same formal form as (CF*), the formal expression of counterfactual conditionals I argued were expressible by the modal agnostic in Section 5.2.2. Therefore, I think that modal agnosticism could provide van Fraassen with the tools he needs to circumscribe the observable from the unobservable in a principled manner, allowing him to avoid Ladyman's objection all together.

6.1 POSSIBLE OBJECTIONS AND CONCLUDING REMARKS

There are a number of objections that could arise from my proposal. These include:

i) If modal agnosticism can express counterfactual claims objectively, then it must entail, like modal realism, inflationary metaphysics. Therefore, modal agnosticism is incompatible with constructive empiricism.
ii) If modal agnosticism was a natural match for constructive empiricism, then the constructive empiricist would already have adopted it.

Someone objecting that modal agnosticism and constructive empiricism are incompatible on the grounds outlined in (i) could only have misunderstood the modal agnostic project. The very aim of the agnostic programme is to garner as many benefits as possible from genuine realism while avoiding the inflationary possible world ontology Lewis' programme entails.

The realist could reply that *all* possible inflation isn't avoided, since the agnostic, unlike the modal error-theorist, acknowledges that other possible worlds *may* exist, even if she holds herself as having no warrant for believing in their existence. While this may be so, I would reply that the agnostic's epistemic attitude remains entirely compatible with that of the constructive empiricist's; both wish to remain agnostic about the existence of ontologically problematic entities that fall in their domain (possible worlds and unobservable phenomena respectively). Recall that while the constructive empiricist initially seemed to be in trouble for being unable to do without theory-laden language, van Fraassen was able to respond to this charge. I expect that if accused of using possible world laden language, the modal agnostic could respond in kind. Consequently, the parallel in epistemic attitudes further suggests that modal agnosticism is a natural partner for constructive empiricism.

---

11 $S_A y \rightarrow A \in y$

In response to (ii), there are various reasons which may explain why the constructive empiricist has not yet adopted modal agnosticism. From a pragmatic perspective, while Divers has published numerous papers developing modal agnosticism, the first of these appeared only after the principle constructive empiricist – realist debate conducted by van Fraassen and Ladyman ended in 2004. Thus, a properly developed version of modal agnosticism was not yet available with which van Fraassen could have responded.

Secondly, modal agnosticism still remains in its infancy. While Divers has worked hard to try and show that the modal agnostic can live with the modal deficit discussed in Section 5.3 (see Divers 2004), future research may reveal that the modal agnostic or constructive empiricist in fact cannot. Alternatively, further investigation may reveal other serious shortcomings with Divers' theory. Consequently, it is no surprise that constructive empiricists like van Fraassen may have been reluctant to adopt modal agnosticism too quickly.

Until shown otherwise however, I think my suggestion is one the constructive empiricist should take seriously. Though it may require the constructive empiricist to 'stick her neck out', I think that adopting modal agnosticism, even at this early stage, is a more promising route for the constructive empiricist to take than for her to abandon the main motivation for her position in light of the realist's challenge.

ACKNOWLEDGEMENTS

I am particularly grateful to Dr. Katherine Hawley for diligently supervising the development of this paper. I am also grateful to Jacob Busch, Dr. Peter Clark, Dr. Stephen Read and Rose Buckler for their comments on previous versions of this paper.

REFERENCES

Blackburn, S. (1984). *Spreading the Word*, Oxford: Clarendon Press.

Brock, S. (1993). 'Modal Fictionalism: A Response to Rosen'. *Mind*. 102/405, pp. 147-50.

Devitt, M. (2005). 'Scientific Realism', in Jackson, F. and Smith, M. (eds), 2005. *The Oxford Handbook of Contemporary Philosophy*, pp. 765-91. Oxford: Oxford University Press.

Divers, J. (2002). *Possible Worlds*. London: Routledge.

