2016, 1 (2)

ARGUMENTA

The Journal of the Italian Society for Analytic Philosophy

First published 2016 by University of Sassari

© 2016 University of Sassari

Produced and designed for digital publication by the *Argumenta* Staff

All rights reserved. No part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from *Argumenta*.

Editor
Masssimo Dell'Utri
(University of Sassari)

Editorial Board
Carla Bagnoli (University of Modena and Reggio Emilia), Clotilde Calabi (University of Milano), Elvira Di Bona (Columbia University), Ian Carter (University of Pavia), Simone Gozzano (University of L'Aquila), Andrea Iacona (University of Torino), Elisabetta Lalumera (University of Milano-Bicocca), Pier Luigi Lecis (University of Cagliari), Olimpia Giuliana Loddo (University of Cagliari), Giuseppe Lorini (University of Cagliari), Marcello Montibeller (University of Sassari), Pietro Salis (University of Cagliari), Giuliano Torrengo (University of Milano)

Argumenta is the official journal of the Italian Society for Analytic Philosophy (SIFA). It was founded in 2014 in response to a common demand for the creation of an Italian journal explicitly devoted to the publication of high quality research in analytic philosophy. From the beginning *Argumenta* was conceived as an international journal, and has benefitted from the cooperation of some of the most distinguished Italian and non-Italian scholars in all areas of analytic philosophy.

Contents

Editorial

The first months of the year 2016 witnessed two great losses for the international philosophical community. In February 19th and March 13th, respectively, Umberto Eco and Hilary Putnam died, leaving a void whose real depth and breadth will only be revealed by the years to come.

The Editorial Board of *Argumenta* decided to pay a tribute to both, in one of the best ways available in circumstances like this—yielding the floor to them, as it were. Following a suggestion by Carlo Penco, in what follows you can find the transcription of the talk Umberto Eco gave at the Round Table that rounded off the 6th National Conference of the Italian Society for Analytic Philosophy—held in 2004 at the University of Genova. We warmly thank Richard Davies for editing the English version of the transcription, and hope that both the English and the Italian versions will manage to convey that fine sense of humour which was one of the chief features of Eco's personality.

As to Hilary Putnam, when he first was told about the launch of a new journal in analytic philosophy, he immediately agreed to contribute an article. His article titled *Reading Rosenzweig's Little Book* was then ready to be issued some months ago, but had to wait for the present issue—which is actually the second one of the inaugural volume of *Argumenta*. Now its appearance takes on a new light, and not only because it marks a further development and clarification of ideas put forward in his book *Jewish Philosophy as a Guide to Life*.

Together with Putnam's piece, the reader will find the articles of four other distinguished philosophers who kindly submitted their papers to welcome a newly born journal. We heartily thank all of them for what we take to be an important mark of encouragement.

Argumenta 1, 2 (2016): 113-114
ISSN 2465-2334
© 2016 University of Sassari

Moreover, this second issue of the inaugural volume features the first special issue of the journal: *New Trends in Philosophy of Mind and Epistemology*, edited by Maria Cristina Amoretti and Francesca Ervas. We are convinced that this special issue adds greatly to the quality of the volume as a whole.

The reply Diego Marconi has written in response to some passages in the article by Peter Hacker which appeared in the first issue, concludes the volume.

As usual, all the articles appearing in *Argumenta* are freely accessible and freely downloadable; once again, we are very grateful to the colleagues who acted as referees.

Buona lettura!

Massimo Dell'Utri
Editor

Millianism and the Problem of Empty Descriptions[1]

Frederick Kroon

University of Auckland

Abstract

Empty names present Millianism with a well-known problem: it implies that sentences containing such names fail to express (fully determinate) propositions. The present paper argues that empty descriptions present Millianism with another problem. The paper describes this problem, shows why Millians should be worried, and provides a Millian-friendly solution. The concluding section draws some lessons about how all this affects Millianism and the problem of empty names.

Keywords: Millianism, empty names, descriptions, negative existentials, pretense.

1. Introduction

Given the continuing interest in the question of the semantic status of names and the impressive evidence for something like a Millian theory of names, the problem posed by non-referring or *empty* names is often thought to be among the most difficult and important problems currently facing philosophy of language. It is hard to disagree. Harder to accept, however, is the all-consuming focus on empty *names*. It can be shown, I think, that empty or non-referring definite descriptions give rise to a related problem on the most plausible theories of definite descriptions, including Russell's. Furthermore, this new problem is one that ought to concern Millians. For not only are Millians among the most ardent supporters of Russell's theory of definite descriptions (although not, of course, as applied to names), but this new problem also affects certain Millian-friendly descriptivist solutions to the problem of empty names (ones that uphold a Millian story about names but allow definite descriptions to do duty for empty names in problem contexts). In short, there is a problem of empty descriptions that is also a problem for Millianism. The present paper describes this problem, shows why Millians should be worried, and provides a Millian-friendly solution. The concluding section draws some lessons about how all this affects Millianism and the problem of empty names.

[1] I am grateful to many philosophers for their useful comments on ancestors of this paper. Special thanks to David Chalmers, Anthony Everett, and Alberto Voltolini.

Argumenta 1, 2 (2016): 115-132
ISSN 2465-2334

© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.KRO

2. Descriptions, negative existentials, and relative clauses

It is clear why empty names present a problem for Millianism. Millianism, after all, declares that the semantic content of a proper name is simply its referent, with sentences containing names expressing (structured) singular propositions that have the referents of these names as constitutents. It follows almost right away that sentences containing empty names do not express propositions and so lack a truth-value—contrary to the seemingly *obvious* fact that sentences like 'If Vulcan exists, then it is a planet', LeVerrier believed that Vulcan was a planet', 'My children expect Santa Claus to visit on Christmas eve', and so on, say something true.[2] Consider in particular true negative existential statements like 'Vulcan does not exist' and 'Santa Claus does not exist'. It is the absence of referents for 'Vulcan' and 'Santa Claus' that seem to make these statements true, yet Millianism claims that this very absence prevents them from saying anything true.

This problem—the 'negative existential problem for names'—is just one of a spectrum of problems that together make up the so-called 'problem of empty names'.[3] But it is widely regarded as among the most pressing for Millianism, and the problem that Millians are often keenest to target in their defense of Millianism. There is one solution in particular that I want to highlight. This is what I shall call the 'descriptive-replacement' solution, adopted by, among others, Ken Taylor and by Fred Adams and his co-authors.[4] These authors argue that Millians are right to deny that sentences containing empty names express (complete) truth-valued propositions, but take this to show that our intuitions about the truth-value associated with sentences containing empty names relate not to any propositions semantically *expressed* but to what utterances of these sentences pragmatically *convey*. The name 'Santa Claus', for example, might be

[2] Some Millians have recently become tempted by the view that many so-called empty names are not really empty but (on at least some prominent uses of these names) stand for special abstract objects. See especially Salmon 1998 and 2002, and, for some recent discussion, Caplan 2004 and Braun 2005. Note that Millians sympathetic to such a view do not extend it to the case of "empty" descriptions; the view is supposed to be a view about names only. If, however, there really is a problem of "empty" descriptions, as I argue in this paper, then some of the arguments for the claim that names like 'Santa Claus' and 'Vulcan' are not genuinely empty may suggest that many so-called empty descriptions are also not really empty—so much the worse for those arguments, in my view.

[3] Thus Braun (2005) talks of The Problem of Meaningfulness for Names, The Problem of Meaningfulness for Sentences, The Problem of Truth Value, The Problem of Attitude Ascriptions (a special case of the last problem), and The Problem of Belief and Sincere Assertive Utterance.

[4] See Taylor 2000; and Adams and Stecker 1994, Adams, Fuller and Stecker 1997, and Adams and Dietrich 2004. Although these authors all accept versions of the view, whose locus classicus is Braun 1993, that atomic sentences containing empty names express gappy propositions, there are signficant differences between Taylor's view and that of Adams *et al*, in particular with respect to the way associated descriptive propositions are generated. Note also that some Millians, notably Nathan Salmon, have argued for a *semantic* version of a descriptive solution to the negative existential problem for names (see Salmon 1998, esp. 303-304). As far as I can see, none of these differences affect what I have to say in this paper. For an excellent discussion and critique of descriptive solutions, see Everett 2003.

associated with the description 'the jolly man who brings presents and lives at the North Pole' and an utterance of 'Santa does not exist' might then pragmatically convey the claim that the jolly man who brings presents and lives at the North Pole does not exist. Our inclination to believe that a speaker speaks truly when she says 'Santa Claus does not exist' shows that we easily mistake what is pragmatically conveyed for what is expressed.

I will return to the descriptive-replacement solution to the negative existential problem for names in the next section. But first I want to make good on my claim that the logical behaviour of empty definite descriptions in some contexts exhibits difficulties that are very similar to the ones affecting empty names. I shall again focus on negative existential statements, this time negative existentials featuring definite descriptions (*descriptive* negative existentials for short).[5]

Consider the following two statements, and imagine them to be uttered assertively:[6]

(1) The golden mountain—the only mountain to be mentioned in Smith's book on myths and legends—does not exist.
(2) The golden mountain does not exist; nor do any of the many other strange and wonderful objects mentioned in Smith's book on myths and legends.

What is characteristic of these two descriptive negative existential claims is that they not only deny the existence of some object, in this case the golden mountain; they also describe this object in other terms, in this case as an object mentioned in Smith's book on myths and legends. Call descriptive negative existentials that have this feature *classificatory*. Classificatory descriptive negative existentials like (1) and (2) strike (most of) us as perfectly coherent and, uttered in appropriate circumstances, as clearly true. It is not hard to see, however, that there is no consistent way of representing (1) in classical accounts of definite descriptions, such as Russell's theory of descriptions or reasonable systems of free logic. Represented in such theories, statements like (1) and (2) turn out to be not just false but logically false. The reason is not hard to see. In order to represent (1), for example, it seems that we should represent 'The golden mountain is the only mountain to be mentioned in Smith's book on myths and legends' as a conjunct of our representation of (1), since this sentence is logically implied by (1). But this component sentence is incompatible with another component sentence of (1), namely 'The golden mountain does not exist'. If the latter is true, there is

[5] The argument is a simplified version of an argument first presented in Kroon 2009. The present version also responds to a number of criticisms.

[6] The reason for restricting our attention to assertive utterances of a sentence like (1) is that in such cases the logical role of a non-restrictive relative clause like '[which is] the only mountain mentioned in Smith's book on myths and legends' is particularly straightforward. If this sentence is uttered in compound constructions, however, there are complexities. In particular, it looks as if the relative clause dominates the main clause in such constructions. Thus, in the sentence 'If the golden mountain—the only mountain mentioned in Smith's book on myths and legends—does not exist, then Smith will win the bet', the antecedent of the conditional is 'the golden mountain does not exist' and 'The golden mountain is the only mountain mentioned in Smith's book on myths and legends' seems to function as a conjunct. The curious behaviour of non-restrictive relative clauses may suggest that we should understand their nature in other ways. (According to Potts 2005, especially chapter 4, they encode conventional implicatures.)

no unique golden mountain, in which case the affirmative sentence 'The golden mountain is the only mountain to be mentioned in Smith's book on myths and legends' is false. This is so whether we adopt Russell's theory of descriptions or a philosophical competitor to Russell's theory like negative free logic.

Thus consider the way Russell would represent (1) when assertively used. On a Russellian analysis (1) contains as conjunct the affirmative identity sentence 'The golden mountain is the only mountain to be mentioned in Smith's book on myths and legends':

(1a) (The golden mountain = the only mountain to be mentioned in Smith's book on myths and legends) & the golden mountain does not exist.

But since for Russell the first conjunct (an identity sentence) requires that there *is* a unique golden mountain, while the second conjunct (a simple negative existential) requires that there is no unique golden mountain, this provides us with a blatantly inconsistent analysis of (1): in Russellian notation,

($1a_R$) $((\exists x)(G!x \,\&\, (\exists y)(S!y \,\&\, x=y)) \,\&\, \neg(\exists x)G!x)$.

Inconsistency is not the only problem with ($1a_R$). ($1a_R$) also appears to misrepresent the logical grammar of (1). Take the non-restrictive relative clause in (1) ('[which is] the only mountain …'). This clause qualifies a single occurrence of the description 'the golden mountain', with the predicate 'does not exist' attached to this single occurrence. (1a) and ($1a_R$), on the other hand, posit *two* occurrences of the same description, with the consequent eliminative analysis of these occurrences severing all connection between them. To get a reading that stays closer to the surface logical grammar of a claim like (1), we need to let 'exists' function as a genuine first-level predicate and use anaphora to refer back to the putative golden mountain, thus yielding something like:

(1b) The golden mountain = the only mountain to be mentioned in Smith's book on myths and legends & *it* [this golden mountain] does not exist.

In Russellian form, letting the second occurrence of 'the golden mountain' in (1) be replaced by a variable bound by the quantifier in first occurrence:

($1b_R$) $((\exists x)(G!x \,\&\, (\exists y)(S!y \,\&\, x=y)) \,\&\, \neg Ex)$.

($1b_R$) too is bound to strike us as an entirely unsatisfactory reading of (1), and for much the same reason as before. For Russell, the only acceptable reading of 'E(xists)' is as a universal predicate (*everything* exists),[7] so ($1b_R$) is again inconsistent. That suggests a familiar alternative. Given that sentences like (1) and (2) imply that there is such a thing as the golden mountain, even if it is also said to be non-existent, we might construe ($1b_R$) in a rather different way. Replace '$\exists$' and '$\forall$' with neutral, non-existentially loaded quantifiers 'Σ' and 'Π', ranging over a domain that encompasses both existent and non-existent objects, and let E be a discriminating predicate of existence. That way we get a consistent Meinongian reading of (1b):

[7] Russell is usually said to believe that 'existence' is not a predicate. That seems to me straightforwardly wrong since he must have known that in his logic 'x exists' can be represented as '$(\exists y)(x=y)$'. What is true is that he thought that in statements of existence there is no substantive role for such a predicate. For evidence that Russell may also have held a version of this view, consider, for example, his claim in the lectures on Logical Atomism that "[i]f there were any fact of which the unicorn was a constituent, there would be a unicorn and it would not be true that it did not exist" (Russell 1956: 248).

$(1b_M)$ $((\Sigma x)(G!x \,\&\, (\Sigma y)(S!y \,\&\, x=y)) \,\&\, \neg Ex)$.[8]

But if $(1b_M)$ is the most reasonable alternative reading of (1), it leaves Millians in an awkward position. For $(1b_M)$ not only contravenes the classical Russellian interpretation of the quantifiers, but also the way the theory of descriptions was, for Russell, supposed to help in the defense of a "robust sense of reality" designed to keep non-existent objects like unicorns and mythological mountains at bay.[9] And while Millians reject the descriptivist thesis that names are descriptions in disguise, they tend to support the theory of definite descriptions in its classic Russellian form, including its refusal to countenance genuine non-existents. They are not likely, then, to want to accept a reading like $(1b_M)$. But they also would not accept a reading like $(1a_R)$, since it is manifestly inconsistent.

Note that inconsistency is not peculiar to a Russellian treatment of definite descriptions in (1b). We get the same result if we adopt the kind of anti-Meinongian, anti-Russellian view of empty descriptions advocated by Mark Sainsbury in his *Reference without Referents* (Sainsbury 2005). On Sainsbury's view, which is based on a version of Negative Free Logic (NFL), 'The golden mountain does not exist' is true just if the term 'The golden mountain' or '[The x: Gx]' is empty. But if the term '[The x: Gx]' is empty the identity claim '[The x: Gx] = [The x: Sx]') is *false*, since in NFL all atomic affirmative sentences containing an empty term are false. It quickly follows that the sentence representing (1), '[The x: Gx] = [The x: Sx] & ¬E [The x: Gx]' is necessarily false.[10]

This is not to say that there are *no* consistent readings of (1) apart from the Meinongian one. There is one important candidate. We could opt for Ed Zalta's abstract-object interpretation of talk about the non-existent.[11] Zalta acknowledges two distinct modes of predication: encoding and exemplifying. The golden mountain is an abstract object that *encodes* the properties of being golden and a mountain, but does not exemplify them. (Exemplification is the more familiar way of having properties; the encoding way is peculiar to abstract objects.) This object in turn *exemplifies*, but does not encode, such properties as being the only mountain to be mentioned in Smith's book on myths and legends.[12] Assuming that the golden mountain does exemplify this property, it is indeed the only mountain to be mentioned in Smith's new book. And because this object is an

[8] For a recent example of a theory of descriptions couched in terms of an extreme anti-Russellian, Meinongian view of the quantifiers, see Priest 2005.

[9] According to Russell, "Logic … must no more admit a unicorn than zoology can; for logic is concerned with the real world just as truly as zoology, though with its more abstract and general features. … A robust sense of reality is very necessary in framing a correct analysis of propositions about unicorns, golden mountains, round squares, and other such pseudo-objects" (Russell 1919: 47-48).

[10] We also get inconsistency, but in a different way, if we follow Potts 2005 in treating the content of the non-restrictive clause in (1) as a conventional implicature rather than as part of the semantic content of (1). For so construed, the relative clause still gives rise to an identity commitment, and whether this commitment is construed in Russellian terms or in the manner preferred by NFL, it will be false exactly when the remainder of (1)—the part expressing what Potts calls the 'at-issue' content—is true.

[11] See Zalta 1983, 1988.

[12] More precisely: being the only object to be mentioned in Smith's new book that has the property of being a mountain, where '*having* property P' ranges over exemplifying P and encoding P.

abstract object it does not have the property of physical existence. Hence (1) counts as true, not false, so long as we assume that existence in (1) is physical existence. (For Zalta, every object, whether abstract or not, belongs to the domain of quantification of the classical existential and universal quantifiers, and so has the trivial property of "logical" existence, definable from by the classical existential quantifier as $\lambda x[(\exists y)(x=y)]$. But for Zalta it is physical existence rather than logical existence that is at stake in a sentence like (1).)

Most Millians, I suspect, would resist this interpretation of (1). After all, it comes at a considerable a cost, both ontological (all the things said not to exist *do* exist, even if not as physical objects) and ideological (the solution comes with a controversial distinction between two modes of predication). The interpretation would certainly be anathema to those with broadly Russellian sympathies.

3. The Millian's impasse

So there we have it: a familiar kind of descriptive negative existential that seems in its own way as embarrassing for a Millian, given what is likely to be her strong ideological commitment to Russell's theory of descriptions, as a negative existential involving an empty name.

For some Millians, the problem strikes even deeper. We saw earlier that some Millians attempt to solve the negative existential problem for names by arguing that a speaker who utters a "true" negative existential like 'Santa Clause does not exist' pragmatically conveys a true proposition that is somehow salient enough to be the real focus when the speaker and her audience characterize the negative existential as true. But such a descriptive-replacement solution quickly becomes subject to the negative existential problem for descriptions. Suppose Mother says to Johnny: 'Santa Claus does not exist'. On the descriptive-replacement solution, there is a salient true proposition that is thereby conveyed to Johnny, say the proposition that the jolly man who brings presents and lives at the North Pole does not exist. But suppose that Johnny seems unable to make the connection; such an item of news would be too hard to accept, so Johnny fails to grasp the appropriate descriptive association and hence the proposition. To drive the message home, Mother adds, as a reminder: 'You know, the nice man you always write letters to at Christmas'. This provides Johnny with enough information to enable him to recall who Santa is: he is the jolly man who brings presents and lives at the North Pole. The message is now all too clear to Johnny, but so is the problem that this presents for any Millian who favours such a solution to the negative existential problem for names. For on its most reasonable interpretation, the pragmatically conveyed proposition that Johnny now succeeds in grasping says that the jolly man who brings presents and lives at the North Pole is the nice man he always writes letters to at Christmas; and this jolly man does not exist. Given the theory of descriptions, or even a free description theory based on NFL, this implies that the most reasonable interpretation of Mother's complex negative existential renders it inconsistent.

The question is what the Millians should do about this problem of empty descriptions. Their options are limited. They will not want to reject Russell's understanding of the theory of descriptions in favour of something more Meinongian in which the existential quantifier is no longer the *existential* quantifier, as in ($1b_M$). Nor, if they are adherents of a system of free logic like NFL, will they want to change to a Meinongian-friendly version of free logic. And

they certainly should not argue that we must stop uttering statements like (1), on the grounds that such statements are philosophically confused. (If such statements are found to be philosophically confused and so unsayable, then so much the worse for the philosophy that finds them so.)

Should Millians, perhaps, back down on any initial distaste for Zalta's abstract-object interpretation of such sentences? After all, such an account provides readings that appear to be both consistent and anti-Meinongian, features that should impress Millians.[13] I am going to argue against any such move. This paper describes and motivates an alternative solution, one that is closer to Russell's in so far as it assumes his account of existence and his resolute repudiation of non-existent objects. The solution involves the thought that the kind of inconsistency displayed by (1) on its various inconsistent readings constitutes a quite general phenomenon, one that has nothing as such to do with existential statements and provides no succour for Meinongians. On the account I favour, the inconsistency of classificatory descriptive negative existentials like (1) is simply an interesting special case of this more general phenomenon.

I shall argue, in short, that Millians are free to accept ($1b_R$) as an appropriate reading for a descriptive negative existential like (1), despite the inconsistency of this reading. (For the remainder of the paper, I shall assume Russell's account of descriptions as quantifiers, although the general strategy seems to be available to negative free logicians like Sainsbury as well.) The argument for the general claim that inconsistency provides no bar to such readings will occupy the next section of the paper. After this, section 5 returns to the particular case of claims like (1) and (2), and uses the framework of section 4 to show why the ensuing Millian-friendly solution to the problem of descriptive existentials should be preferred to an abstract-object solution. The final section of the paper returns us to the problem of negative existentials for proper names.