—— (2004). 'Agnosticism about Other Worlds: A New Antirealist Programme in Modality'. *Philosophy and Phenomenological Research*, 69/3, pp. 660-85.

—— (2006). 'Possible-Worlds Semantics Without Possible Worlds: The Agnostic Approach', *Mind*, 115, pp. 187-225.

Field, H. (1980). *Science Without Numbers: A Defence of Nominalism*. Princeton: Princeton University Press.

Ladyman, J. (2000). 'What's Really Wrong with Constructive Empiricism?: Van Fraassen and the Metaphysics of Modality', *The British Journal for the Philosophy of Science*, 51, pp. 837-56.

—— (2004). 'Constructive Empiricism and Modal Metaphysics: A Reply to Monton and van Fraassen', *The British Journal for the Philosophy of Science*, 55, pp. 755-65.

Lewis, D. (1986). *On The Plurality of Worlds.* Oxford: Blackwell.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. Penguin.

Monton, B. and van Fraassen, B. C. (2003). 'Constructive Empiricism and Modal Nominalism', *The British Journal for the Philosophy of Science*, 54, pp. 405-22.
Psillos, S. (1996). 'On van Fraassen's Critique of Abductive Reasoning', *Philosophical Quarterly*, 46, pp. 31-47.
Rosen, G. (1984). 'What is Constructive Empiricism?', *Philosophical Studies*, 74, pp. 143-78.
—— (1990). "Modal Fictionalism". *Mind*. 99/395, pp. 327-54.
—— (1993). "A Problem for Fictionalism About Possible Worlds". *Analysis*. 53/2, pp. 71-81.
van Fraassen, B. C. (1979). 'Russell's Philosophical Account of Probability', in G. W. Roberts (ed.), *Bertrand Russell Memorial Volume*, London: George Allen & Unwin.
—— (1980). *The Scientific Image*. Oxford: Clarendon Press.
—— (1989). *Laws and Symmetry*. Oxford: Oxford University Press.
—— (2002). *The Empirical Stance*. New Haven: Yale University Press.

How We Each Maintain Our Personal Identity

Mark Pexton

In this essay I will address the broad topic of personal identity. This topic deals with the problem of how we can truly claim that a person is the same person throughout his life or indeed over any period of his life. What is it about a person that means I can refer to him as a continuing entity? I will argue that some traditional approaches to identity miss the real question and propose that identity resides in a self's characteristic interaction with the world.

One's first response may well be that no, a person is never the same from one moment to the next, we refer to people by the same names only for convenience. Our experiences, our environment changes us and also we change ourselves from within. Our body is programmed to change when we grow older. Indeed we are not even the same from moment to moment, we have moods.

And is there even a self to talk about at all? Hume saw us as composed of our experience. We have a perspective but the viewer himself is elusive and indescribable. When we examine ourselves to find our self we find nothing but the present content of our experience.

So when we talk of our own lives what are we talking of, what is it that is born and dies? What is a self?

First let us be clear on some terms. We clearly accept that at no two points is a person identical in body or in mind, 'Maximum similarity within the groupings would limit them to atomic-point-instants. The purpose of the identity notion is wider breadth, but a grouping that included everything would not convey specific information'[12]. But this in my opinion is a different question from that of identity and here the main thrust of this essay differs from the view that, 'the relation of identity is logically one-one: I cannot be identical to two distinct people.'[13] Identity is an abstract

12 Robert Nozick, *Personal Identity Through Time* (pg 108 of *Personal Identity* ed. by Martin and Barresi, 2007)
13 Brian Garrett, *Personal Identity* in *The Shorter Routledge Encyclopedia Of Philosophy,* 2005

term and refers to a thing's essence – what it is that makes it what it is, what distinguishes it from another thing, a description that marks its boundaries, the definition of a thing. So the identity of a person is that which matches the definition of a person. So now we can see that the 'ever-changing' argument against a persisting self misses the point of the question of identity.