4. Descriptive existentials and the role of pretense

To set the scene for the argument that is to follow, consider the following case. Suppose Jones claims to have discovered the golden mountain, contrary to Smith's often-repeated insistence that there is no such thing. Jones even displays what he calls 'irrefutable' photographic evidence for his claim, pointing to what seems to be a golden-coloured cone-shaped object at the top of one photograph. Smith's response is sarcastic:

(3) The mountain that Jones recently discovered—the golden mountain, no less—has one notable property Jones failed to disclose: it is not golden.

(Imagine that Smith thinks that the golden colour in the photograph is due to a reflection of the sun on a perfectly standard granite mountain peak.) (3) presents us with the following problem. On the surface, it expresses a straightforward contradiction because the non-restrictive relative clause '[which is] the golden mountain …' is in logical tension with the negated predicate 'is not golden'. But Smith clearly succeeds in saying something true and informative, so this inter-

[13] It should be said that Zalta's theory of objects has many other theoretical virtues, among them its ability to provide a new neologicist foundation for mathematics (Linsky and Zalta 2006). My concern here is limited to Zalta's well-known account of statements about non-existents.

pretation is unintended. The question is how to interpret the sentence in line with Smith's intentions.

The answer seems simple. In uttering the first part of (3), she pretends to go along with Jones's interpretation of the evidence—she is doing as if the latter's perspective on his discovery is correct, and she is acting as a kind of mouthpiece for Jones and this perspective. Smith *intentionally* mimics the way Jones represents his discovery, because her doing so is a staging post along the way to her saying, in a manner that deliberately mocks both the description and the describer, that it is a misrepresentation. (Indeed, it is easy to imagine this mimicking being continued, to the even greater irritation of Jones: 'Now that Jones has found the golden mountain', Smith might say, 'the fountain of youth surely cannot be too far away.')

If this is right, however, Smith's primary intention in uttering (3) cannot be to assert what (3) expresses, which is a contradiction. But how is her contradictory statement able to impart *any* kind of true, non-trivial information? Not surprisingly, given the role we have already assigned to Smith's doing as if Jones's interpretation of his discovery is correct, the notion of pretense will play an important role.[14] When Smith says that the mountain recently discovered by Jones—the *golden* mountain no less—is not golden, her words express a contradiction within her pretense. She thereby aims to exploit an interpretive tension that faces her audience, for no speaker is likely to want to assert a contradiction. What Smith's audience now understands is that her going along with the way Jones represented the world in his speech is a matter of pretense, not of belief, and that she is in fact disputing the aptness of the way Jones represented the world. She affirms that in reality it is *not* the case that the mountain recently discovered by Jones is golden.

In schematic form, this proposal—call it the *Pretense-Reality* proposal, or (PR)—says the following. Take the relevant core of a statement like (3) to be 'The A, which is B, is not C' (where being B entails being C). Such a statement is put to two rather different purposes by the speaker. First of all, the speaker asserts something from the perspective of her pretense: she *quasi*-asserts that the A is B. We might call this the *quasi*-content of her utterance of 'The A, which is B, is not C'. Secondly, she asserts something about how matters stand apart from the pretense: she asserts that it is *not* the case that the A is C. Call this the *real* content of her utterance.

To see how (PR) helps us to understand claims of non-existence, suppose that Smith's reasoning goes through a few more stages. Her claim (3) is first prompted by Smith's conviction that Jones's photograph features light playing on a granite mountain. She then develops a more radical view: Jones's photograph does not reveal a mountain at all—the thought that it does is caused by a perceptual illusion. This sequence of thoughts might be reported by Smith as follows:

(4) The mountain that Jones recently discovered—the golden mountain, no less —is not golden at all. In fact, if you look closely enough you'll see that it does not even exist. Jones's mountain is a trick played by the light.

[14] See Walton 1990 for a seminal account of the ideas of pretense and make-believe, and their importance for understanding the nature of the representational arts.

Once again, this is a perfectly intelligible sequence of claims, even if the situation is unusual. But how should we characterise what is being said? (PR) suggests the following answer. First of all, note that Smith's use of (4) commits her to all of the following:

(4i) The mountain that Jones recently discovered is the golden mountain.
(4ii) The mountain that Jones recently discovered is not golden.
(4iii) The mountain that Jones recently discovered does not exist.
(4iv) The mountain that Jones recently discovered is a trick played by the light.

In what follows I'll focus on the first three sentences. (I take (4iv) to be some kind of explanatory metaphor, but will have nothing further to say about it.) As before, we can make sense of Smith's commitments by saying that Smith *quasi-asserts* that the mountain Jones recently discovered is the golden mountain (in virtue of (4i)), while she *really* asserts that it is not the case that the mountain Jones recently discovered is golden (in virtue of (4ii)). But note that this way of characterising the real content of Smith's utterance is scope-ambiguous. If the description is assigned wide scope relative to negation, it yields the claim that there is a unique mountain that Jones discovered last month, and it is not golden. But it also has a weaker narrow-scope reading: it is not the case that (there is a unique mountain that Jones discovered last month, and it is golden). At the point at which Smith uttered (3), she intended the stronger reading.

The possibility of a narrow-scope reading bears significantly on the nature and consequent usefulness of claims of non-existence. As in the case of (3), Smith's choice of words in uttering (4) shows that she is engaged in the pretense that Jones's interpretation of the evidence is correct. And as in the case of (3), it may appear that the real content contributed by (4ii) is the internally negated proposition that there is a unique mountain that Jones discovered, and it is not golden. But against the background of (4iii) Smith is clearly trying to deny that Jones succeeded in discovering *any* mountain, let alone a *golden* mountain, and so her assertion must be understood in more austere terms: at the point where she utters (4iii) she must be taken as asserting the externally negated proposition that it is not the case that there is a unique mountain that Jones discovered, one which has the property of being golden.[15]

How does Smith manage to get her audience to understand that this is the reading she intends? And how does (4iii) allow her to communicate her newfound belief that there is in fact no mountain that Jones discovered, let alone a *golden* mountain? To set the scene for an answer to these questions, notice that the standard Russellian account of (4iii) faces a version of the very problem that affected (1). In the context of (4), (4iii) is just a variant of the classificatory negative existential

[15] It is tempting to think that (3) involves something like a referential use of the definite description 'the mountain that Jones recently discovered', based on Smith's (quasi-) perceptual sightings of a mountain on Jones's photo. But there are closely related examples that do not involve reference to (quasi-) perceptual sightings. Imagine, for example, that Smith thinks that Jones simply made up the story about discovering a new mountain. Smith's claim 'The mountain that Jones recently discovered—the golden mountain no less—is not golden; it does not even exist; Jones never left his study, and made the whole story up' raises all the same issues.

(4iii') The mountain that Jones recently discovered—the golden mountain, no less—does not even exist.

Given the presence in (4iii') of the non-restrictive relative clause '[which is] the golden mountain', we can now appeal to the same kind of argument as we used in the case of ($1b_R$) to argue for letting the description 'the mountain that Jones recently discovered' have wide scope in the analysis of (4iii'). And because in the context of (4), (4iii) is just a variant of (4iii'), this suggests that the best Russellian analysis of (4iii) also assigns wide scope to the description. That is,

($4iii_R$) There is a unique mountain recently discovered by Jones, and it does not exist. (In first-order notation, $((\exists x)(J!x \ \& \ \neg Ex)$.)

'E(xists)' is logically universal, however, and so (4iii) turns out to be inconsistent on its best Russellian analysis, just like (1). (As before, an analogous argument is available to those who, like Sainsbury, prefer to use NFL as their background logic.)

So in the context of (4) inconsistency affects not only (4ii) but also the negative existential (4iii). But how, in that case, should we understand what Smith says with her utterance of (4iii)? (PR) yields the following answer. To begin with, note that the use in (4iii) of the adverb 'even' ('does not *even* exist') implies—perhaps conventionally implicates—that the real content of Smith's utterance of (4iii) entails the real content of her utterance of (4ii): if Jones's mountain does not even *exist*, it certainly cannot be golden. This suggests that we should let the broad applicability of 'exists' explain the relationship between what is asserted with (4ii) and (4iii).

(PR) helps us to see how. We saw that, in the context of (4), (4iii) is a simple variant of (4iii'), so consider (4iii') again. Because 'exists' is (necessarily) universal, something's having a property B entails that it exists. So (4iii') can be understood as a degenerate case of a sentence subject to (PR), with B taken as the property of being uniquely a golden mountain and C being the property of existence. In that case, an utterance of (4iii') allows Smith to quasi-assert that the unique mountain recently discovered by Jones is the golden mountain, while at the level of real content she asserts that it is *not* the case that the unique mountain recently discovered by Jones exists. And much the same can be said about (4iii). It differs in not explicitly stating anything about the mountain in question, so we can understand its explicit commitments to be minimal: at the level of quasi-content, Smith says only that there is a unique mountain recently discovered by Jones, that the mountain recently discovered by Jones *exists*.[16] At the level of real content, however, nothing has changed. We can again take the real content of her assertion to be that it is not the case that the unique mountain recently discovered by Jones exists.

[16] In fact, in contexts where the description 'the A' is understood as having wide scope, 'The A does not exist' can be understood as a simple variant of the trivial classificatory negative existential 'The A, *which exists*, does not exist' (on this construal, '[which exists]' does not show up in 'The A does not exist' because it is logically redundant and hence elided). This again suggests that to articulate what is asserted with an utterance of a statement 'The A does not exist', when there is clear evidence that 'the A' is to be understood as having wide scope, we should treat it as a degenerate case of claims subject to (PR). (This is the explanation given in Kroon 2009. I now prefer the explanation given in the text.)

In principle, this statement of the real content of (4iii') and (4iii') allows two possible readings: the description can be assigned wide scope or narrow scope. But this time we have no option. Because it is a priori true that *everything* exists, we cannot reasonably interpret Smith in accordance with the wide-scope reading, that is, as asserting the internally negated claim that there is a unique mountain recently discovered by Jones, and it *does not* exist. The only option, then, is to interpret Smith's utterance as having the corresponding externally negated claim as its real content:

(4iii)$_{RC}$ It is not the case that: there is a unique mountain that Jones recently discovered, and it exists.

Eliding the logically redundant clause 'and it exists', this finally yields:

(4iii)$_{RC}$ It is not the case that there is a unique mountain that Jones recently discovered.

Note that little more than the logic of 'exists' was needed to derive this account of the real content of Smith's utterance of (4iii) (so close, of course, to Russell's own statement of the *meaning* of a sentence like (4iii)). But it is enough to show why, after hearing (4iii), Smith's audience is not able to hear the real content of (4ii) as one in which 'the mountain recently discovered by Jones' is assigned wide scope. The wide-scope reading is ruled out by **(4iii)$_{RC}$**. That was not true at the point where Smith uttered (3), since at that point she still thought that there was a mountain that Jones had recently discovered. Her only aim was to deny that this mountain was *golden*.[17]

5. Solving the negative existential problem for descriptions

By dividing the contribution made by an assertive utterance of an inconsistent statement like (2) or (3) into a quasi-asserted quasi-content and an asserted real content, (PR) gives us the material we need for a Millian-friendly solution to the negative existential problem for definite descriptions. What I have argued is that inconsistency is often just a spur to look beyond literal meanings to what the speaker intends to assert about the real world, against a background of propositions that the speaker entertains only in a spirit of pretense. Applied to the case of assertive utterances of (1) and (2), the lesson is that we can safely adopt an inconsistent reading like (1b$_R$) for (1) (similarly for (2)), and then use (PR) to divide the (quasi and real) contribution made by such utterances.

Here is what we can say about the quasi-content of such utterances, and in what sense we can count this quasi-content as true. With the description 'the

[17] The account on offer is pragmatic rather than semantic. By contrast, philosophers in the Walton tradition (among them Crimmins 1998, Armour-Garb and Woodbridge 2015, and Everett 2013) hold that negative existentials involving names are true in some kind of extended pretense in which 'exists' is used as a discriminating predicate; what makes claims like 'Holmes does not exist' true in this pretense are certain meta-representational facts such as the referential failure of singular *Holmes*-representations. Arguably, such a view can be extended in a non-meta-representational way to descriptive negative existentials. On my view, however, we have no more reason to think that 'exists' is used in a special game-bound way in negative existentials than that 'golden' is used in a special game-bound way in (3). The interpretive tension found in each is enough to force us to recognise in these statements a claim about how things are pretended to be and a different claim about how things really are.

golden mountain' assigned wide scope, we should understand the speaker who assertively utters (1) as someone who is engaged in the pretense that all the names and descriptions used by Smith when describing the claims of various myths and legends genuinely denote objects; she thereby pretends that there exists a unique golden mountain, that there are unicorns, and so on. (As the words 'mentioned by Smith' make very clear, this is pretense involving the *terms* Smith uses. We should not assume that the speaker knows exactly what terms they are.) Facts about the world (in particular, facts about what Smith's book contains) then make certain other propositions fictional or true from the perspective of her pretense. In particular, we can suppose that the nature of the descriptions in this book make it fictional that the golden mountain is the only mountain among the objects mentioned in the book. Hence the quasi-content of the speaker's assertive utterance of (1)—namely, that the golden mountain is the only mountain mentioned in Smith's book on myths and legends—is true from the perspective of her pretense. At the same time, the real content of her utterance is also true, for what the speaker asserts apart from the pretense is that there is no unique golden mountain which exists, and hence that there is no unique golden mountain (a reading derived in the same way as $(\mathbf{4iii})_{\mathbf{RC}}$) On the view I have defended, what informs our sense that the speaker speaks truly when she utters (1), despite the fact that (1) is inconsistent, is the fact that she *quasi*-asserts something that is true in her pretense, and at the same time *asserts* something that is genuinely true.

The case of (2)—'The golden mountain does not exist; nor do any of the many other strange and wonderful objects mentioned in Smith's book on myths and legends'—is more tricky, but only because nothing has been said about the sorts of designators that might feature in Smith's book. To see why this is important, consider the way an abstract-object theorist might try to handle (2). She would say that the second part of (2) declares both that i) there are many strange and wonderful objects mentioned in Smith's book, including the golden mountain, and that ii) none of these exist (that is, none of them physically exist; all of them are abstract encoding objects). On such an account, as on traditional Meinongian accounts, any objects that do not exist are individuated in terms of their own specific set of properties, in this case the properties they encode. Thus assuming that (2) talks of the golden mountain, Zeus, the fountain of youth, and so on, (2) will be a way of stating in general terms that the following are all abstract encoding objects: the golden mountain, Zeus, the fountain of youth, and so on.

But this presents the abstract-object theorist with the following problem. Some of the designators that Smith uses in his book to designate his many "strange and wonderful" objects will be definite descriptions ('the golden mountain' and 'the fountain of youth', say), and some will be mythical proper names like 'Zeus'. Zalta has a well-developed story about the abstract objects designated by such terms, but it is important to note that they may not be the only sorts of terms we use when expressing claims of non-existence. We can imagine, for example, that Smith's book also contains numerous plural descriptions; consider, for example, the term 'wood nymphs' in 'The forests were filled with numerous cavorting wood nymphs'. We can suppose, as is surely likely, that the myth in question does not provide discriminating descriptions of each individual wood nymph, nor that it quantifies their exact number ('numerous' will be considered precise enough). The difficulty this generates is this. In uttering (2), the

speaker may be trying to affirm that there are no such things as the golden mountain, the fountain of youth, Zeus, unicorns, humans taller than tall trees, water dragons, wood nymphs, and so on. On the account I prefer, that is the upshot of the real content of the speaker's utterance of (2), which I take at a first approximation to be something like the following:

(2)$_{RC}$ Fictionally, there are many strange and wonderful objects that the terms in Smith's book on myths and legends denote / apply to. This includes 'the golden mountain'. But there is no (unique) golden mountain,[18] and in reality none of the terms in Smith's book that *fictionally* denote strange and wonderful objects denotes any object.

This account does not assume that the speaker can somehow discriminate in her pretense among the entities to which these various terms apply, or fix on their exact number. Despite this, the account is easily able to say why the real content of (2) may well be true. There are, after all, no golden mountains; no Olympian gods; no unicorns, humans taller than tall trees, water dragons, wood nymphs, and so on.[19] If, fictionally, the objects mentioned in Smith's book are of these various kinds (and of no other kind) and there are many such objects and they are all strange and wonderful, then the real content of (2) is true.

Faced with this uncertainty about the terms Smith uses, Zalta's account faces an obvious difficulty. There are no discriminable encoding objects corresponding to each of the numerous wood nymphs, so no sense can be made of the claim that none of the numerous wood nymphs exist. No doubt Zalta can make sense of the claim that the *collection* of wood nymphs exists as an abstract object (one that encodes the property of being a large set all of whose members are wood nymphs), or that the species/kind *wood nymph* exists as an abstract object (one that encodes the property of being a kind K such that, necessarily, all and only wood nymphs). Assuming Zalta's theory, these objects can then be said to be non-existent. But in affirming (2), these are not the kinds of objects whose existence is denied. (2) is used to deny the existence of individuals like Zeus, the golden mountain, all humans taller than tall trees, each and every wood nymph and water dragon, and so on, not just the existence of discriminable things and kinds like the golden mountain, Zeus, and the species *wood nymph*. (2) might even indicate this quite explicitly by adding something like: 'for example, none of the many cavorting wood nymphs featured in chapter 3 (charming creatures—I wish their creator had told us something more about each one of them)'.[20]

[18] This metalinguistic way of describing the real content of an assertive utterance of (2) suggests that grasping the content presupposed an ability to use semantic ascent and descent on terms. But there may be other ways of describing the real content, say by appeal to there being properties associated with terms that, fictionally, pick out objects, and in reality pick out nothing.

[19] I assume that in each case we know properties whose instantiation is sufficient and necessary for there to be such objects. Where names and natural kind terms are concerned, causal descriptivism of the kind I favor tells one kind of story about such properties. Millians tend to be more guarded, either espousing a causal theory of some kind or invoking meta-representational properties (see, for example, Walton 1990 and Everett 2013).

[20] On the problem that non-discriminable fictional object present for a view like Zalta's, see, for example, Azzouni 2010. I should emphasise that the present kind of pretense

6. Back to Millianism

Earlier I pointed out that the negative existential problem for descriptions also affects the Millian-friendly *descriptive-replacement* solution to the negative existential problem for names, since the latter must hold that an utterance of a statement like 'Santa Claus—you know, the nice man you always write letters to at Christmas—does not exist' pragmatically conveys the problematic claim 'The jolly man who brings presents and lives at the North Pole is the nice man you always write letters to at Christmas; and this jolly man does not exist'. On the view I have advocated, the problem that such a claim poses for both Russell's theory of descriptions and a free description theory based on NFL is best tackled by acknowledging the role of pretense. On the basis of Mother's engagement in a piece of existential pretense, she quasi-asserts that there is a unique jolly man who brings presents and lives at the North Pole, the same person, furthermore, as the nice man Johnny always writes letters to at Christmas. We can assume that this quasi-content is true from the perspective of her pretense. Having alerted Johnny's attention to the first description by explicitly citing the second, she is now able to stand apart from the pretense, and declare that there is in fact no unique jolly man who brings presents and lives at the North Pole (the real content of her utterance, and a claim that is genuinely true, not just pretend-true).

There is, however, something rather disingenuous about such an account when presented in the context of the descriptive-replacement solution to the problem of negative existentials for names. After all, Millians who accept this solution do so because they think that a sentence like 'Santa Claus does not exist' fails to express a (complete) proposition, and so they try to capture our intuition that the sentence expresses a truth by descending to the level of associated descriptions. But if they are allowed to appeal to pretense to solve the resulting negative existential problem for descriptions, it is not clear why they should have bothered to descend to the level of descriptions in the first place when tackling the negative existential problem for names. They could simply have appealed to pretense at the level of *names*. Thus consider a speaker who pretends that she is successfully referring to someone with the name 'Santa Claus'. From the perspective of this pretense, the speaker continues to be able to contemplate alternative ways of identifying the person designated with her use of 'Santa Claus' (for example, as the kind man Johnny writes to every Christmas). Furthermore, she is now able to entertain the proposition expressed by the sentence 'Santa Claus does not exist' (a fully fledged proposition, this time, although only from the perspective of the pretense—a *pretend*-proposition).[21] Pretense theorists think that Millians should focus on what speakers might be held to assert through their use of negative existentials when empty names are understood from this kind of pretend perspective; they should not focus merely on propositions in the descriptive neighbourhood of these negative existentials—especially not given the rather tenuous nature of the connection between these negative ex-

view of claims about the non-existence of non-discriminable objects like individual wood nymphs is consistent with the claim that the individual wood nymphs do not exist because the species *wood nymph* does not exist.

[21] Among prominent works that feature the idea of a pretend-proposition are Kripke 2013 and Walton 1990.

istentials and the descriptive claims that have been suggested by supporters of the descriptive-replacement solution.[22]

Not surprisingly (given their general attachment to the Kripkean revolution in the philosophy of language), most Millian supporters of this kind of pretense theory argue that the real content of negative existentials involving names is given by non-descriptive claims: in the case of 'Hamlet does not exist', for example, the claim that singular *Hamlet* representations fail to refer (Everett 2013). It is important to emphasise, however, that Millians *need* not refuse a substantial role to descriptions in such cases. One reason why they might want to acknowledge such a role is that names and definite descriptions are often used in the same *kind* of negative existential construction (as in 'Neither Zeus nor the golden mountain exists'), and on the surface it seems odd to have a radically disjunctive account of what such a sentence is used to say. But there is also a philosophically more weighty reason, although one that many Millians will contest. There is good reason, in my view, to think that speakers have a grasp of the conditions under which names refer, or fail to refer. Presented with various possible scenarios involving the use of a name, they will know which ones contain the referent and which do not (and which are borderline).[23] Assuming something like a causal-historical account of reference, that suggests that speakers have implicit knowledge of the causal-historical conditions under which a name successfully refers to an object, which in turn suggests that the right theory of reference-determination is in fact not strictly causal but causal descriptivist (even if the right theory of the semantic *content* of a name is not descriptivist at all, but Millian).

If so, the pretense account alluded to two paragraphs ago can be fleshed out as follows. In pretending that the Santa-Claus tradition is factual, the speaker pretends that the reference-determining condition underlying her use of a name like 'Santa Claus' (we can stipulate this to be something of the form: 'the actual individual I am more or less reliably acquainted with under the name 'Santa Claus' as being such-and-so') singles out a real person. What she then asserts when she declares that Santa Claus does not exist can be derived in the same way as **(3iii)**$_{RC}$. It will be something like: It is not the case that there is a (unique) actual individual I am more or less reliably acquainted with under the name 'Santa Claus' as being such-and-so. This is true, and indeed necessarily true (given the rigidifying role of 'actual'); to that extent, it is necessarily the case that Hamlet does not exist.[24] Millians, given their historical commitment to both

[22] Consider, for example, Everett's objection that different speakers—and the same speaker at different times—are very likely to associate different descriptive contents with the same empty name (Everett 2003, 5.3). Adams and Dietrich respond to this and other worries in Adams and Dietrich 2004 (II.B), but in my view do not take the full measure of the worry (Kroon 2014). Not only can there be substantial variation among associated descriptions in the case of empty names, but such names are subject to a version of the problem of error—speakers can associate the wrong information with names because of misreading, mishearing or misremembering what was passed on to them.

[23] The locus classicus of this argument is Jackson 1998. I elaborate this argument in relation to the topic of negative existentials in Kroon 2014.

[24] Everett (2013: 72 fn. 45) criticises this account on the grounds that a) we do not hear negative existentials as making meta-representational claims, and b) it generates the wrong modal profile for a statement like 'Holmes does not exist' since it counts an utterance of 'Holmes exists' as true with respect to a world w in which, intuitively, Holmes

Russell's theory of descriptions and something like a causal theory of reference, should find such a conclusion congenial.[25]

References

Adams, F. and Dietrich, L. 2004, "What's In a(n Empty) Name?", *Pacific Philosophical Quarterly*, 85, 125-48.

Adams, F., Fuller, G. and Stecker, R. 1997, "The Semantics of Fictional Names,'' *Pacific Philosophical Quarterly*, 78, 128-48.

Adams, F. and Stecker, R. 1994, "Vacuous Singular Terms," *Mind & Language*, 9, 387-401.

Armour-Garb, B. and Woodbridge, J. 2015, *Pretense and Pathology: Philosophical Fictionalism and its Applications*, Cambridge: Cambridge University Press.

Azzouni, J. 2010, *Talking about Nothing. Numbers, Hallucinations, and Fictions*, Oxford-New York: Oxford University Press.

Braun, D. 1993, "Empty Names," *Noûs*, 27, 449-69.

Braun, D. 2005, "Empty Names, Fictional Names, Mythical Names", *Noûs*, 39, 596-631.

Caplan, B. 2004, "Creatures of Fiction, Myth, and Imagination", *American Philosophical Quarterly*, 41, 331-37.

Crimmins, M. 1998, "Hesperus and Phosphorus: Sense, Pretense, and Reference", *Philosophical Review*, 107, 1-47.

Evans, G. 1982, *The Varieties of Reference*, J. McDowell (ed.), Oxford: Clarendon Press.

Everett, A. 2003, "Empty Names and 'Gappy' Propositions", *Philosophical Studies*, 116, 1-36.

Everett, A. 2013, *The Nonexistent*, Oxford: Oxford University Press.

Jackson, F. 1998, "Reference and Description Revisited", in Tomberlin J. (ed.), *Language, Mind, and Ontology*, *Philosophical Perspectives*, 12, Malden: Blackwell, 201-18.

Kripke, S. 1973, *Reference and Existence: The John Locke Lectures*, Oxford: Oxford University Press.

fails to exist but John Perry is called 'Holmes'. But a) and b) are based on a confusion. It is reference-determining properties that do the work, not the meta-representational property of being picked out by a reference determining property. And these are certainly not properties like *being called 'Holmes'*. On my account, they are best thought of as given by rigidified descriptions that (typically) appeal to causal-historical relationships or relationships of acquaintance.

[25] Note that such an account suggests that different speakers (and the same speaker at different times) assert different propositions when they utter the same negative existential 'N does not exist'. I attempt to resolve this problem in Kroon 2014. Other theories of negative existentials tend to inherit a version of this problem, and so will need their own way of dealing with the problem. (Everett, for example, writes that "utterances of 'Holmes does not exist' will carry the information that singular representations which count as referring to Holmes within the scope of the pretense fail genuinely to refer" (Everett 2013: 72), a claim that is centred on the pretense *of a particular speaker*.)

Kroon, F. 2004, "Descriptivism, Pretense, and the Frege-Russell Problems", *The Philosophical Review*, 113, 1-30.

Kroon, F. 2009, "Existence in the Theory of Definite Descriptions", *The Journal of Philosophy*, 106 (7), 365-89.

Kroon, F. 2014, "Content Relativism and the Problem of Empty Names", in García-Carpintero, M. and Martí, G. (eds.), *Empty Representations: Reference and Non-existence*, Oxford: Oxford University Press, 142-64.

Linsky, B. and Zalta, E. 2006, "What is Neologicism?", *The Bulletin of Symbolic Logic*, 12/1, 60-99.

Potts, C. 2005, *The Logic of Conventional Implicatures*, Oxford: Oxford University Press.

Priest, G. 2005. *Towards Non-Being*, Oxford: Oxford University Press.

Russell, B. 1919, *Introduction to Mathematical Philosophy*. London: Allen and Unwin.

Russell, B. 1956, "The Philosophy of Logical Atomism", in Russell, B., *Logic and Knowledge*, Marsh, R.C. (ed.), London: George Allen & Unwin, 175-281.

Sainsbury, M. 2005, *Reference Without Referents*, Oxford: Oxford University Press.

Salmon, N. 1998, "Nonexistence", *Noûs*, 32, 277-319.

Salmon, N. 2002, ''Mythical Objects,'' in Campbell, J.K., O'Rourke, M. and Shier, D. (eds.), *Meaning and Truth: Investigations in Philosophical Semantics*, New York: Seven Bridges Press, 105-23.

Taylor, K. 2000, "Emptiness Without Compromise", in Everett, A. and Hofweber, T. (eds.), *Empty Names, Fiction, and the Puzzles of Non-Existence*, Stanford: CSLI Publications, 17-36.

Walton, K. 1990, *Mimesis as Make-Believe*, Cambridge, MA: Harvard University Press.

Zalta, E. 1983, *Abstract Objects. An Introduction to Axiomatic Metaphysics*, Dordrecht: Reidel.

Zalta, E. 1988, *Intensional Logic and the Metaphysics of Intentionality*, Cambridge, MA: MIT Press.

Happiness, Luck and Satisfaction

Kevin Mulligan

University of Geneva

Abstract

In some of its many forms, happiness is no emotion. But there is also an emotion of happiness which, like other emotions, has correctness conditions. The correctness conditions of happiness differ in several respects, formal and non-formal, from those of emotions such as admiration, fear and indignation. The account given here of the correctness conditions of happiness suggests an account of happiness as a species of satisfaction and an account of the relation between happiness and affective rationality or reason.

Keywords: Happiness, Luck, Satisfaction, Axiological Ascent

1. Dimensions and Objects of Happiness

Happiness has been the object of an enormous amount of philosophical reflection for a very long time. More recently, it has also been the object of a great deal of empirical work. Some of the philosophical or conceptual questions which have been raised are: What sort of an affective episode is happiness, the sort of thing which may but need not be felt at a time? What is the relation between happiness as an affective episode, and happiness as an enduring state or disposition? Is happiness, understood as an affective episode, ever an emotion? If so, what is its object, what is it about? Is happiness invariably a good thing?

In what follows, I shall put forward and argue for some answers to these questions. I shall consider the last three questions in greater detail than the others since it seems to me that they have been neglected in the philosophy and science of happiness. My answers to these three questions rely on a number of assumptions which I shall make explicit but which must here remain mere assumptions.

Is happiness always a positive emotion? Consider two fairly clear cases of happiness. Sam, who has just fallen in love, is blissfully happy. Roger, a religious believer who loves his God and takes himself to be loved in return, is blissfully happy. The first case is sometimes described as a state of felicity, the second as a state of beatitude. And both may be described in German as examples of *Glückseligkeit*. Bliss is positive, if anything is. But are the two examples of bliss examples of positive emotions?

Argumenta 1, 2 (2016): 133-145
ISSN 2465-2334

© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.MUL

Suppose that an emotion is a mental or psychological episode which is *about* or *of* something, a person, an action, a situation. Fear of a dog, admiration of a picture, being ashamed of a past deed are then examples of emotions. Is bliss about anything? The question is not a question about the reasons one has, if any, for being in a state of bliss or about the causes of such a state. One is afraid of the dog because it is or seems to be dangerous, admires the picture because it is or seems to be beautiful or interesting. But what one is afraid of is the dog, what one admires is the picture. If we distinguish in this way between what an emotion is about and one's reasons for feeling the way one does, then the happiness of bliss is not about anything, and so not an emotion.

But the happiness of bliss is not the only form happiness takes. Happiness *is* sometimes an emotion. Maria is happy about her impending promotion, Roger about the fact that he has passed an exam. Some of us at some time have been happy about aspects of our lives or situations. Sometimes such happiness takes the form of joy and rejoicing. One particularly important type of happiness or joy has as its object goods and situations which we are lucky to have or be in. These include lovers, spouses, children, power, wealth and many abilities. We may not always be aware of just how lucky we are as lovers, spouses, parents, children, citizens of functioning states and so on. But the quality of our happiness about these aspects of our lives emerges with startling clarity when we lose these goods, when a happy or lucky situation comes to an end. The unhappiness of loss is perhaps the most central type of sadness and grief is one of the most intense forms it may take.[1] Happiness about one's good fortune, like its opposite, the unhappiness of sadness over the loss of such good fortune or over one's bad or indifferent luck are emotions. Bliss, like its opposite, despair, is not.

The two forms of happiness identified so far are easily confused with cases which are sometimes described, more or less loosely, as cases of happiness or as cases of what makes us happy:

Sam is happy because he is experiencing many pleasurable sensations or feelings on the back of his neck
Sam is happy because he is pleased that he has passed his exams
Sam is happy because he got what he wanted
Sam is happy because he is enjoying the film
Sam is happy because he is enjoying watching the film/skiing/reading a novel…
Sam is happy because he is having fun
Sam's joy about the news he has received makes him happy
Sam is happy because he is moved by Maria's generosity
Sam is happy because he is very interested in, fascinated by, the film

The fauna of happiness are very varied. Amusement, being care-free, comfort, content, delight, elation, enthusiasm, exhilaration, affective fusion and ecstasy, fun, gaiety, gladness, interest, mirth, satisfaction, serenity, tranquillity, feeling well, felt well-being are all sometimes described as forms of happiness or as happiness-makers.

One fundamental type of happiness, then, has an object. In order to understand just what precise form its objects take I shall rely on an assumption about

[1] On sadness, grief and loss, cf. Roberts 2003: 234-36.

emotions in general. The assumption is that emotions may be correct or incorrect. Consider, first of all, an analogy. Our beliefs, convictions, opinions and judgments are often incorrect. If you believe that *p* and it is false that *p*, then your belief is incorrect. May emotions also be incorrect? Consider the following examples. Sam is afraid of a dog which is not in fact dangerous. Russell is indignant about some situation which is not in fact unjust. Roger is ashamed of some past deed which was not in fact shameful. Hans despises someone who is not in fact contemptible. Paddy is angry with someone who has not in fact offended him. How should the distinction between fear of a dog which is dangerous and fear of a dog which is not dangerous, between contempt for someone who is contemptible and contempt for someone who is not contemptible be characterized? We very often make such distinctions. Indeed the concept of *phobia* presupposes something like this distinction. Plato and Aristotle sometimes characterize non-intellectual and non-perceptual states and acts as correct and incorrect. They refer to correct and incorrect love (*eros*), desire and choice. Following their example, many philosophers have distinguished in passing between correct and incorrect emotions, although the distinction is rarely exploited in any detail except by Brentano and some of his students. Related assessments, also to be found in Plato and Aristotle and, for example, in the writings of Wolfgang Köhler and C.D. Broad, characterize some emotions as more or less (*in*)*appropriate* or (*un*)*reasonable*, *fitting*, *required* or *permitted*. We do not nowadays, in my experience, often call emotions correct or incorrect outside philosophy seminars. But we do describe some emotions as appropriate or inappropriate. We also say that someone has absolutely no reason, or no good reason, to be afraid or to be ashamed of this or that. Reasons which speak for or against something are typically defeasible reasons; they are reasons which can be trumped or defeated by better reasons. But there are perhaps non-defeasible reasons for or against something, reasons which cannot be defeated. Thus one may think that if Christian is really contemptible, then that is a non-defeasible reason for despising him. One way of understanding the correctness of emotions is in terms of non-defeasible reasons: to say that fear of a dog is correct is to say that there is a non-defeasible reason for being afraid of it, its danger. The fact that the dog is dangerous makes fear of the dog correct.[2]

Since the terms "incorrect emotion" and "correct emotion" are unusual and irritate many, particularly the politically correct, I shall employ them here in order to focus attention on an important and neglected distinction, one which may however be expressed in other terms.

Emotions, as we have seen, are not the only affective phenomena. Bliss, we said, is no emotion, because it is not about anything. Preference is not an emotion either but on some views it is an affective phenomenon or supervenes on or is determined by emotions.[3] The concept of preference is one of the most important concepts in the study of human behaviour. And preference, like fear, shame and admiration, may be correct or incorrect. If Susie prefers *x* to *y*, her

[2] If the distinction between non-defeasible and defeasible reasons is accepted, it is tempting to understand the latter in terms of the former and even to understand *being a reason for* in terms of *correct making*. But this is not the path taken by most contemporary philosophies of reasons.

[3] Cf. Mulligan 2015.

preference is correct only if x is better (for her) than y. Love of a person, unlike falling in love, is not an episode and so not an emotion, as we have defined the term. But it is love *of* a person. May it be correct? On one influential view, the Romantic or troubadour view, it cannot be correct or incorrect. In Shakespeare's words, "Love is not love/Which alters when it alteration finds" (Sonnet 116). For suppose Sally loves Sam because he possesses some valuable quality. Then her love would cease to be correct when he loses this quality. The non-Romantic view is formulated by an Irish aristocrat:

> I loved thee, beautiful and kind,
> And plighted an eternal vow.
> So altered are thy face and mind
> 'Twere perjury to love thee now
> (Robert Nugent)[4]

Are all emotions correct or incorrect? Consider sympathy. Sam has the impression that Maria is suffering or that she is unhappy and reacts sympathetically. He suffers *with* her (*sym*pathy, *com*passion) or is unhappy *with* her. Might his sympathy be incorrect? His reaction might be inappropriate because, although he has the impression that Maria is suffering, she is not in fact suffering. But suppose she really is suffering or is unhappy. There is no obvious candidate for the role of a value the exemplification of which by Maria's suffering or unhappiness would make Sam's sympathy correct and the non-exemplification of which would make Sam's sympathy incorrect. Sympathy, it seems, is an emotion which is neither correct nor incorrect. If this suggestion is correct, it suggests that sympathy, whatever useful functions it may have, for example in reducing the amount of what Max Scheler called "practical solipsism"—behaving *as though* (as opposed to believing *that*) other people are zombies—in the world, does not belong, as do other emotions and preference, to the sphere of affective rationality.

The distinction between correctness and incorrectness, as applied to beliefs, emotions and preferences, raises many philosophical questions. Are there beliefs which are neither correct nor incorrect? Are there emotions other than sympathy which are neither correct nor incorrect? Does correctness, like appropriateness, admit of degrees? Is the concept of correctness a normative concept? What is the relation between the *correct-incorrect* couple and the *right-wrong* couple which plays such an important role in (Anglophone) ethics? In what follows I shall not attempt to answer these questions. I shall assume that there are clear cases in which an emotion may be—correctly—said to be correct and other equally clear cases in which an emotion may be—correctly—described as incorrect. But I shall consider one worry about the distinction since my reply to the worry will play a role in what follows.

Is the distinction between correct and incorrect emotions not a piece of intolerable moralism? No—"correct" and "incorrect" are not obviously moral or ethical terms. This is clearly the case if the concepts of correctness and incor-

[4] What I have here called the troubadour view of love was expounded and forcefully defended by Scheler and McTaggart almost a century ago. Cf. Ronnow-Rasmussen 2011, ch. 6.

rectness are not normative concepts. But even if they are normative concepts they are not obviously ethical or moral concepts. After all, there are many non-ethical values—such as the value of health, pleasure or happiness and aesthetic values such as beauty, prettiness and the sublime. And there are many non-ethical norms, such as traffic rules and the normative principles of prudential rationality. So even if correctness and incorrectness are normative properties, they are not obviously moral or ethical properties.

The distinction between correct and incorrect emotions points to an ambiguity in the notion employed so far of the object of an emotion. Shame about a past deed seems to have two objects, to be about two things, the deed and its shamefulness. Similarly, indignation about some situation, for example the way Hans has treated Maria, is about two things, the situation or action and its injustice. Following Husserl, we may call the past deed *the proper object* of shame and its shamefulness the *improper object* of shame. Similarly, the way Hans has treated Maria is the proper object of indignation and injustice its improper object. The improper objects of shame and indignation are values, more exactly, value-properties or qualities. But the proper objects of shame and indignation are characterized without any reference to values. Within the philosophy of emotions two rival accounts of what I have called the improper objects of emotions can be distinguished. There is the view that emotions reveal or disclose their improper objects, value-properties. And there is the view that emotions are reactions to prior awareness of such value-properties.[5] Since I favour the second view, I employ it here, although what I say can be easily formulated in terms of the first view.

We are now almost ready to answer the question: what is happiness about when it is about something? The answer will rely on the assumption that emotions may be correct or incorrect and the assumption that emotions have both a proper and an improper object.

What are the proper and improper objects of happiness? Many ascriptions of happiness are of the form: *x* is happy about the fact that *p*. Sam is happy about the fact that he has many friends, is successful, is admired by his children, that he has children, this or that ability. These, it seems, are the proper objects of Sam's happiness. But is this answer correct? If Sam takes himself to have many friends, the appropriate emotional reaction will, for example, be gratitude towards his friends for their affection, or simply being pleased that he has so many friends. If Sam takes himself to be admired by his children, the appropriate emotional reaction is relief or pride or, again, being pleased by this fact. There is a possible view of happiness according to which it is a constellation of positive emotions such as being pleased, satisfied, pride, relief and so on.[6] On this view, the objects of happiness would be the objects of emotions other than happiness

[5] For the first view, cf. Tappolet 2000, Johnston 2001, Deonna and Teroni 2012. For the second view, cf. Mulligan 2009, 2010. The two views correspond to two alternatives within the framework of appraisal theories of emotions in psychology, cf. Mulligan & Scherer 2012.

[6] It is an analogue of the view of emotions put forward by the Genevese psychologist Claparède, in a discussion of James' view of emotions: an emotion is a *Gestalt* consisting of a variety of bodily sensations and feelings, rather than an unstructured sum of such feelings.

and the (in)correctness of happiness would be a function of the (in)correctness of emotions other than happiness.

If the proper object of Sam's gratitude is his friends and their affection for him, the improper object of his gratitude is the positive value to him of their friendship. If Sam's happiness is to be distinguished from such positive emotions as gratitude, pride or being pleased, it must have its own distinctive proper and improper objects. What might these be? My answer will make use of a further assumption. The two main assumptions introduced so far, the distinctions between correct and incorrect emotions and between the proper and improper objects of emotions, belong to the philosophy of mind. The assumption to be introduced now comes from the philosophy and logic of value.

Suppose something exemplifies some particular, positive value: an ornament is beautiful, an action is just, a face is pretty, a person is healthy, a landscape sublime, a gait or a handbag elegant, a wine pleasant, a novel interesting. Then, so the assumption, it is positively valuable, good, that each of these particular, positive values is exemplified. If something is beautiful, then it is good that that thing is beautiful. If an act is just, then it is good that that act exemplifies justice. (An alternative, weaker and in many ways more plausible claim is that if something exemplifies a positive value, then it is better that this is the case than not. I do not employ the weaker claim here simply because it requires formulations more complicated than those needed for the stronger claim.) Goodness, badness and betterness are sometimes called *thin* values, as opposed to *thick* values such as evil, justice, elegance, and pleasantness. Using this terminology, the principle may be formulated as follows: if some thick value is exemplified, that fact has thin value, it is good that this is the case. The principle might be called the principle of *axiological ascent*.

The proper object of happiness, I suggest, is the goodness of the exemplification of certain types of thick, positive value. Happiness differs in this respect from many other positive emotions. Emotions such as gratitude, being pleased, interest, respect and admiration have, in the simplest cases, persons and objects as their proper objects and a variety of thick values as their improper objects. But the proper object of Sam's happiness is not merely his children, their admiration, his abilities and success in life, nor the values of these. It is the positive value of the exemplification of different positive values by his children, their admiration and his abilities. One attractive feature of this view is that it allows for the fact that many different types of objects, creatures, relations and situations as well as many different thick values may go to make up the proper object of happiness.[7] Another attractive feature of the view is that it assigns a distinctive and invariable proper object to happiness. Unlike many other emotions, happiness has value as part of its proper object. A third feature of the view is that it does justice to the fact that happiness and unhappiness are reflective emotions, the result of standing back from one's life or aspects of it.

Are there emotions other than happiness the proper object of which is the thin, positive value of the exemplification of thick, positive values? One candidate is an emotion already mentioned as an emotion which is often said to make us happy, in a loose sense of the word—being moved.[8] What are we moved by?

[7] On this variety, cf. Kenny and Kenny 2006, Kazez 2007.
[8] Cf. Cova and Deonna 2014.

By the positive value of the exemplification of different thick positive values of, for example, the birth of a child, generosity, heroism, the affection of an ageing couple, weddings, flags—and a variety of kitsch objects, scenes and displays. The reaction to the birth of a child may be wonder, to a display of the flag patriotic stirrings, to the fidelity of an ageing couple, heroism or generosity admiration. But what moves us is the positive value of the exemplification of the different thick, positive values by these items.