To support this use of the term identity, Locke's consciousness theory[14] clearly defines personal identity in terms of a continuation of the same consciousness (or memory) which nonetheless does change over time. And psychological reductionism argues for a similar thing – former selves share the same identity with future selves in virtue of sharing memories and beliefs and a development occurring from former to future states. Traditional responses to fission thought experiments treat identity as being a matter of sameness[15]. Consider the case of a single person's brain being split into two so that two brains are made that share exactly the same characteristics and are then placed into two bodies that share exactly the same characteristics. The puzzle asks whether the two descendents of the original brain, or the descendents and the original, share the same identity. Some contemporary theorists argue that the descendents cannot both be the same as the original since then they would both be the same as each other. I argue that this does not reflect the meaning of identity; I will return to my answer to this question.

In addition I am not attempting to find the criteria for ascertaining identity but what identity consists of in itself. Sydney Shoemaker in his 'Personal Identity and Memory'[16] treats memory as a criterion for ascertaining identity. He found that both bodily and mental criteria are used to ascertain identity in the third person and I am sure this is right but it does not answer the question of what it is to be someone and not someone else. He uses Locke's cobbler and prince thought experiment in which a prince wakes up in a cobbler's body (a case of bodily transfer) and the question is whether this person is now the prince or the cobbler. Shoemaker finds the case ambiguous since the two criteria bodily and mental are in conflict. I will argue however that it is not ambiguous.

Now to what we mean by self: I propose that it is the interaction between the internal and external. The self is the predictable result of interaction with the environment. It is an amorphous mass which forms into different shapes according to what environment it faces. Further, consider what things we think do not have selves. These include minerals and plants. Some people consider animals to have selves but on a much more limited basis than people. Indeed we do not consider animals to be persons. The notion of person is I think closely connected to a self. It is a matter of agency, the imposition of action on the environment. Harry Frankfurt has a similar thought, 'It is only through our recognition of a world of stubbornly independent reality, fact, and truth that we come both to recognise our selves as beings distinct from others and to articulate the specific nature of our own identities.'[17] Selves are things that have agency, causal power in the way rocks and plants do not. A self is distinct from its environment; rocks and plants are just considered members of the environment rather than a thing which has a separate, contained being.

So from this we can now conclude the identity of the self, the person: I suggest that it is the particular, predictable reaction to the environment, whether this is in terms of behaviour or internal mental events. With regard to the fission puzzle; if the situation were extended to imagine that the two fission descendents were placed into two different worlds then I would argue that they do share the same identity despite being distinct entities and they also share the same identity as the original.

14 John Locke, Chapter 27 of *Essay Concerning Human Understanding* (2nd edition), 1690 and in *Personal Identity,* ed. by John Perry, 2008 (first published 1975)
15 p2 of *Personal Identity,* edited by Raymond Martin and John Barresi, 2007
16 Sydney Shoemaker, "Personal Identity," *The Journal of Philosophy,* Vol.56, No.22 (October 22, 1959). And in *Personal Identity,* ed. by John Perry, 2008 (first published 1975)
17 Harry G. Frankfurt, *On Truth*, 2006 (p101, published by Pimlico, 2007)

When faced with different environments they will react in the same way and therefore have the same identity. In the case of the prince and the cobbler it follows that the person who wakes up is the prince since it is in the mind that the characteristic interaction with the environment is produced. The body is just a tool for interacting with the environment.

For me the problem of personal identity is treated by Nozick and others in a confused manner. A distinction must be made between quantitative and qualitative identity. It is the case that two people can be qualitatively identical since they may both have the same properties but for sure no two people can be quantitatively identical since they are by definition two separate whole people. Nozick's closest continuer theory of personal identity[18] concerns *qualitative* similarity between a person at different points in time. The causal connection and similarity of and between two 'person stages' means that they share the same identity, and whichever 'person stage' in the future is most closely related to the original takes precedence and continues to hold that identity. However there is no reason, according to the earlier distinction in terms, why there cannot be two or more holders of the same qualitative identity. Does it matter that two merely quantitatively different people had different pasts and so causes for their beings? If they share the same makeup (which contains within it a history of thought) then this is all that is needed to discern their shared identity in the present.