If the proper object of happiness is the thin, positive value of the exemplification of certain types of thick, positive value, what is its improper object? To ask this question is, as we have seen, to ask what condition must hold if happiness is to be correct.[9] The assumptions introduced so far require that there be an answer to this question. I suggest that a person's happiness is correct only if its proper object really constitutes good luck for her. Good luck is always someone's good luck, the "objective" happiness or flourishing of a person, and it is the improper object of happiness. Happiness about this or that, whether the happiness is an episodic emotion or an enduring state, may be correct or incorrect because it has an improper object. Suppose Sam is happy about his abilities, his financial position or his relationships. His happiness is correct only if he is in fact lucky. The link between good luck or fortune in life and happiness as an emotion or state is marked in many languages. (*Glück*, like *heureux*, may refer to a psychological state or to good luck or fortune.) In English, a person's situation or prospects may be described as happy; Australia is the happy country and Austria-Hungary was *Felix Austria*. The relation between "subjective" happiness and "objective" happiness, as I have presented it, resembles the relations we have already considered between fear and danger, indignation and injustice, being ashamed and shamefulness. Just as one may be indignant about a situation which is not in fact unjust, so too one may be happy about the positive value of abilities and relationships and situations which are not in fact examples of one's good luck. For the proper object of one's happiness may hide great deceptions, ill-fortune and unluckiness: one's abilities may be tragically incompatible, children, friends and lovers may turn out to be treacherous, and ex-students and colleagues to be ungrateful or malevolent wretches. Indubitably good things may turn out to be *bad for one* or neither good nor bad for one, but *indifferent for one*. Novels provide an inexhaustible panorama of the ways in which apparent good luck may turn out to be merely apparent. If it is true that no one should be called happy until she is dead, it is precisely because apparent good luck may turn out to be no such thing or even very bad luck.[10] And if it is true that one may be un-

[9] Utilitarian accounts of the ethical rightness of actions traditionally understand the latter in terms of happiness or of happiness and its value. An utilitiarian account of ethical rightness in terms of correct happiness or of correct happiness and its value would lead to a drastic modification of traditional utilitarianism. Similarly, some accounts of well-being in terms of the satisfaction of desires or preferences distinguish between such satisfaction and the satisfaction of informed desires or ideal preferences, cf. de Sousa 1987: 167-69, Skorupski 1999: 130-133, Rodogno 2015.

[10] What is the difference between happiness and being moved? The value which is the proper object of being moved—for example, the positive value of the exemplification of courage—is an impersonal value. And being moved is correct only if the positive value is a high or important impersonal value. Being moved by kitsch is an example of incorrect emotion because the kitsch object or scene is good *for* the consumer of kitsch. Happiness,

lucky after death (slander, the destruction of a reputation), a person's death is not the right time for ascriptions of luck either.

Relying on four by no means uncontroversial assumptions, I have provided an account of the objects, proper and improper, of happiness. As far as I can see, the account given of the improper object of happiness, good luck, remains an attractive option even if the account given of the proper object of happiness is rejected. One virtue of the account presented above is that it does justice to the not uncommon view that happiness differs in several respects from most other positive emotions. That happiness differs from other positive emotions, that it is indeed is more important than other positive emotions, is perhaps one reason why it has enjoyed so much philosophical attention since Antiquity.

But philosophers and psychologists are rightly never very impressed by common views. Two important recent attempts to understand happiness in terms of pleasure are due to Kahneman and Feldman.[11] According to Feldman's carefully argued "reductive account of happiness",

> a person's momentary happiness level at a time is the amount of … attitudinal pleasure he takes in things at [a] time, minus the amount of attitudinal displeasure he takes in things at that time. On my account, to be happy is to be on balance attitudinally pleased about things.[12]

I implicitly rejected such an account when I suggested that being pleased may be said to make us happy in a loose sense of the word, but no more. But is Feldman really wrong? Is being happy not just being pleased? If we assume that each type of emotion has a distinctive value-property as its improper object, then it is plausible to think that being pleased stands to the value of pleasantness or agreeableness as happiness to the positive value of good luck. But this consideration will not move anyone who wants to understand emotions in abstraction from any account of values. Are there any reasons for thinking that happiness cannot be reduced to being pleased which are independent of our assumptions about the relations between emotions and values?

Being pleased and being displeased, unlike happiness, come in three distinct varieties:

Sam pleases Maria
The fact that she has passed her exam pleases Maria
Maria is pleased that she has passed her exam

The third example is an example of what Feldman calls attitudinal pleasure and a propositional attitude. The pleasure to which Feldman wants to reduce happiness is pleasure as a propositional attitude. The first example is not an example

on the other hand, is correct only if its proper object is a form of high value *for* the happy subject, personal rather than impersonal value. On personal value, being good *for* someone the seminal work is Ronnow-Rasmussen 2011.

[11] Kahneman 1999, Feldman 2010. For criticisms of such "hedonic" accounts of happiness, see Haybron 2008, Massin 2011. Feldman presents a number of criticisms of Kahneman's views. Kubovy 2015 presents an alternative to Kahneman's view that the value of a life is just the sum of the value of its moments, an alternative which in several respects is congenial to the view sketched here.

[12] Feldman 2010, 110.

of propositional pleasure. It is an example of the simplest form of intentionality. It resembles seeing someone and remembering someone rather than seeing that *p* or remembering that *p*. It is an example of what is sometimes called non-propositional intentionality. On some views, this simplest form of intentionality need not involve any exercise of thought or concepts. The second example has some of the features of the first example and some of the features of the third example. Like the first example, it has the form of a relation. Like the third example, it may be held to involve the attribution of some thought or conceptual representation. And "the fact that she has passed her exam" is clearly a nominalisation of "she has passed her exam". Happiness, then, does not display the same multiplicity as being pleased. One may be happy that *p* or happy about the fact that *p*. But happiness does not display the simplest form of intentionality. (If Sally is happy with Sam, then she is satisfied with him, perhaps by his performance in the office, or she is happy while she is with him.) And that is just what we would expect if the above account of the objects of happiness is correct.

The main reason for rejecting the reduction of happiness to being pleased which is independent of claims about the relation between emotions and values is that happiness, unlike being pleased, is always beyond our control. In the figurative formulations employed across the centuries to capture this point, happiness is said to be a gift, something which falls from heaven, something we stumble upon. It is, it is often said, non-figuratively, a *by-product*, something we should not aim at or cannot aim at or should not aim at because we cannot aim at it. Whatever the correct view about such claims is, it seems very plausible to say that correct happiness is much less easy to bring about than incorrect happiness. For good luck is by nature not something we can manipulate.[13]

2. Satisfaction and its Determinates

Happiness and being moved, I have argued, differ from many other positive emotions in that their proper object is the thin value of the exemplification of thick values. This is not surprising if we bear in mind that happiness and being moved are two ways of being satisfied.[14] Satisfaction, in this Rolling Stones sense, is not to be confused with the satisfaction of mental states and acts which figures prominently in the accounts of intentionality given by Husserl and Searle. The former has a polar opposite, dissatisfaction, the latter has no polar opposite. The former is personal and may be felt, the latter is impersonal. But the impression that one's desires, for example, have been satisfied or fulfilled (realized, *erfüllt*) is often the basis of the reaction of satisfaction (*Befriedigung*), in particular of felt satisfaction. Personal satisfaction may be based on the impersonal satisfaction not only of desires but also of drives, strivings, needs, intentions and projects, just as impersonal non-satisfaction of these may trigger personal dissatisfaction.

But other determinates of satisfaction do not depend on realizations. Thus there is the satisfaction of the vital state of well-being (as opposed to being ill, cf.

[13] Luck has been much discussed recently in ethics and epistemology. For a pioneering comparison of what philosophers and psychologists have said about luck and an attempt to pin down the notion, see Pritchard and Smith 2004.

[14] For the view that happiness is a species of satisfaction, see Tatarkiewicz 1976.

wohlsein vs *unwohlsein*) or vigour, the satisfaction of comfort and the satisfaction of bliss, of subjective non-intentional happiness.

Satisfaction and dissatisfaction can often be classified in terms of their proper objects. Consider, for example, self-directed dissatisfaction and its opposite. Self-dissatisfaction, the state based on the impression of one's worthlessness or disvalue, and often described in terms of the operations of conscience or of a super-ego, comprehends cases ranging from the non-realisation of projects and desires in which what dissatisfies is the fact that *I* have not realised this desire or that project, to states of affairs well beyond the subject's control, such as dissatisfaction rooted in the negative value properties one exemplifies. Self-satisfaction covers a similar range of cases. Some of the abilities which are the objects of self-satisfaction are innate, others are not. There is, for example, the satisfaction of the miser who is aware of his ability to buy anything he likes.

Suppose Sam feels worthless. He is a fool, a coward, mean and ugly and knows this. He is unhappy and so dissatisfied. What exactly is the proper object of his unhappiness? We may appeal to the type of consideration already employed. Sam's awareness of his foolishness, we may suppose, triggers self-contempt. His awareness of his cowardice triggers self-hatred. He is merely annoyed and angry and sometimes displeased by his meanness, and this only occasionally. His awareness of his ugliness is the source of recurrent bouts of self-pity. Finally, he bitterly regrets that he has rarely seized any opportunity to modify the traits he exemplifies.

Self-directed contempt, hate, pity, displeasure and anger, like regret, are not species of dissatisfaction or of unhappiness. If Sam is dissatisfied with himself, the proper object of his dissatisfaction is in part the various disvalues he takes himself to exemplify. But the full proper object of his dissatisfaction is the disvalue of his exemplification of the disvalues of foolishness, ugliness, meanness and cowardice. Sam might well be subject to reactions of self-contempt, self-hatred, self-pity without ever being dissatisfied with himself. Such dissatisfaction involves taking a step-backwards, reflection. Instead of simply hating himself, despising himself because he a useless, mean, coward, Sam forms the thought or has the impression that his exemplification of these various disvalues is itself a bad thing. Dissatisfaction with oneself is a cool emotion compared to the heat of self-hatred, self-contempt and self-pity. The latter are more exhausting than the former.

This brief account of satisfaction and dissatisfaction suggests the following diagnosis of some popular accounts of happiness. Happiness and satisfaction are not the same thing; happiness is a determinate of satisfaction.[15] The satisfaction due to the realisation of desires, projects and preferences is not happiness either, although such realisations may be an ingredient of happiness. Finally, as already indicated, if "well-being" has the meaning it has in ordinary language, then it is not the same thing as happiness either. It is a distinct species of satisfaction and at best an ingredient of happiness.

[15] On determinates, determinables, species and emotions, cf. Johansson 2001.

3. The Dangers of Happiness

Can happiness or its pursuit be harmful? Are positive emotions dangerous? Are they always a good thing? Recent discussion of happiness often seems to ignore the possibility that happiness might sometimes be, on balance, a bad thing. But economics tells us that the inability to defer gratification or "present happiness" will make one worse off. Are there other ways in which happiness or positive emotions in general may be harmful?

Incorrect beliefs can certainly be harmful. If someone knows that you believe what is not the case, she may be able to bankrupt you. Incorrect emotions and preferences are also harmful:

they are a waste of psychic energy;
they make you vulnerable;
they waste your time and that of others;
they motivate projects which are doomed to failure and lead to disappointment.

All incorrect emotions, preferences and beliefs may be harmful. But incorrect happiness is perhaps more harmful than any other sort of incorrect emotion. In order to make this claim plausible, let us begin by considering Agathe:

Agathe prefers her beliefs to be correct rather than incorrect and prefers her emotions and preferences to be correct rather than incorrect.
Agathe's actions and mental life are guided by these two preferences.
So Agathe is intellectually virtuous, an epistemic heroine or saint.

It is easy to *imagine* someone like Agathe. She is clearly an admirable person (although she is probably not very happy). But why call her *intellectually* virtuous? Because of the similarity between preferring correct to incorrect beliefs, on the one hand, and preferring correct emotions and preferences to their incorrect counterparts, on the other hand. The similarity consists in the fact that just as Agathe prefers her beliefs to track the way the world really is, so too, she prefers her emotions and preferences to track the way the world of values and of what has value, positive or negative, really is. She prefers fear of dangerous dogs to fear of non-dangerous dogs, indignation about unjust situations to indignation about situations which are not unjust. And so on. To care about tracking the way the world is, the world of fact and of values, is a central component of intellectual or cognitive, as opposed to ethical, virtue. Care of this kind and preferences like those of Agathe are consistent with stupidity, with being slow. But when they successfully guide a person's actions and mental lives they are inconsistent with foolishness. Stupidity is not foolishness. The opposite of stupidity is intelligence, of foolishness wisdom.[16] If a philosopher loves wisdom, she should perhaps also hate foolishness.

Some people, everyone will admit, are not like Agathe. They often prefer incorrect beliefs and emotions. In particular, they prefer incorrect, self-flattering beliefs. And they often seem to prefer incorrect, self-flattering beliefs because this allows them to revel in the emotional reactions which would be appropriate to these beliefs, were they correct. People who can believe what they want, says Lichtenberg, are happy creatures. Of course, those of us who do not resemble

[16] Cf. Mulligan 2014.

Agathe do not typically prefer incorrect beliefs, preferences and emotions *under this description*; we do not typically set out to form incorrect beliefs and preferences. As Aquinas put it, no one wants to be a fool, but everyone wants the consequences of foolishness (*stultitia*). Many accounts of many varieties of foolishness—vanity, sour grapes and sweet lemons (sloes)—rely on the attractiveness of incorrect beliefs and impressions and of the incorrect emotions and preferences to which they give rise. Thus the vain man wants to be applauded and praised whether or not the applause and praise are justified. Just how the relation between incorrect beliefs, emotions and preferences and self-deception should be understood is a very controversial matter in both philosophy and psychology.[17] But if we assume that some of us differ from Agathe in the ways described, then it becomes plausible to think that incorrect happiness can be more harmful than all other incorrect emotions. Insofar as one prefers that one be lucky in life rather than unfortunate one is tempted to form incorrect beliefs about one's situation, life, relations and capacities, about matters well beyond one's control. For such beliefs are constituents of felt or subjective happiness. The illusion that one is lucky in life, that Fortuna is on one's side, is one of the worst illusions, one of the *worst* things *for* a person, just because being lucky in life is the highest form which positive personal value, being good for a person, can take.[18]

References

Cova, F. and Deonna, J.A. 2014, "Being Moved", *Philosophical Studies*, 169 (3), 447-66.

Deonna, J. and Teroni, F. 2012, *The Emotions. A Philosophical Introduction,* London: Routledge.

De Sousa, R. 1987, *The Rationality of Emotions*, Cambridge, MA: MIT Press.

Feldman, F. 2010, *What Is This Thing Called Happiness?*, Oxford: Oxford University Press.

Haybron, D.M. 2008, *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*, Oxford: Oxford University Press.

Johansson, I. 2001, "Species and Dimensions of Pleasure", *Metaphysica*, 2 (2), 39-72.

Johnston, M. 2001, "The Authority of Affect", *Philosophy and Phenomenological Research*, LXIII, I, 181-214.

Kahneman, D. 1999, "Objective Happiness", in Kahneman & al. 1999, 3-25.

Kahneman, D., Diener, E. and Schwarz, N. (eds.) 1999, *Well-Being: The Foundations of Hedonic Psychology*, New York: Russell Sage Foundation Press.

[17] Trivers 2011 (4, 68 ff., 135 ff.) thinks there is not much truth in the idea that self-deception makes us happy. Its function is to make us better at deceiving others.

[18] Thanks to Riccardo Braglia, CEO and Managing Director, *Helsinn Holding SA* and the *Fondazione Reginaldus* (Lugano) for financial support of the work published here. To Julien Deonna, Ingvar Johansson, Hélène Leblanc, Olivier Massin, Graham Oddie, Raffaele Rodogno, Toni Ronnow-Rasmussen and Fabrice Teroni for criticisms and discussion. And to audiences in Basel (1996), Ascona (2014), Milan (2015) and elsewhere for their reactions.

Kazez, J. 2007, *The Weight of Things: Philosophy and the Good Life*, Oxford: Blackwell.

Kenny, A. and Kenny, C. 2006, *Life, Liberty & the Pursuit of Utility*. *Happiness in Philosophical and Economic Thought*, Exeter: Imprint Academic.

Kubovy, M. 2015, "A Deep Structure of Lives: A Prolegomenon to an Ontology", *Philosophia Scientiae*, 19, 3, 5-28.

Massin, O. 2011, "Joies amères et douces peines", in Tappolet, Ch., Teroni, A.F. and Konzelmann-Ziv, A. (eds.), *Les ombres de l'âme, penser les émotions négatives*, Geneva: Markus Haller, 55-70.

Mulligan, K. 2009, "On Being Struck by Value – Exclamations, Motivations and Vocations", in Merkel B. (ed.), *Leben mit Gefühlen. Emotionen, Werte und ihre Kritik*, Paderborn: mentis Verlag, 141-61.

Mulligan, K. 2010, "Emotions and Values", in Goldie, P. (ed.), *The Oxford Handbook of Philosophy of Emotion*, Oxford: Oxford University Press, 475-500.

Mulligan, K. 2014, "Foolishness, Stupidity and Cognitive Values", *The Monist*, Nanay B. (ed.), *The Philosophy of Robert Musil*, 97, 1, 66-85.

Mulligan, K. 2015, "Is Preference Primitive?", in Persson, J., Hermerén, G. and Sjöstrand, E. (eds.), *Against boredom*. *17 Essays presented to Nils-Eric Sahlin*, Lund: Fri tanke förlag, 169-80.

Mulligan, K. and Scherer, K.R. 2012, "Toward a Working Definition of Emotion", *Emotion Review*, 4, 345-57.

Pritchard, D. and Smith, M. 2004, "The Psychology and Philosophy of Luck", *New Ideas in Psychology*, 22, 1-28.

Roberts, R. 2003, *Emotions. An Essay in Aid of Moral Psychology*, Cambridge: Cambridge University Press.

Rodogno, R. 2015, "Prudential Value or Well-Being", in Brosch, T. and Sanders, D. (eds.), *Oxford Handbook of Value: The Affective Sciences of Values and Valuation*, Oxford: Oxford University Press, 287-311.

Ronnow-Rasmussen, T. 2011, *Personal Value*, Oxford: Oxford University Press.

Skorupski, J. 1999, *Ethical Explorations*, Oxford: Oxford University Press.

Tatarkiewicz, W. 1976, *Über das Glück*, Stuttgart: Klett Verlag.

Tappolet, Ch. 2000, *Emotions et Valeurs*, Paris: Presses Universitaires de France.

Trivers, R. 2011, *The Folly of Fools*. *The Logic of Deceit and Self-Deception in Human Life*, New York: Basic Books.

The Democratic Riddle

Philip Pettit

Princeton University and Australian National University

Abstract

Democracy means popular control, by almost all accounts. And by almost all accounts democracy entails legitimacy. But popular control, at least as that is understood in many discussions, does not entail legitimacy. So something has got to give. Democratic theories divide on what this is, so that the question prompts a taxonomy of approaches. The most appealing answer, so the paper suggests, involves a reinterpretation of the notion of popular control.

Keywords: Democracy, Control, Legitimacy

There are three plausible assumptions that are commonly made about democracy and that we should be loath to reject; they are axioms of democratic discussion. Yet those assumptions are inconsistent with one another, and constitute a riddle for democratic theory. This paper presents and motivates those assumptions and shows how standard approaches to democracy resolve the riddle in an unsatisfactory *ad hoc* manner, restoring consistency by the blunt rejection of one or another axiom. The paper goes on to outline a different strategy of resolution, involving the reinterpretation rather than the rejection of an assumption, and it uses this to identify a novel way of conceptualizing democracy as a regime of deliberative regulation. Under this conception the assumptions become jointly consistent, while remaining individually plausible.

1. The riddle

The assumptions that give rise to the riddle are these:

- *The equivalence axiom*: democracy means popular control.
- *The legitimacy axiom*: democracy entails legitimacy.
- *The tyranny axiom*: popular control does not entail legitimacy.

Argumenta 1, 2 (2016): 147-160
ISSN 2465-2334

© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.PET

These assumptions are clearly inconsistent with one another. Two items, democracy and popular control, are presented as logically equivalent; that is how 'means' is to be understood. Yet one of those items, democracy, is said to entail legitimacy whereas the other, popular control, is said not to entail legitimacy. Clearly something has got to give: one assumption has to be rejected or the assumptions have to be reinterpreted so that consistency is restored. But the assumptions all look to be unambiguous and compelling. And so the inconsistency between them presents something of a riddle.

The first axiom elaborates on the meaning of democracy, as that is registered in the very etymology of the word. 'Democracy' comes from two Greek words, 'demos' meaning people, and 'kratos' meaning power. The idea in any democracy is that the people have influence, where the extension of the people may be taken as given and unproblematic; specifically, the people have influence in or over the exercise of government. The influence they have must not be partial and wayward, however, like the influence of the weather on government. It must be significant in degree, and it must be directed to a systematic, equally welcome end; otherwise it will not give the people the requisite power or control. The first, equivalence axiom equates democracy with a regime in which the people have an influence of that controlling sort.

The second axiom formulates the intuition behind a standard move in democratic debate. I will make this move if I argue in minimal defense of a law or policy that since it was introduced under democratic arrangements, it has the backing of a legitimate regime and you ought to recognize this. The move is hard to resist and that is why the second assumption has a natural appeal. The assumption does not offer a full defense for a given law or policy, of course, for in agreeing that the regime is legitimate, you do not necessarily agree that its laws and policies are invariably just or for the best. You will have to acknowledge is that since they were adopted within the regime—and, we may assume, in accord with the regime's procedures—the government has the right to enact and enforce its laws and policies, and you will have to accept your obligation, if you are to oppose them, to do so within the system.[1]

Opposing a decision within the system will mean exploiting the resources of democracy to challenge it in the courts, in the forums of politics, in the media, or even on the streets. At the limit it may mean resorting to civil disobedience. Such disobedience remains within the system insofar as you accept the right of the government to penalize you for disobeying. By displaying such an acceptance of the regime—and therefore an acceptance of censure and punishment—you underline the intensity of your feeling and, in consequence, the strength of your claim to be given a democratic hearing.[2]

[1] Notice that I define the legitimacy of a regime without connecting it to an obligation, pro tanto or otherwise, to comply with the dictates of the regime. For the record, I think that such legitimacy connects rather with an obligation, if you oppose the laws, to oppose them within the system. See Pettit 2012 and on related matters Simmons 1979.

[2] It may also be best to oppose an illegitimate regime within the system, of course, especially via civil disobedience. Civil disobedience in this case would mean treating the regime as if it

I describe the third axiom as one of tyranny because it picks up the common claim that popular control may involve a tyranny of the majority. This claim is as old as democracy itself, marking the fact that a majority may serve its own interests in government, and act in a way that treats the members of one or another minority as less than equal. The word 'tyranny' is significant for it clearly suggests that a system of popular control may be, not just suboptimal or unjust, but downright illegitimate. The idea behind the assumption is that popular control does not entail legitimacy, precisely because certain majoritarian decisions may systematically offend against a minority in a way that makes them illegitimate.

2. Standard responses

There are three standard responses to the democratic riddle. More exactly, there are three standard positions in democratic theory that present different responses to the riddle, each denying a different axiom. For reasons that will become clear in discussion, I describe them as populism, liberalism and eclecticism.

The populist response accepts that democracy means popular control and that it also entails legitimacy but avoids the riddle by denying the third, tyranny assumption and asserting that a system implements popular control then it is a legitimate regime. The idea is to live with the logic of majoritarianism and embrace the conclusion that majority rule is always legitimate rule, even if it offends against a minority in some way. The position is populist insofar as it prioritizes the collective rights of the people over the individual rights of members, including members in minority categories.

The liberal response accepts that democracy means popular control, as in the first axiom, and that popular control does not entail legitimacy, as in the third, but avoids the riddle by rejecting the second assumption that democracy entails legitimacy. I describe this response as liberal, because it is associated in particular with William Riker's (1982) case for liberalism against populism. He accepts the equation between democracy and popular control but argues that as it is exemplified in standard voting systems, popular control does not guarantee legitimacy; it is consistent with a violation of minority rights that would, intuitively, make a regime illegitimate. Thus he embraces the conclusion that democracy does not entail legitimacy. In order for a regime to count as legitimate, he suggests, public decision-making will not only have to emanate from a democratic source but also respect individual rights on the part of those whom it affects. Democracy, on this approach, is not a guarantor of legitimacy. It will have to be balanced by a counter-majoritarian regime

were legitimate, while not thinking that it actually is. The tactic might put pressures in place that would push the regime towards an increasingly legitimate pattern of organization and behavior. Holding an individual or institution responsible when they are not properly fit to be held responsible may help to make them responsible; it may 'responsibilize' them (Garland 2001; Pettit 2007). Similarly treating a regime as legitimate, say in civil disobedience, may elicit support from others and expose the government to publicity in such a way that the system performs in an ever more legitimate pattern.

of individual rights in any legitimate system; this regime will operate via a constitution that restricts the scope of the democratic will.

Where populism rejects the third assumption, and liberalism the second, eclecticism rejects the first. It argues that democracy should not be equated with popular control but only with an eclectic dispensation of popular control constrained by an independent factor: usually, some set of individual rights. It embraces a compound conception of democracy, as it has been called (Mayerfield 2009). In order to count as democratic a regime will have to display both the control element and the constraint element; the control element alone will not ensure democracy. This being so, the second and the third axioms can stand; democracy in the eclectic sense will entail legitimacy but popular control on its own will not. This position is quite commonly adopted in contemporary legal and political theory. It is exemplified in the view of democracy taken by Ronald Dworkin (2006: 131) when he writes: 'a community that steadily ignores the interests of some minority or other group is just for that reason not democratic'.

The pattern in the three positions given is straightforward. Each affirms two of the axioms and derives the falsity of the third from the two affirmed. Eclecticism holds that democracy entails legitimacy but that popular control does not and concludes that democracy is not equivalent to popular control: it also requires a regime of individual right. Liberalism holds that democracy is equivalent to popular control but that popular control does not entail legitimacy and concludes that democracy does not entail legitimacy either; the respect for certain rights is also required. Populism holds that democracy is equivalent to popular control and that democracy entails legitimacy and so concludes that popular control entails legitimacy. The positions map out as follows.

	Equivalence axiom	Legitimacy axiom	Tyranny axiom
Eclecticism	No	Yes	Yes
Liberalism	Yes	No	Yes
Populism	Yes	Yes	No

Should we be attracted by any of these positions? I do not think so. They each suffer from a glaring deficiency, which is that the presumptive justification for rejecting the targeted axiom in each case is just that doing so will resolve the riddle; the rejection is in that sense ad hoc. Eclecticism appeals on the grounds that it saves the legitimacy and tyranny axioms, liberalism on the grounds that it saves the equivalence and tyranny assumptions, and populism on the grounds that it saves the equivalence and legitimacy assumptions.

This might not be a problem for any of these theories if the axiom it rejected were not persuasive. But, intuitively, each of the axioms is quite compelling. The equivalence axiom spells out the core idea in democracy that the people rule by exercising control in or over government. Rejecting it by requiring that the people rule

according to an independently given set of rights looks like redefining democracy to one's purposes: in effect, rigging the books. Things would be different if the argument were that the very notion of popular control, in a relevant sense, requires a regime of rights; this is the sort of view for which I argue later.[3] The eclectic position is that democracy is not just a matter of popular control, despite the etymology and the meaning of the word; it requires an extra element that has nothing to do with the notion of control.[4]

The legitimacy axiom articulates an assumption that figures on two fronts in regular democratic discussion. It appears, first, in the common assumption that if you admit that a regime is democratic but deny that it is legitimate, then you are an elitist who rejects the terms of democratic exchange. You say that it is not enough for legitimacy the people rule; the people must rule rightly: that is, rule rightly by the lights of a presumptively higher authority. The legitimacy axiom appears, second, in the assumption commonly made in the classification of regimes that if a system of government is intuitively illegitimate, say because of not giving independence to the courts, then it is not properly democratic. If illegitimacy entails the absence of democracy, then democracy must entail legitimacy.

The intuitive grip of the tyranny assumption is probably even stronger than the grip of either of the other axioms. This axiom figures prominently in the canonical commentaries on democratic practice, classical, modern and contemporary. It has always seemed obvious that the rule of the many can be deleterious for the life of the few and it has always been taken for granted that any systematic abuse of the few will make for illegitimacy. Electoral control can make for a tyranny of a majority, where a tyranny is the very paradigm of an illegitimate regime.

The three responses rehearsed here, and the criticisms just leveled against them, are most naturally interpreted in terms of the electoral or representative conception of the people as the members of the society, considered as independent subjects. But it is worth noticing that they also apply under the rival, participatory conception of the people as an incorporated assembly in which members are committed to acting as one. This is the people considered en bloc, rather than en masse, in the corporate image under which Hobbes (1994) and Rousseau (1973) depicted the sovereign people. Hobbes thought that the people could legitimately assume this self-incorporating, sovereign status, though he did not think that was necessary for legitimate government; in fact he preferred the arrangement under which the people were represented by an individual sovereign. In contrast, of course, Rousseau held that the people in assembly constituted the only legitimate sovereign.

[3] For the record, Ronald Dworkin's phrasing occasionally suggests that individual rights should be introduced, not as an independent constraint on popular rule, but as a constraint required by the very notion of popular rule itself.

[4] Stephen Holmes (1995: 206), who adopts something like the eclectic position, softens it by arguing that while the notion of democracy does not require a regime of individual rights, the organization of democracy does; the idea is that a democratic regime would be overburdened with decision-making if individual rights did not put a broad range of questions off the table. This is probably the best defense possible for what remains a distinctively eclectic line.

As interpreted in the electoral model, eclecticism says that it is not enough for democracy that the people are in electoral control, liberalism that while this is enough for democracy it does not ensure legitimacy, and populism that electoral control does not allow for an illegitimate tyranny of the majority. But those claims are also defined under the assembly model. Eclecticism would say that it is not enough for democracy that the people rule in assembly; liberalism that while assembly rule is enough for democracy it does not guarantee legitimacy; and populism, most brazenly, that assembly control does not allow an illegitimate tyranny of the majority.

3. The reinterpretation strategy identified

Each of the three positions identified is tenable but in view of the problems raised, none is particularly attractive. Eclecticism requires a rigged redefinition of democracy, liberalism a rejection of the fundamental idea that democratic approval is a guarantee of legitimacy, and populism an acceptance of the repugnant idea that no collective impositions can intrude illegitimately in the lives of individuals. So where, then, are we to turn?

The strategies discussed each involve the rejection of one assumption, with the cost attendant on letting that assumption go. An alternative strategy for resolving the riddle would support reinterpretation rather than rejection. It would find a base for arguing that, properly understood, the assumptions are individually true and yet jointly consistent. So is there any hope on this front?

If the reinterpretation strategy is to work, then at least one of the crucial terms—'democracy', 'popular control' or 'legitimacy'—has to have a different sense in the two axioms where it figures. But 'democracy' and 'legitimacy' are used in a way that is supposed to be appropriate, no matter how the term is interpreted; the idea is that however democracy is understood it involves popular control and entails legitimacy, and that however legitimacy is understood it is entailed by democracy and not entailed by popular control. Thus the only hope for the reinterpretation strategy is to be able to argue that 'popular control' has one sense in the equivalence axiom and another in the tyranny axiom.

In the tyranny axiom popular control is clearly associated with the rule of numbers—roughly, with majoritarian rule; and taken in that way, the axiom is clearly true. So that means that the reinterpretative strategy will work just in case the idea of popular control in the equivalence assumption can be interpreted in a non-majoritarian manner; specifically, in a manner that preserves the entailment from democracy to legitimacy. If popular control is understood in one way within the equivalence assumption, and in another within the tyranny assumption, then there will be no problem of inconsistency among the three assumptions.

We saw earlier that if control is exercised democratically by the people, then the people must exercise a significant degree of influence on government and the influence must be systematically directed to a welcome, shared end. The majoritarian exercise of influence can certainly be significant, whether in the electorate or in an assembly, and the reason it is problematic is the likelihood of its being directed systematically to a majoritarian, unshared end: the satisfaction of the many, in relative

indifference to the few. This end may be systematically furthered because voters act as factions for group ends and the end of the majority faction wins out, or because voters each act as individuals for their own ends and the end sought by most—that is, by the majority faction—is successful. But if the tyranny assumption interprets popular control as majoritarian influence for majoritarian ends, then perhaps we can understand control in the equivalence assumption as a form of influence that differs from any such majoritarian process.

If we can do this plausibly, then we can run a credible version of the reinterpretative strategy. We will be able to recast the three axioms in a form that makes them individually persuasive but collectively consistent. The formulation will look like this.

- *The equivalence axiom*: democracy means popular, non-majoritarian control.
- *The legitimacy axiom*: democracy entails legitimacy.
- *The tyranny axiom*: popular control does not in itself entail legitimacy.

Is there a plausible interpretation of popular, non-majoritarian control such that we can think of that control as answering to the idea of democracy? The challenge is to find a non-majoritarian means-end package that would allow us to say that the people exercise the popular control associated with democracy when they use those means for that end. In particular, the challenge is to find a package that we can hope to implement under a suitable institutional design; it should not be a utopian package that feasible institutions would be unable to promote reliably.

Once the question is posed in this way, it becomes clear that there is at least one candidate answer available in the literature. It is found in the assembly model put forward by Rousseau in *The Social Contract*, published in 1762. Rousseau thinks that, strictly speaking, democracy requires that the people not only legislate in assembly, as his assembly model has them do, but also that they should act as an executive of their own laws—something that he thinks would not be desirable. But, putting aside his terminological idiosyncracy, we can reasonably treat his assembly rule as a model of democracy.[5]

Rousseau accepts the second axiom according to which democracy in the relevant, assembly sense entails legitimacy and he also agrees, in line with republican tradition, with the third axiom that popular control on its own—majoritarian, assembly control—does not entail legitimacy. What he does in order to resolve the riddle is to reinterpret the first, equivalence assumption according to which democracy is equivalent to popular, assembly control. He argues that this assumption is true only when popular, assembly control is constrained so that it is exercised by the assembly as a whole in pursuit of the will of the assembly as a whole—the general or

[5] Hobbes (1994) takes a different line, of course. He argues that the third axiom should be rejected insofar as he maintains that there is no such thing as tyranny; any individual or body that exercises sovereign rule effectively, guarding against the state of nature, is a legitimate ruler. This position derives from an overall theory of the nature of the state or commonwealth. Hobbes thinks that democracy might give rise to legitimacy, without falling foul of the democratic riddle, but he also thinks that other effective regimes would be equally legitimate.

corporate will. Constrained to that effect, the assembly decisions will be proof against the sub-corporate or particular interests of factions or coalitions. Thus Rousseau can represent the equivalence axiom as consistent with the assumption that assembly control as such, since it can be unconstrained, does not entail legitimacy. He recasts the three axioms in the following consistent form.

- *The equivalence axiom*: democracy means suitably constrained, assembly control.
- *The legitimacy axiom*: democracy entails legitimacy.
- *The tyranny axiom*: assembly control does not in itself entail legitimacy.

Rousseau's resolution is not of great interest in its own right, for at least three reasons. First, it presupposes, unrealistically, that an assembly of all the members of a society can take charge of government; there isn't a plausible version that would apply in an electoral rather than an assembly democracy. Second, the corporate or general will that is to be enacted in democratic decision-making remains an obscure target.[6] And third, the measures proposed for constraining the assembly, allowing the enactment of the general will, are just not very persuasive. One proposal is that the assembly should vote only on general matters, not on issues of special interest to some individuals, and the other that there should be no discussion of the kind that might lead to the formation of faction. It is not clear how either measure could be reliably implemented.

The key idea in the Rousseauvian approach is that the control exercised by the people should be constrained so as to reflect their corporate or general will, where this need not always be reflected properly in majority voting. The idea is that the means of popular control should be constrained so as to target the enactment of that will, thereby avoiding a tyranny of the majority and implementing a regime that ensures legitimacy. Given that the Rousseauvian general will does not look like a feasible target on which to train popular control, however, is there any alternative available to do the same job? What other popular ends, and associated popular means, might be identified as essential to the sort of popular control that is equated in the first axiom with democracy?

[6] Actually the preference or will of a group agent—the corporate will—cannot be identified with that which a majority supports. Take simple on-off preferences over whether p, whether q, and whether p&q, among a group of three individuals, A, B and C; the scenario parallels a scenario with judgments that is later discussed in the text. A and C may prefer that p, so that a majority support it, with only B against. B and C may prefer that q, so that a majority support it, with only A against. But A and B will prefer in consistency that not-p&q, so that a majority will prefer that p, a majority that q, and a majority that not-p&q. The only resolution for the group, if it is to act as a corporate agent, is to go against the majority preference on some matter: say, as a group, to form a will that p, that q and also, against the majority, that not-p&q. See (Pettit 2001, Ch.5; List and Pettit 2002; Pettit 2003b; List 2006; List and Pettit 2011). Rousseau might welcome this result insofar as it shows that the corporate will is not necessarily identical with the majority will. But of course the result also shows, contrary to Rousseau's claims, that the corporate will need not be identical with the majority will, even when that will is formed under suitable constraints. And it does not show that the corporate will, so formed, will be free of factional influence.

4. The reinterpretation strategy applied

The most salient alternative to making the general will into the target of popular control is suggested by the traditional opposition between will and reason. One of the great divides in thinking about the control that an individual subject has in his or her life is between the voluntarist school of thought that associates it with the control of the will, however that is conceived, and the intellectualist approach that would identify it with the control of reason. This suggests, then, that if we are to find an alternative to the Rousseauvian approach that can evade the democratic riddle in a parallel way, then we should think about the possibility of a form of popular control that is designed, in the means and ends adopted, to implement something that we might think of as popular reason.

The idea of a reason-based system of popular control divides into two varieties, depending on whether the system is meant to empower the reasoning of people about what they should collectively do, or to empower the reasons that they validate as considerations pertinent to questions of what they should collectively do. The term 'deliberative democracy' is generally used of the first of these proposals, though the proposal is often applied only to democracy at one or another local site: in a town meeting, or voluntary association, or formal committee (Cohen 1989; Gutmann and Thompson 1996; Bohman and Rehg 1997; Elster 1998). The second proposal, which I find more attractive, is sometimes in the background of discussions of deliberative democracy but is rarely spelled out properly. Where the first argues for a deliberatively conducted democracy, the second comes out in favor of a deliberatively regulated democracy. It puts the emphasis, not primarily on the exercise of reasoning among people, but on the fact that government, be it representative or participatory, is required to account for its decisions on the basis of reasons that are invoked in popular debate and pass muster among the people generally.

Those who espouse a deliberatively conducted democracy recommend reasoning among members of the relevant group, in particular reasoning about what option in a given decision will serve the group well, not about what will promote their self-interest (Coleman and Ferejohn 1986). But since democracy never escapes disagreement, this individual deliberation will have to give way at some point to voting. And since voting on connected issues is always liable to generate an inconsistent set of judgments on the part of the group, the ideal of deliberative democracy requires members to be prepared to deliberate as a group, not just individually, about how best to resolve such inconsistencies. Such collective deliberation will be impossible outside of a relatively small group, however, and so the strict ideal of a deliberatively conducted democracy is bound to be infeasible at the national level (Pettit 2003a).

The only point in this argument that may be obscure is the claim that majority voting on connected issues is liable to generate inconsistent group judgments. But the point is demonstrable (Pettit 2001: Ch. 5; List and Pettit 2002; Pettit 2003b; List and Pettit 2011). In order to illustrate it, let a group of just three people, A, B and C have to vote, at the same or different times, on whether p, whether q, and whether p&q. A and C may vote that p, B that not-p. B and C may vote that q, A that not-q. And so a majority, A and B, will vote in consistency that not-p&q; A thinks that not-q, after all, and B that not-p. The group, if it goes on straight majority voting,

will have to hold that p, that q and that not-p&q. No sensible group agent will be happy to settle for such an inconsistent set of judgments, however, at least not if the judgments relate closely to decision-making. And so the group will have to deliberate as a group about which of the majority judgments to reject in order to ensure that the group will act on a coherent vision.

But if the idea of popular control via deliberative conduct does not look feasible, what of the related idea of popular control via deliberative regulation: that is, regulation by popular reasons? In order to introduce this possibility, we need to answer three questions. What are popular reasons? How might they regulate government? And would such a system of democratic regulation enable us to resolve the democratic riddle?

First question

Suppose, as routinely happens in an electoral democracy, that the members of a society indulge in constant debates about what they or their government should do in this or that predicament. We may think of the discussions as occurring on shop floors, office corridors and public houses, as well as in more formal debates in the media, at party meetings or on the floor of parliament or congress. Such discussions do not normally come to blows in a successful, ongoing democracy, even as they cross different lines of division among members and generate quite sharp disagreements. That this is so means that participants must be able to identify considerations that they can each present in arguing for their views, and expect others to find relevant or pertinent to collective issues. The disagreements will be generated by people's weighting those considerations differently or by their having different views on related matters of fact—or, of course, by failures of coherence or logic on one or another side. But however much disagreement materializes from such sources, the fact that discussions continue among the different parties, not giving way to straight conflict, means that the parties must be able to find or forge reasons for their proposals that pass muster on all sides as considerations that may be taken into account in collective decision-making. Those reasons will be relevant, by all lights, to resolving issues of what the collectivity should do; they will not be reasons that cite sectional interests only, or that depend on beliefs that are held only in some sections of the community, or that fail to connect in some other way with shared concerns.

As parties learn their way around in discussions about collective decisions, including discussions outside their immediate circle or across different circles in the society, they are bound to familiarize themselves with the considerations that have this standing. Those considerations will count as reasons for or against certain collective decisions that each recognizes as relevant, each recognizes as having this recognized status, and so on in the usual hierarchy. They will constitute a shared currency, in principle capable of constant development, for the evaluation and regulation of collective decision-making.

How variable are the common or popular reasons that we might expect to emerge in a society and serve in the regulation of government? Some reasons will have to be recognized as a condition for the very possibility of a currency of popular reasons emerging, developing and gaining regulative influence. These will include

considerations of respect that rule against treating people as something less than conversable persons who can be reached by reasoning—they will argue against restricting free speech, for example, and in favor of limiting the reach of government—as well as equality considerations that rule against treating any parties as worthy of less than equal respect; these may rule against denying them equal rights of speech and association, for example, or against denying them equal immunity to certain government actions. Other considerations will be more or less universally compelling, if not required as conditions for the very possibility of deliberative regulation, such as those that rule in favor of providing for the basic education, material welfare and legal security that are essential for commanding respect. Others again may be culturally variable considerations, bearing on limits to private property, or the position of a certain religion in public life, or the need for cultural homogeneity.[7]

Second question

How might popular reasons of this kind be empowered in political life and be used to regulate government? Elections are clearly going to be of the first importance, both for guarding against takeover by a dynasty or clique, and for providing a context for the sort of public discussion and deliberation that allows popular reasons to materialize and gain recognition.[8] But equally clearly there will have to be non-electoral safeguards and institutions that help to ensure, first, that no collective decisions are made in favor of options that popular reasons would outlaw; and, second, that collective decisions between other, more or less acceptable options are made on the basis of procedures that popular reasons would support: depending on context, the procedures supported may be a parliamentary vote, referral to an impartial or expert body, a general referendum, or whatever.