In addition, the view that identity is an extrinsic property, something that depends on persons other than the person in question (for example in the case of the two identical fission descendents), strikes me as very odd. It does not make sense to say that a person's identity depends on another thing for it to obtain or that a person can cease to continue his identity because there is a tie between two closest continuers. Whether I am who I am should depend on something about me and not on something else otherwise my identity is not a fact *about me* at all it is just the trivial dividing of the world into separate entities.

I think a lot of the concern over the nature of personal identity and in particular with regard to fission cases is the question over subjective continuation of experience. How can one center of experience result in two or more centers of experience? This is exemplified in the case of teletransportation. A replica of a man is on Mars after his cells were copied. Or in the case where half a brain can continue the functionality of a person and each half is separated into two different bodies. For we regard one center of experience (one experiencer) as mutually exclusive of all others. I do not propose to answer this question here as I am providing an objective account of personal identity but this issue may account for why it is argued that only one person can have one identity.

It does happen that we consider the identity of a person to have changed. This can only come from the environment. This compromises the person's agency by itself imposing a change in the characteristic interaction of the person. An example of this would be the onset of Alzheimer's or the loss of sight (though not necessarily). The objection to this may be that despite the extreme nature of the change, the self is just interacting with the environment in the same way as a less extreme external situation. The difference I argue is that the environment can act to diminish the agent's causal power without the agent having autonomy over the change.

Any change brought about by the person of his own characteristic interaction with the environment (a 'self forming action') and causal power is a result of the original person acting. The different future interaction is therefore a characteristic of the same identity. This raises however the question of free will and whether in fact as Kane suggested there is a process by which we can change

18 Robert Nozick, *Personal Identity Through Time* (In *Personal Identity* ed. by Martin and Barresi, 2007)

ourselves and our identities[19]. Are there situations where a person could have gone in one of two different directions or do we owe our present self to our original self? In terms of identity I think it is the case that we are the sum of the different decisions held together in a chain linked to our original selves. Our present self can be traced back to the original self and therefore holds the same identity. All the decisions made would not have been made had it not been for the nature of that original self.

In conclusion I have found that the identity of a self is held in that self's characteristic interaction with the environment and that two quantitatively different selves hold the same qualitative identity. In addition I have found that the environment can impose a change on the identity of a self and that a self and the environment are in a conflict of causal powers.

Bibliography

Personal Identity ed. by Raymond Martin and John Barresi, 2003, Published by Blackwell Publishing Ltd, 2007. Containing:

> Robert Nozick, 'Personal Identity through Time' in *Philosophical Explanations.* Cambridge, Mass.: The Belknap Press, imprint of Harvard University Press, 1981.

Brian Garrett, 'Personal Identity' in *The Shorter Routledge Encyclopedia Of Philosophy,* ed. by Edward Craig, Published by Routledge, 2005.

Personal Identity ed. by John Perry, 1975, Published by University of California Press, 2008. Containing:

> John Locke, Chapter 27 of *Essay Concerning Human Understanding* (2nd edition), 1694
>
> Sydney Shoemaker, "Personal Identity," *The Journal of Philosophy,* Vol.56, No.22 (October 22, 1959)

Harry G. Frankfurt, *On Truth*, 2006, Published by Pimlico, 2007

Daniel C.Dennett, *Freedom Evolves, 2003,* Published by Penguin Books, 2004 Containing reference to:

Robert Kane, *The Significance of Free Will,* 1996, Published by Oxford University Press, 1996

19 Robert Kane, *The Significance of Free Will,* 1996 and in Daniel C.Dennett, *Freedom Evolves (Chp. 4), 2003*