The non-electoral devices used will be various, reflecting the requirements of implementing deliberative regulation and the influences of the particular reasons validated in common exchange. They are almost certain to include rule-of-law constraints on how government acts; individual-right constraints on what it does; institutional restrictions such as the separation of powers, accountability measures and

[7] John Rawls (1999) may often have popular reasons in mind when he speaks of public reasons and my ideas have clearly been influenced by his discussion. I prefer to speak of popular reasons, emphasizing points that are not made in Rawls and might even be rejected by him: first, that they are generated as a byproduct of ongoing debate; second, that they are relevant to such debate, no matter at what site it occurs, private or public, informal or formal; and third that in principle the popular reasons that operate in a society, or even in the international public world, may not be reasons that carry independent moral force: we may disapprove of their having the role they are given in debate. The language of popular reasons, as used here, may be more in the spirit of Habermas (1984, 1989) than Rawls. I am grateful for a discussion on this topic with Tim Scanlon.

[8] Knights 2005 makes a persuasive case that it was the frequency of elections in the late seventeenth and early eighteenth century, together with the existence of a culture of pamphlets, journals and coffee houses, that gave rise to the formation of a discursive politics: a politics in which interest-based factions gave rise to ideological parties that focused on deliberatively negotiable differences public policy.

the depoliticization of certain decisions; and, perhaps most important, exposure to a power of effective invigilation and contestation on the part of ordinary people and their representatives. Electoral institutions will combine with these arrangements to generate a characteristically democratic dynamic of exchange between those in government, those out of government, and ordinary people. If the dynamic works well, then it will help to ensure that government does not act against popular reasons and that when popular reasons require that it act, then it actually does so. It does not violate the requirements of popular reasons either in the way it acts or in the way it fails to act.

Third question

Does the model of a deliberatively regulated regime give us a plausible image of popular control and enable us to resolve the democratic riddle? I believe it does. It involves a system in which people exercise influence on government in many different forums, as we have seen, and at many different levels. All going moderately well, this influence ought to have the systematic effect, though perhaps only over the long run, of disciplining government so that policies or processes that offend against popular reasons get to be sidelined. And that effect ought to be welcome on all sides insofar as it is generally accepted that such reasons should prevail in public life; this acceptance will show up in how people expect others, including those in government, to behave in pushing their various causes.

If the electoral and non-electoral institutions that prevail by courtesy of general acceptance do actually facilitate the deliberative regulation of government, therefore, then it is plausible to think of the network that they constitute as mediating the control of the people. Various actors will play a part within that network, ranging from citizens in their electoral and contestatory roles to those they elect to legislative or executive office, to those who are appointed in judicial roles, to those on central banks or electoral commissions to whom public decisions are outsourced. The network involved will generally sustain a global, more or less coherent pattern of outcomes—else the society would be anarchic—and in doing so it will have to be faithful to the sorts of reasons that are valorized in public discussion; it will have to rule out policies or processes that are inconsistent with such considerations.

We are now in a position to see how the approach sketched here can serve to resolve the democratic riddle in the reinterpretative manner that parallels the otherwise very different approach found in Rousseau. Deliberative regulation counts as a form of popular control, by the argument just given, and the popular reasons it empowers promise to impose a regime that ensures legitimacy; certainly it does not hold out the danger of a tyranny of the majority. Thus we can reinterpret the three democratic axioms in a way that makes them individually persuasive and jointly consistent.

- *The equivalence axiom*: democracy means popular, deliberatively regulated control.
- *The legitimacy axiom*: democracy entails legitimacy.
- *The tyranny axiom*: popular control does not in itself entail legitimacy.

The point of the discussion in this last section has not been to provide a proper defense of the idea of a deliberatively regulated democracy, or of the institutional means whereby it might be advanced; I tackle those goals in (Pettit 2012). I think that it is important to have a conception of democracy that solves the democratic riddle and my only wish has been to show that there is a candidate that promises to be able to achieve this goal: one that stands in for a family of candidates, indeed, since deliberative regulation can be variously interpreted.[9] Democracy should not be thought of in purely electoral, majoritarian terms, despite the preference of many professionals (Przeworksi 1999), at least not if the democratic riddle is to be resolved. And it need not be thought of in that way. There is an alternative approach available, although it has not had as much attention as it deserves, and needs.[10]

References

Bohman, J. and Rehg, W. (eds.) 1997, *Deliberative Democracy: Essays on Reason and Politics*, Cambridge, MA: MIT Press.

Cohen, J. 1989, *Deliberation and Democratic Legitimacy*, in Hamlin, A. and Pettit, P. (eds.), *The Good Polity*, Oxford: Basil Blackwell, 17-34.

Coleman, J. and Ferejohn, J. 1986, "Democracy and Social Choice", *Ethics*, 97, 6-25.

Dworkin, R. 2006, *Is Democracy Possible Here? Principles for a New Political Debate*, Princeton: Princeton University Press.

Elster, J. (ed.) 1998, *Deliberative Democracy*, Cambridge: Cambridge University Press.

Garland, D. 2001, *The Culture of Control: Crime and Social Order in Contemporary Society*, Chicago: University of Chicago Press.

Gutmann, A. and Thompson, D. 1996, *Democracy and Disagreement*, Cambridge, MA: Harvard University Press.

Habermas, J. 1984-1989, *A Theory of Communicative Action*, Vols. 1 and 2, Cambridge: Polity Press.

Habermas, J. 1996, *Postscript to Between Facts and Norms*, in Deflem, M. (ed.), *Habermas, Modernity and Law*, London: Sage, 135-50.

Hobbes, T. 1994, *Leviathan*, Curley, E. (ed.), Indianapolis: Hackett.

Holmes, S. 1995, *Passions and Constraint: On the Theory of Liberal Democracy*, Chicago: University of Chicago Press.

[9] If Ronald Dworkin's comments are interpreted in a non-eclectic manner, as noted earlier, then his approach might be situated in this broad family. And so might the approach espoused by Jürgen Habermas (1996: 142) when he argues that a regime of individual rights is required for deliberative participation in government; it depends on whether 'deliberative participation' is understood broadly enough to include or overlap with deliberative regulation.

[10] A version of this paper was presented as the Inaugural Edmund Burke Lecture at Trinity College, Dublin, in 2007, where I benefited greatly from the comments of those in the audience but in particular from the remarks of my respondent, Leif Wenar.

Knights, M. 2005, *Representation and Misrepresentation in Later Stuart Britain: Partisanship and Political Culture*, Oxford: Oxford University Press.

List, C. 2006, "The Discursive Dilemma and Public Reason", *Ethics*, 116, 362-402.

List, C. and Pettit, P. 2002, "Aggregating Sets of Judgments: An Impossibility Result", *Economics and Philosophy*, 18, 89-110.

List, C. and Pettit, P. 2011, *Group Agency: The Possibility, Design and Status of Corporate Agents*, Oxford: Oxford University Press.

Mayerfield, J. 2009, "Democratic Legitimacy of International Human Rights Law", *Indiana International and Comparative Law Review*, 19, 49-88.

Pettit, P. 2001, *A Theory of Freedom: From the Psychology to the Politics of Agency*, Cambridge and New York: Polity and Oxford University Press.

Pettit, P. 2003a, *Deliberative Democracy, the Discursive Dilemma, and Republican Theory*, in Fishkin, J. and Laslett, P. (eds.), *Philosophy, Politics and Society*, Vol. 7, *Debating Deliberative Democracy*, Cambridge: Cambridge University Press, 138-62.

Pettit, P. 2003b, *Groups with Minds of their Own*, in Schmitt, F. (ed.), *Socializing Metaphysics*, New York: Rowan and Littlefield, 167-94.

Pettit, P. 2007, "Responsibility Incorporated", *Ethics*, 117, 171-201.

Pettit, P. 2012, *On the People's Terms: A Republican Theory and Model of Democracy*, Cambridge: Cambridge University Press.

Przeworksi, A. 1999, *A Minimalist Conception of Democracy: A Defense*, in Hacker-Cordon, C. and Shapiro, I. (eds.), *Democracy's Value*, Cambridge: Cambridge University Press, 23-55.

Riker, W. 1982, *Liberalism against Populism*, San Francisco: Freeman and Co.

Rousseau, J.-J. 1973, *The Social Contract and Discourses*, London: Dent & Sons.

Simmons, A.J. 1979, *Moral Principles and Political Obligations*, Princeton: Princeton University Press.

Reading Rosenzweig's Little Book

Hilary Putnam

Harvard University

Abstract

In this article the author addresses the issues that Franz Rosenzweig raises in his *Büchlein* as they affect the former's own very personal manifestation of Judaism. The article therefore covers not only the contents of the "little book", but aims more generally to say something about aspects of Rosenzweig's thought that the author finds problematic. The article begins by looking at three notions that are often used in connection with the sorts of issues Rosenzweig raises (atheism, religion, and spirituality), goes on to stress the importance of Rosenzweig's "religious existentialism", and ends by keeping its distance from some of Rosenzweig's central claims, with an eye to both reconciling the author's religiosity with a plausible naturalism and salvaging a suitable space for philosophical speculation about God.

Keywords: atheism, religion, spirituality, religious existentialism, Rosenzweig

In this article I will discuss Rosenzweig's *Büchlein*,[1] but I do not want to simply interpret the "little book"—that I did in chapter one of my *Jewish Philosophy as a Guide to Life*.[2] I want to address the issues that Rosenzweig raises as they affect my own very personal manifestation of Judaism. So my article will range more widely than just the contents of the "little book", and also say more about aspects of Rosenzweig's thought that I find problematic. To do that, I will begin by looking at three notions that are often used in connection with the sorts of issues Rosenzweig raises: atheism, religion, and spirituality.

1. Atheism, Religion, and Spirituality

The terms "atheism", "religion", and "spirituality" are terms just about everyone thinks they understand. Yet all of them have many meanings. To the Greeks and the

[1] Rosenzweig 1921a.
[2] Putnam 2008.

Argumenta 1, 2 (2016): 161-168
ISSN 2465-2334

© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.PUT

Romans, "atheism", for example, meant disbelief in the gods (in the plural), and to the Jews, Christians, and Muslims, denial of the existence of "God" (in the singular), but "God" too has many possible understandings. (The situation is not improved by the fact that some people insist that their own understanding of these terms is the only right one.) For example, for most traditional Jews, Christians and Muslims, God was conceived of as a person or personality who was aware of and sometimes answered petitionary prayers. In other words, he was a supernatural being who could and sometimes would help you. The football player who crosses himself when he runs out to the field is acting as if he believed in such a God, and perhaps he does. From his point of view, someone who denies that is an atheist. Yet there are great theologians, including Maimonides, who thought that belief in a "magic helper" God is, in effect, *idolatry* although the traditional Jewish petitionary prayers are, nonetheless, obligatory, according to them, not because they effect God in any way, but because of their positive effect on the person who utters them. (Of course, Maimonides' theology was, in many ways, intellectually elitist: only those capable of philosophical meditation are capable of receiving God's "overflow", according to this great medieval Jewish sage.) For Maimonides, it is thus the "worshipper on the street" who is an atheist—an atheist without knowing it, since all idolaters are atheists without knowing it, in traditional Judaism. One person's "atheism" may be another person's "belief in God"!

"Religion" is another term with many understandings. For many years, I was a member of what corresponds to a "religion department" at Harvard, the interdisciplinary "Committee for the Study of Religion", and in that committee, as in, as far as I know, all religion departments in non-denominational schools in the United States, Confucianism is one of the faiths that is studied, along with Buddhism, Taoism, and many others. The Committee avoids speaking of "religions", however, preferring a finer-grained subdivision of what are often referred to by that term into "religious traditions" and "communities of faith" within those traditions. I once heard the great chairman of the Committee at the time I became a member, Wilfrid Cantwell Smith, make the bold claim that "I could show you as much variety in Methodist communities in London in 1815 as is supposed to exist among the 'world religions'". Yet Confucianism is not theistic at all, nor are many varieties of Buddhism. Yet all of these are unmistakably forms of spirituality. But what is "spirituality"?

What I am referring to by that term has two aspects. One I will illustrate with a reference to a French scholar, Pierre Hadot. A mutual friend who was very close to Hadot and also a former student of mine, Arnold Davidson, made me aware of Hadot's remarkable writings about the history of philosophy many years ago, a selection from which Arnold edited under the title *Philosophy as a Way of Life; Spiritual Exercises from Socrates to Foucault*, and reading Hadot made me aware that ancient philosophers in particular saw philosophy not as a set of propositions to be discussed and criticized—although, to be sure, they did that too—but as a regimen of spiritual exercises, or as he puts it "a mode of existing-in-the-world, which had to be

practiced at each instant; and the goal of which was to transform the whole of the individual's life".[3]

Referring to the practices involved in such a "mode of existing in the world" as *spiritual* exercises points to another aspect that they have: that the transformation in question is experienced as putting the individual in touch with something (or somethings) *higher* than herself. While that "something higher" may be conceived of as God, it also may not be. The "way of heaven" of which Confucianism speaks is not God, nor is the Tao, nor the mystical "emptiness" (which is somehow also a fullness) of Zen Buddhism, nor the natural forces and powers of paganism and neopaganism.

2. A digression about my own religiosity

As I explain in the Introduction and the Afterword to *Jewish Philosophy as a Guide to Life*, my own religious life involves saying the Jewish prayers and studying Jewish texts, including the Jewish Bible (the "Tanach") and the Talmud as well as Jewish philosophy. And I conceive of this study and prayer as a system of spiritual exercises in Hadot's sense. But I do not believe that the Jewish Bible, or any book, for that matter, is "the word of God". I am aware that the God I pray to is a human construct, although, in my view, we construct our various versions of "the available God"[4] in response to deep human needs that we do not "construct". I say this because this is absolutely not the view of Franz Rosenzweig. The idea that God is a human projection was anathema to Rosenzweig, and when he found it in some of Buber's early writing, he denounced it as "atheistic theology". Yet, even if mine is an "atheistic theology", I find much of value in Rosenzweig, as well as in Buber and Levinas, whom I also wrote about in my book. For me they are examples of different but wonderful expressions of the human need for a spiritual dimension to life. And the very diversity of those expressions is of value for me, as it is for another Jewish thinker I very much admire, Jonathan Sacks, who wrote *The Dignity of Difference* while he was Chief Rabbi of Great Britain.[5]

But I digress. Coming back to Rosenzweig, the very fact that Rosenzweig did not approve of the kind of theology, naturalistic theology, that best reconciles my own spiritual needs with my critical intelligence, has led me to ask just how Rosenzweig managed to reconcile, or at least to combine, the existentialist side I describe with sympathy in my book, and what I will call his "metaphysical theism". I will focus mainly on Rosenzweig's *Understanding the Sick and the Healthy* (*Büchlein vom guten und kranken Menschenverstand*, in German). That means, of course, that I am going to discuss things in Rosenzweig that I disagree with, and not only things I admire.

[3] Hadot 1981, transl. p. 265. I quote these words on p. 12 of chapter one of my *Philosophy as a Guide to Life.*
[4] Available God, from Kaufman 1993.
[5] Sacks 2003.

3. Rosenzweig's "Little Book" (*Büchlein*)

Since few if any of you will have read *Understanding the Sick and the Healthy* itself, and not everyone will have read chapter one of my book, the one that discusses it, I will briefly sketch the contents of Rosenzweig's *Büchlein*. The "illness" that Rosenzweig's imagines the "patient" to be suffering from is *philosophy*. (The likening of philosophy to an illness is the similarity between Rosenzweig and Wittgenstein, that I also discuss in my chapter.) And Rosenzweig explains that although the number of philosophers is small, the philosophical disease is capable of striking non-philosophers too, at any time. But what is the philosophical disease?

The philosophical disease that Rosenzweig attacks throughout the *Büchlein* is looking for "essences": wanting to know the "essence" of the world, the "essence" of the human being ("Man") and the essence of God. And in the various chapters of the *Büchlein*, Rosenzweig wittily describes and attacks all the philosophical positions known to him, especially but not exclusively in German philosophy. The strategy is to show that each of the metaphysical theories reduces its supposed subject—the world, the human being, God—to something so utterly different from what "common sense in action" takes the world (or anything in it—say "a slab of butter"), a human being (say, one's beloved), or God to be, that, for all intents and purposes, the world turns out to be *nothing*, the particular human being turns out to be *nothing*, and God turns out to be *nothing*. For example, the empiricist philosopher-scientists whom Rosenzweig knew about held that statements about matter are, when properly understood, statements about sensations. For great German scientists like Mach and Boltzmann—and this was true of the analytic philosophers of the time, especially the logical positivists, as well—science does not describe an independently existing reality, it just tells you that "if I have such and such sensations, then I expect such an such sensations to follow", e.g., if I (seem to myself to) put a burning match to a piece of paper (or have visual impressions as of doing that), then I can expect to have visual impressions of the paper burning. [Here I have expanded on an argument on p. 69 of the *Büchlein.*] In other words, science itself, according to its great positivist representatives, does not even pretend to tell you the *essence* of physical reality; it only tells you about *appearances*. To ask for more is to talk "nonsense" according to the positivists. And the idealist philosophers (Hegel and Co.)? For the Hegelians, the empiricists' and positivists' "sensations" are aspects of Mind. But what is Mind? If the world is Mind, it obviously cannot be just *my* Mind, that would be too solipsistic. So it must be "consciousness in itself", or perhaps "consciousness for itself"—something we do not experience the world or anything in it *as*. The philosophers' world ends up being *nothing*, or at least nothing that common sense in action is aware of.

This will have to serve to give you flavor of Rosenzweig's critique of metaphysics, minus the wonderful language and the lightly worn erudition. I have to skip over Rosenzweig's critique of philosophers' accounts of the essence of "Man" for reasons of space. The critical point comes when Rosenzweig comes to philosophers' account of the nature (or "essence") of God. Rosenzweig has no trouble disposing (to his satisfaction) of what he calls "mysticism" (by which he means one contemporary German variety of mysticism), a doctrine which insists that God is "wholly

other". Like the doctrines that make the world and its matter into Mind, and Mind into consciousness-in-itself, this makes God "nothing", according to Rosenzweig. And idealism, which makes God, Man and World all into Mind (= "nothing"), and Spinozistic pantheism which makes God into Nature (= "nothing"), are all convicted of making God into "nothing" as well.

What makes philosophy a disease, in Rosenzweig's eyes, is that while it starts from something valuable—namely, "wonder", that is, a sense of the miraculousness and mystery of the existence of the world, of humans, and of God—*freezes* that wonder by turning it into philosophical wheel-spinning.

[*A critical remark:* Rosenzweig seems to believe that accepting any of these philosophical conceptions of God is incompatible with a meaningful spiritual experience of God's reality, but I don't believe this is true. Maimonides, for example, thought that God has no properties, that is, no predicate that applies to anything else applies to God. Even "exist" does not apply to God. Nor does "not-exist". Nor does "either exist or non-exist"—God does not have to obey the Law of the Excluded Middle! This should make Maimonides' God a "nothing" for Rosenzweig (who does not discuss this negative theology, as far as I know). Yet anyone who has read *The Guide to the Perplexed* must know that Maimonides did feel a profound sense of spiritual connection with God, an "overflow" from God, as he describes it. And I am sure that many 19th century and early 20th century Idealists found *spiritual* and not only intellectual sustenance in the Hegelian idea that everything that exists, including one's own consciousness and the world outside it, is part of God. It is just not true that one has to have the "right" theology, if there is such a thing, to have a meaningful spiritual life.]

4. Existentialism

In my book, I refer to Rosenzweig, Buber, and Levinas as "existentialists", because I take the heart and soul of religious existentialism, beginning with Kierkegaard, to be a battle against turning religion into mere dogma. That is not to say that a religious existentialist must lack a creed; Kierkegaard was, after all a Christian, and one with Christian beliefs. But for him, if Christianity becomes *only* a set of beliefs, then it is no longer a religion, no longer Christianity, no longer of any value at all. What the European bourgeoisie of Kierkegaard's time regarded as Christianity he viewed with much of the same disgust as Nietzsche did.

In Christianity, although not in Judaism, the dichotomy between "faith" and "works" is fundamental, and Christians have spilled a lot of the blood of fellow Christians over disputes about "salvation through faith" versus "salvation through works". But from Kierkegaard's point of view, the 19th century bourgeois "Christians" misunderstood both "works" (which they understood as "good deeds") and "faith" (which they understood as "belief"). For an existentialist, Christian or Jewish, the "transformation of the whole of an individual's life" of which Hadot speaks involves a transformation of both the notion of "faith" *and* the meaning of "good deed". Faith ceases to be equated with belief in some propositions, and good deed ceases to be equated with some conventional notion of doing one's duty.

I say this, because if Rosenzweig had accepted such doctrines as "Torah misinai" (the Torah was dictated to Moses on Mount Sinai), this would *not* in itself have been incompatible with his being an existentialist. But what makes him a difficult thinker—and makes his disgust with theologies according to which God is a human projection, theologies which he denounces as "atheistic theology", while insisting on certain fundamental theses about God and Redemption, problematic—is the difficulty of determining just why and how he tolerates certain deviations from traditional belief and not others. If we can determine that, then I think we will get a better picture of this complex thinker.

5. Rosenzweig's toleration of (or indifference to?) certain traditional beliefs

Here are some examples of Rosenzweig's non-traditional side. I discuss these in the second chapter of my book, the chapter (mainly) on Rosenzweig's *The Star of Redemption*.[6] I point out that although, in a letter to the leaders of the Free Jewish Institute of Learning of which he was one of the founders, he dismisses text-critical approaches to Judaism (Wellhausen), psychological approaches (William James and Freud), and sociological approaches (Weber), he does *not* say that the text-critical scholars are *wrong* (as Hertz, does, for example, in many of the comments in his widely used edition of the Pentateuch). What he says is that "we" (practicing Jews) know Judaism "differently". And he goes on:

> What do we know when we do? Certainly not that all these historical and sociological explanations are false. But in the light of the doing, the right doing in which we experience the reality of the Law, the explanations are of superficial and subsidiary importance.[7]

So the truth or falsity of the doctrines of the infallibility and divine origin of the Bible, which text-critics like Wellhausen call into question, is "of superficial and subsidiary importance!"

Nor does Rosenzweig believe in reward and punishment in an afterlife. (Which he never mentions.) Both in the last chapter of the *Büchlein* and in the first part of *The Star of Redemption,* death is something to be faced bravely, not something to be "eluded". Nor does redemption, in Rosenzweig, involve a future Messianic or post-Messianic era. In one sense, redemption is simply God's love, which we can experience *now*. Redemption is a *present* event. In another sense, we do perhaps anticipate redemption in some form in the future, but what that means we do not know and do not need to know.

> Thus man may act unconcerned with the outcome; he may act according to the requirements of the world as it is today. That day, the day when action is required, lets

[6] Rosenzweig 1921b.
[7] From Glatzer 1961, p. 245.

> him understand what he must perform. The realm of time is the proper arena for his action. He does not need to wait until truth has risen from the depths.[8]

6. Putting this together

On the positive side, Rosenzweig insists on the reality of God, World, and Man. Each is "something", and not the vacuous "nothing" that the philosopher substitutes. The world is the world, and not "appearance", or "consciousness *an und für sich*, or mere "as-if"; Man, the name-giver, is the bridge between God and the world; and God is simply God. The *non-identity* of World, God (no form of pantheism is tolerated by Rosenzweig, any more that idealism), and Man (who is both a part of the world and more than a part of the world) is absolute. These are, I claim, philosophical dogmas for Rosenzweig, however much he denounces philosophy. He may say that these dogmas, indeed the whole of Part I of the *Star*, are of secondary importance, not worth anything by themselves, what is important is "experiential philosophy", the "new thinking"; but experiential philosophy is, it seems to me, a name for a set of spiritual exercises for those, and only for those, who accept these fundamental premises.

One thing that Rosenzweig's non-traditional side enables him to do is maintain the irrelevance of science and empirical studies generally to his religious faith. Psychology of religion, sociology of religion, text-critical and other historical studies of the history of Judaism are free to discover what they may. It is all "of superficial and subsidiary importance."

This separation of the spheres is helped by Rosenzweig's positivistic understanding of science as just concerned with regularities in the "appearances", on which I remarked earlier. At the same time, what I have called Rosenzweig's "religious existentialism" enables him to preach and teach, to experience Judaism and teach others to experience Judaism as (in Hadot's language) "a mode of existing-in-the-world, which had to be practiced at each instant; and the goal of which was to transform the whole of the individual's life." I have not concealed the fact that it is with his religious existentialism that I sympathize, and not with the residual philosophical dogmas.

7. A question I would put to Rosenzweig, if I could

Rosenzweig criticizes so many philosophies of religion as making God a "nothing", that I cannot help wondering, what justifies Rosenzweig in thinking that his "God is something" really makes God more than a mere "nothing"? What he tells us is that regarding God as "something" is not to be understood as a claim about God's essence, it is an expression of a return to "common sense in action". We are to use the *name* "God", and the other names we have for God, including YHWH and "elohim", without thinking about or seeking a theory of God's essence, to use these names as "healthy human understanding" (*guter Menschenverstand*) requires.

[8] Rosenzweig 1921a: 93.

But just how does "healthy human understanding" require us to use the name "God"? We are supposed to realize that God *loves* us; that is the heart and soul of Rosenzweig's faith experience and his theology. And it is, apparently, to be taken literally—Hands off! You philosophers! But does that mean that God *thinks about us*? Does God have *feelings*? Do God's thoughts and feelings occur in *time*? No doubt Rosenzweig would say that when we ask these questions we have moved from "common sense in action" to "philosophy", from wonder at God's love for us to "frozen" speculation. But is the idea that we can think of God as *loving* (and think of that as simply true, every bit as true as that what I bought at the grocery was a piece of butter) while not thinking any of these "philosophical" thoughts about God really plausible? Like Wittgenstein, Rosenzweig wants a sharp line to separate "healthy" uses of words and "unhealthy" ones, but I am suspicious of the idea that such a sharp line exists.

References

Glatzer, N. 1961, *Franz Rosenzweig: His Life and Thought*, New York: Schocken.

Hadot, P. 1981, *Exercices spirituels et philosophie antique*, Paris: Études augustiniennes; transl. *Philosophy as a Way of Life: Spiritual Exercises from Socrates to Foucault*, Oxford: Blackwell 1995.

Kaufman, G. 1993, *In the Face of Mystery: A Constructive Theology*, Cambridge, MA: Harvard University Press.

Putnam, H. 2008, *Jewish Philosophy as a Guide to Life: Rosenzweig, Buber, Levinas, Wittgenstein*, Bloomington: Indiana University Press.

Rosenzweig, F. 1921a, *Das Büchlein vom gesunden und kranken Menschenverstand*, *Understanding the Sick and the Healthy*, New York: Noonday Press, 1953; repr. Cambridge, MA: Harvard University Press, 1999, with an Introduction by H. Putnam.

Rosenzweig, F. 1921b, *The Star of Redemption*, Notre Dame: Notre Dame Press, 1985.

Sacks, J. 2003, *The Dignity of Difference: How to Avoid the Clash of Civilization*, London: Continuum.

Unreasonableness and Rights: On Quong's *Liberalism without Perfection*[1]

Hillel Steiner

University of Manchester

Abstract

This article argues that Quong's *Liberalism without Perfection* errs in claiming that the grounds for enforceably prohibiting unreasonable conduct are that it is unreasonable. What grounds that prohibition is, rather, that such conduct violates independently determined distributively just rights. Political liberalism *presupposes* a theory of distributive justice.

Keywords: compossibility, conflicts of interest, property rights, unreasonableness.

Let me start with two questions: Do we have a moral right to do wrong—that is, to engage in morally wrong conduct? And do we have a moral right to be unreasonable—that is, to engage in unreasonable conduct?[2] Quong's answer to the first question is 'yes', and to the second question, 'no' (*LWP*: 305 ff.). I am inclined to agree with both of Quong's answers here, but not with the reasons he advances in support of his second answer. And this disagreement may, I think, shed some further light on what we might more generally call the 'political liberalism project'.

First, however, a word of clarification before we get into that. For, more strictly, what both questions are asking is: 'Do our moral rights include Hohfeldian claims to other persons' forbearance from interfering with our morally wrong, or unreasonable, conduct: that is, interfering *solely on the grounds that* this conduct is morally wrong or unreasonable?'.[3] There are obviously many forms

[1] Quong 2011; cited hereafter as *LWP*.

[2] The Rawlsian conception of *reasonableness*, which Quong deploys, is extensively articulated in Rawls 1993, pp. 48-54: "Persons are reasonable in one basic aspect when, among equals say, they are ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so. Those norms they view as reasonable for everyone to accept and therefore as justifiable to them; and they are ready to discuss the fair terms that others propose." (p. 49).

[3] A Hohfeldian claim, or what Hohfeld called 'a right in the strict sense', is an entitlement held by one person that correlatively entails the presence of a duty in another: 'X has a claim that Y pay her five dollars' entails 'Y has a duty to pay X five dollars'. It is to be

Argumenta 1, 2 (2016): 169-175
ISSN 2465-2334
© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.STE

of morally wrong conduct: Quong identifies three of them as rudeness, ingratitude, and selfishness (*LWP*: 306), and many others could be added to that list. But, clearly, one that *cannot* sensibly be added in the present context is injustice or moral right-violation itself. Whereas we can make sense of the idea of a moral right to do wrong by acting rudely, we cannot here sensibly ask whether we have a moral right to do wrong by violating moral rights.

Of course, we can easily envisage circumstances where, in order to *prevent* one right-violation, it is necessary to violate another right. This is, for example, what Amartya Sen's cases of *multilateral interdependences*[4] were meant to illustrate: the only way to stop time-bomber A from killing innocent B is for preventer C to break into innocent D's locked office, in order to telephone a warning through to B. Part of what is going on in such cases is the pitting of D's first-order right against B's presumed second-order right to have her first-order right protected. Enforcing first-order rights, by complying with the duties entailed by second-order rights, all too frequently involves violating others' first-order rights. And this does indeed pose a familiar set of problems, to which liberalism *per se* offers no distinctive solution. For while it is pretty clear that stopping A from killing B is, in itself, morally permissible, it is at least less clear that conscripting D or his belongings for that exercise is also morally okay.

In any case, though, these are *not* the problems that Quong is alluding to, in his discussion of rights to do wrong. There, he is talking about wrongdoings that are *not* rights-violations: I am going to call these wrongdoings *not unjust* wrongdoings, or *NUs*. And what Quong is saying is that *NUs* are immunised against permissible interference by the rights of the wrongdoer. That wrongdoer is said to be acting entirely within what Quong calls his own *domain* and, hence, his undesirable conduct must be tolerated. However, the same is *not true* of those who engage in unreasonable conduct. And the question is '*Why not?*'.

Evidently, any answer must depend on our understandings of 'rights', 'domains', and 'unreasonable conduct'. How do our rights constitute domains which are immunised against others' permissible interference with our *NUs* but not with our unreasonable conduct?

In this regard, Quong's proffered example of unreasonableness here is one concerning the acknowledged right to freedom of religion. And what he says is that

> The right to freedom of religion … cannot be used to justify your theft of my laptop computer, even if your religious beliefs really do require you to steal my laptop computer on pain of eternal damnation. Your theft of my laptop cannot be an exercise of your right to religious freedom, even if the theft was sincerely and religiously motivated. Rights are only intended to permit or protect choices made within a limited domain. Your theft of my laptop computer is not part of the domain that is protected by your right to religious freedom, but the important question for our purposes here is this: how are the limits of a given right-protected domain defined? That is, how do we *know* that the laptop theft clearly falls outside that domain? (*LWP*: 307).

distinguished, in particular, from a liberty which is an entitlement of one person that entails the absence of a duty in that same person; 'Y has a liberty to refrain from paying X five dollars' entails 'Y has no duty to pay X five dollars'. Cf. Hohfeld 1919: 36 ff.

[4] Sen 1982.

This is, indeed, the important question. And one answer to this question that Quong rules out is that

> the right to religious freedom cannot protect your religiously motivated theft of my laptop computer for the simple reason that doing so will violate my right to private property. Rights can only protect actions that respect the boundaries of other people's rights—once an act ceases to respect the rights of others, it is no longer possible for that act to be itself protected by a right. As Rawls explains: 'justice is prior to the good in the sense that it limits the admissible conceptions of the good, so that those conceptions the pursuit of which violates the principles of justice are ruled out absolutely: the claims to pursue inadmissible conceptions have no weight at all' (*ibid*.).

Quong correctly deems this answer to be unsatisfactory because

> it begs the crucial question—it *assumes* that it is unjust for you to steal my laptop computer, but we want to know *why* this is the case. If we agree that people normally have a right to religious freedom, why should my right to private property make the exercise of your religion impermissible? If this really was a conflict of rights, surely the right to religious freedom is more important than the right to a laptop? (*LWP*: 307-308).

So what we have here is a genuine conundrum. But, before we proceed to look at possible solutions to it, let us put a little more flesh on the bones of Quong's example.

Here you are, a devout adherent of a religion that stringently requires you to send an e-mail to God each day before the close of business: an e-mail reporting and repenting all the *NUs* you have committed during the previous 24 hours; your failure to do so will result in eternal damnation. On this particular day, your faithful computer decides to crash, leaving you with no option but to steal my laptop in order to transmit the required message punctually. So the next day, in court, you are charged with theft and your defence is that, in taking my laptop, you were exercising your just right to freedom of religion. More precisely, your argument is that, had I managed to prevent your taking it, I would have been guilty of violating that right of yours. And my counter-claim is that, in taking it, you were violating my just property right to possess that laptop. Faced with these mutually opposed rights claims, what should the judge do?

Now, I have pretty much made my living by worrying about conflicts of rights and how to avoid them or, at least, minimise their frequency. And one conclusion I have reached is that certain ways of *conceiving* rights do not help in that regard. Specifically, the view which Quong takes from Jeremy Waldron, namely that

> a right exists when an agent has a sufficiently strong interest to justify holding some other person or persons to be under a duty (*LWP*: 306)

is *not* a way of conceiving rights that is going to be of much assistance in this case. After all, eternal damnation is a pretty heavy sentence and, for those who believe in its possibility, avoiding it has to come fairly near the top of their scale of strong interests.

Nor, I think, will it do for me to assert *my* interest in untrammelled possession of my laptop, and thereby invite the judge to compare the strengths of our competing interests. For, unless she misguidedly imagines her task to be one of maximising social utility, any judgement she delivers on the basis of *these* arguments would, ineluctably, amount to nothing less than an illiberal endorsement of one contestable conception of the good at the expense of another.

Regrettably enough, our world is heavily populated by interpersonal conflicts of strong interests, with the consequence that rights conceived as protections of strong interests are unlikely to deliver the protection they promise. In other words, the rights constituting our respective personal domains—the domains that are said to set limits on what we may do without others' permissible interference—had better *not* be ones determined by reference to our strong interests. Otherwise, our court hearing may itself be destined to carry on for eternity.

So how *can* these rights be determined? How can we get a set of personal domains that are *compossible*? Quong's sensible solution to our conundrum is to invite us to return to the foundational ideal from which our just rights all derive. What he says is that

> To successfully answer our question we need … to ask whether the particular act that is alleged to be protected by a right is consistent with the overall moral ideal which the system of rights is meant to uphold. That moral ideal, I assume, is the ideal of society as a fair system of social cooperation for mutual benefit amongst free and equal citizens. Thinking about the question in this way yields a coherent solution to the laptop example. Private property rights (including rights over laptop computers) are perfectly consistent with the idea of fair cooperation amongst free and equal people; indeed it seems unlikely that this moral ideal could be realized without property rights. My right over my laptop is thus consistent with our foundational moral ideal. The right to religious freedom, however, though also securely derived from that moral ideal, cannot be coherently extended to protect the theft of others' property. The right to religious liberty is meant to protect each person's religious choices from the interference of others, to provide each person with their fair share of *moral space* within which they can make religious choices free from the interference of others. It cannot be used to give a person the moral right to appropriate other people's property whenever this is required for their religious purposes….. (*LWP*: 308; italics added)

I think this is almost exactly right. In particular, I think Quong's concept of 'moral space' is an extremely useful one, and I take it to be pretty much identical with what he refers to as persons' domains. These are the personal spaces within which people can conduct themselves in whatever way they choose—religiously, rudely, or what have you—free from permissible interference by others.

Where I think this quoted passage goes astray is in its suggestion that it is the *right to religious freedom* that is providing each person with that personal space. The right that is providing that space, not least in Jon's own laptop example, is *not* the right to religious freedom but rather the right to private property. To see how and why this must be so, let us return to our courtroom hearing and revise it a little.

Again, your defence is that, in taking the laptop, you were simply exercising the general right we all have, to religious freedom: you were acting within the personal space reserved for you by that right. But now, instead of my

counter-argument being that you were violating my right to private property, what I advance is the claim that you were violating my equally acknowledged *right to free expression*. I was about to engage in an important discussion on the internet, and your taking the laptop prevented me from doing so: it trespassed on the personal space reserved for me by *that* general right, which is also one that is securely derived from our overall moral ideal. After all, does not the right to free expression similarly 'give each person their fair share of moral space within which they can make *expressive* choices free from the interference of others'? And, if so, then how is the judge to determine where the moral space for religious choices ends and the moral space for expressive choices begins? I do not see how the very good idea of consulting our foundational ideal—society as a fair system of social cooperation for mutual benefit amongst free and equal citizens—is going to yield an answer to that demarcation question.

What *is* going to give us that answer is a line of reasoning about how that foundational ideal generates the right to private property and, beyond that, what set of just private property rights is derivable from it. In that regard, Quong is absolutely correct in saying that "indeed it seems unlikely that this moral ideal could be realized without property rights". For it is those very property rights that are doing the heavy lifting when it comes to determining each person's fair share of moral space, each person's interference-free domain of protected choices. In the present case, it is my right to private property that trumps—that sets the limit on—your right to religious freedom, and that gives us the verdict which Quong endorses: namely, that your taking the laptop is *unreasonable* and *not* a protected choice.

The point here is that, for liberals more generally, the right to private property is *not* one right that simply sits alongside others, such as the rights to freedom of religion, free expression, free contract and free association. Rather, it is the *platform* on which these other rights are erected, exercised, and more significantly, *demarcated* one from another. Ranging over rights to persons and to extra-personal objects, it is *that* right that determines the boundaries of our protected domains, within which each of us can behave religiously, expressively, rudely, and so forth, free from one another's permissible interference. For it is only by intervening in persons' dispositions of themselves and extra-personal objects that interference can occur.

So it turns out that what makes your taking the laptop *unreasonable* has less to do with the fact that its justification invokes an esoteric religious belief which many of us do not—and perhaps cannot—share, and much more to do with the fact that it is a violation of my just property rights. If I *were* to share that belief and to regard avoidance of eternal damnation as far more important than participation in that online discussion, I would have had good reason—*ex ante* or *ex post*—to waive my right against your taking my laptop and it would, consequently, not amount to a violation.[5] Nor, I think, would it then count as *unreasonable*, notwithstanding the fact that its justification does invoke that esoteric religious belief. It would not be unreasonable because no one other than a fellow believer is being expected to bear the cost of your action.

[5] Though what is also true is that, if I was an adherent of the *very same* religion as yours, we could still have a rights-conflict, since my consequent strong interest in sending a similar email at the same time would conflict with your interest in doing so.

So, as was said at the outset, I *agree* with Quong that unreasonable conduct does not fall within the personal domains—the moral spaces—that are protected against others' permissible interference. But my reason for agreeing is that such conduct violates just property rights: what it amounts to is a *distributive injustice*. The set of unreasonable actions is a proper subset of the set of rights violations. And this implies that *an account of unreasonableness logically presupposes some conception of distributive justice*.

Precisely what that conception should be is, indeed, something that requires us to consult our foundational ideal of society as composed of free and equal individuals. In that regard, I've tended to follow H.L.A. Hart and others[6] in seeing that ideal as simply and straightforwardly entailing a foundational *right to equal freedom*—a right which can, I think, be shown on analysis to generate a comprehensive set of compossible property rights, a set of non-conflicting personal domains.[7]

But other conceptions are also possible, depending on what further assumptions are introduced into the derivation of property rights. So if, for instance, we want to assume that all persons are rationally self-interested and non-envious and risk-averse and veiled in ignorance when exercising their right to equal freedom to design just basic structures, then it's likely that the thereby derived set of property rights will be ones modified to reflect the demands of the Difference Principle. And if—and to the extent that—persons are *reasonable*, they will *not* insist on modifying the set of property rights in such ways as to reflect the demands of esoterically-held belief systems.

That said, however, it remains true that unreasonable people would still be vested with protected personal domains—domains constituted by, *and limited to*, whatever property rights they secure from exercises of their foundational right to equal freedom. Unreasonable people *do* have rights. What they don't have are rights to behave unreasonably.[8]

[6] Including Herbert Spencer, Henry George and, arguably, Locke, Kant, and Nozick.

[7] Cf. Steiner 1994, *passim*.

[8] An arguable exception: Nicola Mulkeen (correspondence) asks whether persons' domains protect choices which are racially discriminatory. Is their conduct free from permissible interference if they will only help white people with their broken shopping bags, or will only permit a white person to use their laptop to send a message to God? On the one hand, such conduct does look to be unreasonable inasmuch as the justification for it invokes *ipso facto* esoteric reasons; on the other hand, it does not violate anyone's property rights. My inclination is to see instances of such conduct as *NUs*, rather than as rights-violations: morally wrong, but not unjust. Assuming that persons with broken shopping bags, or with the need to use another's laptop, do not have *rights* to others' assistance, it seems to me that liberalism must allow individuals to confer or withhold their benevolence on whatever grounds they choose. Two reasons: (1) there seems to be a non-interruptible slippery slope connecting choices about benevolence recipients and choices about, say, marriage partners—and I presume that liberals would regard the latter as properly protected; and (2) racially discriminatory benevolence, deplorable as it is, entails no denial of all persons having "the same moral status as free persons—as people who are not naturally under the authority of someone else" (*LWP*: 2), nor of "their possession of two moral powers: a capacity for a sense of justice and a capacity for a conception of the good" (*LWP*: 38). On what I take to be a just way of addressing—and offsetting—the disadvantages of suffering racial discrimination, see Steiner 1994: 276. But see Carter 2013 for an argument that such racial discrimination *can* imply a denial of such Rawlsian agential status to its victims.

In short, the impartiality of political liberalism—its capacity to resist dispositive comparisons of the conflicting interests generated by rival perfectionist conceptions of the good—rests upon the set of distributively just property rights that prevail among free and equal individuals.[9]

References

Carter, I. 2013, "Are Toleration and Respect Compatible?", *Journal of Applied Philosophy*, 30, 195-208.

Hohfeld, W.N. 1919, *Fundamental Legal Conceptions*, Cook, W.W. (ed.), New Haven: Yale University Press.

Quong, J. 2011, *Liberalism without Perfection*, Oxford: Oxford University Press.

Rawls, J. 1993, *Political Liberalism*, New York: Columbia University Press.

Sen, A. 1982, "Rights and Agency", *Philosophy & Public Affairs*, 11, 3-39.

Steiner, H. 1994, *An Essay on Rights*, Oxford: Blackwell.

[9] This paper has greatly benefited from discussion at the 2013 Churchill College, Cambridge, conference on *Liberalism without Perfection* and from comments supplied by Ian Carter, Matthew Kramer, and Jonathan Quong.

Argumenta 1,2 (2016)
Special Issue

New Trends in Philosophy of Mind and Epistemology

Edited by

Maria Cristina Amoretti and Francesca Ervas

The Journal of the Italian Society for Analytic Philosophy

Contents

New Trends in Philosophy of Mind and Epistemology: An Overview

Maria Cristina Amoretti* & Francesca Ervas**

*University of Genova

**University of Cagliari

The seven papers included in this special issue of *Argumenta* might be ideally divided into two parts. On the one hand, this issue collects four contributions dealing with some important topics in Philosophy of Mind and Philosophy of Language: the modularity of mind (the connections between the "pragmatic" module and epistemic vigilance mechanisms), the problem of perception and its link with action (the alleged anti-representational character of enactivism), the nature of phenomenal content (the plausibility of naïve realism in explaining the phenomenology of veridical visual experience), and the alleged irreducibility of consciousness (the claim that anti-physicalist intuitions are just a by-product of certain epistemological features of phenomenal concepts). On the other hand, there are three more contributions discussing some relevant themes in Logic and Epistemology: the actuality of the ancient Master Argument (its consistency and relationship with contemporary tense logic), the problem of evidence (the kind of evidence, psychological or non-psychological, intuitions actually provide), and that of counterevidence (the possibility that undermining defeaters, contrary to overriding defeaters, require the subject to engage in some higher-order epistemic reasoning).

In the ideal first section, the paper "Pragmatics, modularity and epistemic vigilance" adopts the modular view of the mind and focuses on the connection between the pragmatics module and epistemic vigilance mechanisms, which filter the incoming information and assess the reliability, competence, and benevolence of the informer. The author, Diana Mazzarella, aims at showing not only that epistemic vigilance mechanisms may directly affect the comprehension process, but also that their emergence may correlate with different developmental stages in pragmatics. As to the problem of perception and its link with action, a successful approach is enactivism, a thesis which exploits the union of action and perception in order to claim that perception is direct, i.e. not mediated by representations. In "Enactivism, Representations and Canonical Neurons", Gabriele Ferretti and Mario Alai question this inference arguing that even though the union of action and perception is well-confirmed by wide empirical evidence in neuroscience, it can only be explained involving subpersonal representations. However, this would mean that perception is indirect, contrary to the enactivist's conclusion. As to the nature of phenomenal content, naïve realism claims not only that the phenomenology of veridical visual experience is

Argumenta 1, 2 (2016): 179-180
ISSN 2465-2334
© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.AMO

explained by acquaintance or perception, an irreducible mental relation between the subject and some environmental objects, but also that the visual phenomenology of veridical experience is wholly constituted by the environmental objects perceived by the subject. Takuya Niikawa's paper "Naïve Realism and the Explanatory Role of Visual Phenomenology" aims to show that naïve realism is what best captures the explanatory roles of the phenomenology of veridical experience. Finally, the paper "Conscious Experiences as Ultimate Seemings: Renewing the Phenomenal Concept Strategy" presents the current versions of the Phenomenal Concept Strategy, which argue that anti-physicalist intuitions concerning consciousness (especially that of conceivability) are just a by-product of certain epistemological features of phenomenal concepts. The author, François Kammerer, raises some questions against these versions of the strategy and defends the idea that phenomenal concepts are concepts of unjustified justifications, or "ultimate seemings".

In the ideal second section, the paper "The ancient Master Argument and some examples of tense logic" discusses the Master Argument of Diodorus Cronus, which—being halfway between ancient logic and metaphysics—has been long debated by both logicians and philosophers. The author, Fabio Corpina, mainly deals with Prior's reconstruction, which marks the beginning of tense logic. More specifically, he evaluates and criticizes an argument by Øhrstrøm and Hasle trying to prove the inconsistency of the Master Argument, and then compares their strategy with that adopted by Prior. As to the problem of evidence, in "Williamson on the psychological view" Serena Maria Nicoli defends the classical assessment of intuitions according to which the nature of the evidence they provide is psychological. Against Williamson—who thinks that as the subject matter of philosophy is non-psychological, the evidence collected by intuitions must be understood as non-psychological as well—she adopts a Wittgensteinian perspective on the aims of philosophy and argues that conceiving the subject matter of philosophy as conceptual does not necessarily amount to conceive it as psychological. Finally, in "Undermining Defeat and Propositional Justification", Giacomo Melis discusses the problem of counterevidence. Defeaters can be understood as pieces of counterevidence: "overriding" defeaters give a subject *S* a reason to believe not-*p* while "undermining" defeaters give *S* merely a reason to give up *p*. The author defends the idea that undermining defeaters, contrary to overriding defeaters, requires the subject to engage in some higher-order epistemic reasoning. In particular, he shows that this proposal can be not only applied to doxastic justification, but also extended to cover propositional justification.

Pragmatics, modularity and epistemic vigilance

Diana Mazzarella

Institut des Sciences Cognitive 'Marc Jeannerod'
University College London

Abstract

This paper addresses the question of the place of pragmatic abilities in the overall architecture of the mind. Until recently, pragmatics was assumed to be part of a non-modular, unencapsulated, central system. Sperber and Wilson (2002) have proposed that pragmatics is to be conceived of as a sub-module of the mind-reading module, with its own principles and mechanisms. This is in line with an increasingly modular view of the mind (Cosmides & Tooby 1992, 1994; Sperber 1994b, 2001, 2005; i.a.), according to which cognition consists of many dedicated domain-specific mechanisms or 'conceptual modules', highly interconnected with each other. This paper focuses on the connection between the pragmatics module and epistemic vigilance mechanisms, that is, mechanisms that assess the quality of the incoming information and the reliability of the individual who dispenses it (Sperber *et al.* 2010). The latter take as their proprietary input the output of the pragmatics module and assess its believability. This paper makes two original proposals: first, that epistemic vigilance mechanisms may directly affect the comprehension process, and, second, that the emergence of epistemic vigilance mechanisms targeted at assessing the communicator's competence and benevolence may correlate with different developmental stages in pragmatics.

Keywords: modularity, pragmatics, epistemic vigilance

1. Introduction

The cognitive revolution, which from the early '60s shaped the domains of linguistics, anthropology, psychology and related disciplines, manifested its effect in the field of pragmatics with the seminal work of Sperber and Wilson (1986/1995). Among many other issues, Sperber and Wilson brought to the attention of the pragmatics community the question of the place of pragmatic abilities in the overall architecture of the mind. At that time, Fodor had already suggested that human cognitive architecture is partly modular (Fodor 1983) by introducing the functional and architectural distinction between modular perceptual and linguistic processors, on the one hand, and non-modular higher-

Argumenta 1, 2 (2016): 181-193
ISSN 2465-2334
© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.MAZ

level central processes, on the other. This gave rise to the interesting question of whether pragmatics is to be thought of as a domain-specific modular system or as part of a domain-general central system.

Fodor's (1983) distinction between modular input systems (perception and language) and non-modular central thought processes is based on a precise characterisation of the nature of 'mental modules'. Modular mental systems are task-specific and relatively autonomous systems which contingently share the following properties: they operate in a mandatory and fast way, they are domain-specific, informationally encapsulated and generally associated with a fixed neural architecture, and they exhibit specific breakdown and developmental patterns. Fodor places particular emphasis on informational encapsulation: "The informational encapsulation of the input systems is […] the essence of their modularity" (Fodor 1983: 70). A mental system is informationally encapsulated (or cognitively impenetrable) if it is rigidly restricted in its access to the full range of the organism's knowledge, beliefs and desires. Such a system cannot take account of (potentially relevant) information that does not belong to its proprietary database, that is, which lies outside its own task specific body of information for processing its particular domain of stimuli.

The Fodorian notion of 'module' seems hardly applicable to pragmatic processing given its undisputed property of sensitivity to a wide range of contextual or background information. A constant tenet of post-Gricean approaches to pragmatics (e.g. Relevance Theory and other contextualist accounts) is the assumption that linguistic meaning underdetermines not only what is meant, but also what is said or explicitly communicated, that is, the truth-conditional content of the utterance (Carston 2002). Consider the following example:

(1) Neil has broken his leg.[1]

The hearer of (1) must decide who the referential expressions 'Neil' and 'his' refer to. Does 'Neil' refer to $Neil_1$ (the hearer's son) or $Neil_2$ (her colleague in the linguistics department)? Does 'his' refer to the referent of 'Neil' or to a different male individual who got a broken leg from a scuffle with $Neil_1$/$Neil_2$? Furthermore, the hearer must decide when the event took place (at some generic time in the past? a few days ago? this morning?). An utterance of (1) may also convey some implicatures (intended implications): for instance, that $Neil_2$ cannot participate in the staff meeting because he is still at the hospital, or that $Neil_1$ cannot run in the school marathon because he has not entirely recovered from the bad accident in which he broke a leg. The linguistically encoded meaning of the utterance, thus, plays a minor, albeit crucial, role in the recovery of the communicated content of the utterance, both at the explicit and at the implicit level. Crucially, there is no principled restriction on the kinds of information that pragmatic processing can call on: perceptual information, background information stored in long-term memory, information that is part of the linguistic context of the utterance. Pragmatic processes appear to be 'informationally unencapsulated' and an account of the structure and function of pragmatics in the mind need to respect this feature of utterance interpretation.

In line with this, pragmatics has traditionally been conceived of as a non-modular central inferential process, i.e. a non-deterministic process of rational

[1] Carston 2007: 25.

belief-fixing. Specifically, it has been described as the process of arriving at an interpretation of the utterance, that is, the process of fixing a belief about the speaker's communicative intention. This has indeed been the standard position of Relevance Theory up to the late '90s (Wilson and Sperber 1986, Sperber and Wilson 1986/1995).[2]

The main aim of this paper is to explore recent developments on the issue of the place of pragmatics in the cognitive architecture of the mind from a relevance theoretic perspective. The structure of this paper is as follows: in Section 2, I present the relevance-theoretic account of pragmatics as a module (Sperber and Wilson 2002, Wilson 2005) and discuss the non-Fodorian notion of module that it relies on. In Section 3, I investigate the relationship between pragmatics and mechanisms of 'epistemic vigilance', that is, those mechanisms that assess the quality of incoming information and the trustworthiness of the individual who dispenses it (Sperber *et al*. 2010) and suggest that the interaction between these two systems goes beyond what is currently acknowledged. Finally, I present some developmental implications of this proposal and sketch directions for future research.

2. The pragmatics module

Sperber and Wilson (2002) suggest that the interpretative process is carried out by a dedicated pragmatics or comprehension 'module', with its own principles and mechanisms. How does this fit with the traditional view of pragmatics as an informationally unencapsulated system? In order to answer this question, we first need to consider the conceptual transformation that has characterised the notion of 'mental module' itself. Although pragmatics is now conceived of as a module, it is not a Fodorian module. For this reason, it is worth taking a step back to look at the wider picture of the mind within which Sperber and Wilson's (2002) proposal is located. This is the view of the mind as 'massively modular', a position pioneered by the evolutionary psychologists Cosmides and Tooby (1992, 1994) and advocated by Sperber himself (e.g. Sperber 1994b, 2000, 2005).

Starting from evolutionary considerations, Sperber (1994b) suggests that the mind is modular through and through, that is, that cognition is based on dedicated domain-specific mechanisms, as opposed to domain-general central processes. This view, which takes the name of 'massive modularity', subverts Fodor's architectural taxonomy of psychological processes, and introduces 'conceptual modules' in addition to perceptual ones. While a discussion of the massive modularity thesis goes beyond the purpose of the present paper, it is worth focusing on some of its implications: first, the introduction of a revised and looser notion of mental module, and, second, the hypothesis of a complex network of perceptual and conceptual modules "interconnected in ways that would make an engineer cringe" (Sperber 1994b: 46).

The assumption that cognitive mechanisms, like every biological mechanism, are adaptively specialised for the solution of particular kinds of task is what grounds the conception of 'mental modules' as domain-specific and autonomous computational mechanisms. They are attuned to the regularities of their specific domain and employ dedicated procedures which are justified by

[2] But see Kasher 1991 for a modular view of certain components of pragmatic knowledge (i.e. basic speech acts and talk-in-interaction). For a critical discussion, see Carston 1997.

those regularities. Significantly, they do not (necessarily) manifest all the properties that Fodor attributes to mental modules: their operations may not be mandatory (in the sense that an appropriate input may not be sufficient to trigger its own processing (Sperber 2005)) and their informational encapsulation may be conceived of as a matter of degree:

> it may be that we have to rethink the concept of module and allow for a kind of continuum, from peripheral perceptual systems, which are rigidly encapsulated (not diverted from registering what is out there), through a hierarchy of conceptual modules, with the property of encapsulation diminishing progressively at each level as the interconnections among domain-specific processors increase (Carston 1997: 20).

This passage interestingly highlights the connection between this new notion of mental module and the hypothesis that modules are highly interconnected. The output of a perceptual or conceptual module can be fed to other conceptual modules, whose outputs can in turn be fed to further conceptual modules, and so on and so forth. The result consists of a chain of inferences that integrates the contribution of each individual module.

With this picture in mind, let us turn to Sperber and Wilson's (2002) proposal that utterance interpretation is carried out by a dedicated inferential mechanism or 'comprehension module'. Sperber and Wilson suggest that pragmatics is a sub-module of the general mind-reading module, which is responsible for providing explanations of individuals' behaviours in terms of attributed mental states (e.g. beliefs, intentions, etc.).[3] Recognising the intention behind the speaker's communicative behaviour is a particular case of mind-reading (as Grice pointed out long ago). But, while utterances are a type of action and a speaker's meaning is a type of intention (i.e. a communicative intention, which is a second-order informative intention), according to Relevance Theory, the domain of overt communication exhibits such specific regularities and is so important in human life that, instead of employing general mind-reading procedures, it deploys its own dedicated comprehension procedure.

The investigation of such regularities requires the introduction of some technical notions, among which the central one is 'relevance'. Relevance is a property of inputs to cognitive processes (e.g. percepts, utterances) and it is a cost-benefit notion: the smaller the processing effort (cost), the greater the relevance; the greater the cognitive effects (benefit), the greater the relevance. Sperber and Wilson (1986/1995) identify three kinds of cognitive effects: contextual implications (i.e., implications that can be derived from the input and the context, but from neither input nor context alone), strengthening of existing assumptions, and contradiction and elimination of existing assumptions. According to Sperber and Wilson (2002), human cognition has been subject to a continuous evolutionary transformation towards greater cognitive efficiency, so that it tends to be geared to the maximisation of the relevance of the information processed (this is the First, or Cognitive, Principle of Relevance). Given this

[3] For a detailed defence of this claim see Sperber and Wilson 2002 and Wilson 2005. According to this view, mind-reading is not a single, homogeneous system but a collection of autonomous dedicated mechanisms, or sub-modules (e.g. the Eye Direction Detector module).

universal tendency to maximise relevance, an audience will pay attention to a stimulus only if it seems relevant enough. By producing an *ostensive* stimulus (e.g. an utterance), a communicator raises a particular expectation in the audience that is not raised by other stimuli. Specifically, her ostensive stimulus conveys a presumption of its own optimal relevance (this is the Second, or Communicative, Principle of Relevance), which is defined as follows:

(2) **Presumption of optimal relevance**
The ostensive stimulus (e.g. the utterance) is presumed to be (i) at least relevant enough to be worth the hearer's attention and (ii) the most relevant one compatible with the speaker's abilities and preferences.

The Communicative Principle of Relevance expresses the kind of regularity that characterises the domain of overt communication and it is this which, according to Sperber and Wilson (2002), motivates the adoption of the following dedicated comprehension procedure (that works according to an in-built presumption of optimal relevance):

(3) **Relevance-guided comprehension procedure**
(a) Follow a path of least effort in computing cognitive effects: Test interpretative hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.
(b) Stop when your expectations of relevance are satisfied.

This procedure guides the construction of appropriate hypotheses about explicit content, intended contextual assumptions and intended contextual implications (implicatures) within the overall comprehension process. Note that the regularity expressed by the Communicative Principle of Relevance is specific to the domain of overt communication: in general, an observer is not entitled to expect that the intentional behaviours of others would have any particular level of relevance to him. But since this expectation is warranted in the case of overt communicative behaviours, the relevance-guided comprehension procedure tends to yield reliable conclusions. Here is an illustration of how the relevance-guided comprehension procedure applies to example (1). Consider the following scenario (described by Carston 2007: 25). Robyn is in one of her students' company. At some point during the conversation, the student, Sarah, addresses to Robyn the following utterance:

(1) Neil has broken his leg.

Suppose that Robyn knows two people called "Neil": $Neil_1$, her young son, and $Neil_2$, a colleague in the linguistics department. While Sarah does not know $Neil_1$, she is acquainted with $Neil_2$, who teaches her syntax. Assume that Robyn is a very apprehensive mother and that she is always worried about her son $Neil_1$. When hearing Sarah's utterance, then, the most accessible referent for "Neil" is $Neil_1$. Furthermore, because $Neil_1$ is both a very clumsy and fearless child, the first interpretative hypothesis to come to her mind is that $Neil_1$ has broken his own leg (after she left him at the kindergarten that morning). In following a path of least effort, this is the first hypothesis to be tested. The comprehension procedure stops when it reaches an interpretation that satisfies the hear-

er's expectations of relevance. The interpretative hypothesis that $Neil_1$ has broken his leg this morning, however, does not satisfy Robyn's expectation of relevance. The reason is that she knows it is not compatible with the speaker's abilities (Sarah does not know that Robyn has a son called "Neil"). The interpretative hypothesis is thus discarded in favour of a less accessible one, e.g. $Neil_2$ has recently broken his own leg.

Interestingly, while the construction of interpretative hypotheses calls on various sources of contextual information, the relevance-guided comprehension procedure clearly suggests that not all available contextual information has to be actively taken into consideration. Rather, pragmatic interpretation exploits "whatever information is most highly activated by the automatic workings of the cognitive system at the time" (Wilson 2005: 1141).

3. The pragmatics module and epistemic vigilance

3.1 The output of the pragmatics module

The massive modularity framework of the mind recognises that modules are highly interconnected with each other, that is, they form a network of systems and subsystems connected in such a way that they may take as input the outputs of several other modules. In what follows, I focus on the pragmatics module and its connection with epistemic vigilance module(s) (Sperber *et al*. 2010).

The pragmatics module takes as input an ostensive stimulus and delivers as output an interpretative hypothesis about the communicator's meaning. Importantly, pragmatic interpretation corresponds to the process of fixing a belief about an interpretative hypothesis, i.e. about which propositions the speaker communicated ('comprehension'), rather than to the process of fixing a belief in the propositions themselves ('(doxastic) acceptance'). Sperber *et al*. (2010) have recently investigated the distinction between comprehension and acceptance and suggested that the latter does not automatically follow from the former. Rather, they claim, humans have developed a suite of cognitive mechanisms (the 'epistemic vigilance module(s)'), which assess the believability of a piece of communicated information and act as a filter at the entrance of the 'belief box' of the interpreter. Since communication is open to the risk of misinformation (be it accidental or intentional), the only way for it to remain advantageous (on average) for both communicator and audience is for its outcomes to be assessed by mechanisms that monitor the quality of the incoming information and the reliability of the individuals who dispense it before accepting it.[4]

The metarepresentational output of the pragmatics module, i.e. 'The $speaker_x$ meant that p_1, .., p_n' (where p_1, .., p_n is the set of propositions communicated), provides the input to two different kinds of epistemic vigilance mechanisms: mechanisms that focus on the source of information (*who* is to be believed) and mechanisms that focus on the informational content itself (*what* is to be believed). While the former assess the reliability of the speaker, that is, whether the speaker is competent (epistemically reliable) and benevolent (morally reliable),

[4] It may seem that epistemic vigilance mechanisms have developed entirely for the sake of the interpreter but they can be advantageous for the communicator too: "from the communicator's point of view, a vigilant addressee is better than one who rejects her testimony outright" (Sperber *et al*. 2010: 376).

the latter assess the degree of believability of the incoming information (the propositions p_1 …, p_n), independently from its source.

Epistemic vigilance mechanisms towards the source of information can deliver either general impressions of trustworthiness (e.g. on the basis of facial clues, see Willis and Todorov 2006) or more costly assessments that result from context-sensitive evaluations of the reliability of the speaker. For instance, given an input of the form 'The speaker$_x$ meant that p_1, .., p_n', they may assess whether the speaker may want the audience to believe the set of propositions p_1, .., p_n for reasons that do not concern their (alleged) truth (e.g. because of some deceptive intention). Or they may detect that the set of propositions p_1, .., p_n is warranted by some beliefs of the speaker that are, in fact, false. In both cases, they would prevent p_1, .., p_n from entering the belief box of the interpreter.

Epistemic vigilance mechanisms towards the content, on the other hand, assess the believability of the incoming information relative to the context of the addressee's existing beliefs (which are themselves, of course, open to revision). Specifically, according to Sperber *et al*. (2010), the beliefs against which the communicated information is tested are those that are automatically activated by the comprehension process and used in the pursuit of relevance. These are a subset of the mental encyclopaedia of the addressee, and provide the ground for an "imperfect but cost-effective epistemic assessment" (Sperber *et al*. 2010: 374). When the result of this assessment is a contradiction, there are three possible outcomes: (i) if the source is taken as trustworthy and the background beliefs of the interpreter that conflict with the incoming information are not held with much conviction, these beliefs are corrected; (ii) if the source is not regarded as trustworthy, the new information is rejected; (iii) if the source is regarded as authoritative and the conflicting background beliefs are held confidently, some process of (typically conscious) coherence checking is triggered. Interestingly, the choice among (i), (ii) and (iii) partly depends upon the output of epistemic vigilance mechanisms focused on the source (the speaker).

To sum up, the pragmatics module is conceived of as interconnected with the epistemic vigilance module(s), whose mechanisms assess the reliability of the source of information and the believability of its content. These mechanisms are geared towards preventing the interpreter from being misinformed by filtering the communicated information that he ends up believing. In the next section, I explore the possibility of extending the scope of interaction between the pragmatics module and epistemic vigilance mechanisms (see also Mazzarella 2013) and highlight its implications for the cognitive architecture of the mind.

3.2 Does epistemic vigilance affect the comprehension process?

My proposal is that, not only do epistemic vigilance mechanisms assess the believability of a piece of communicated information (as suggested by Sperber and colleagues), but they also contribute to the assessment of whether an interpretative hypothesis about the speaker's meaning is to be retained and attributed to the speaker as the intended interpretation. In other terms, they play a role in both 'acceptance' of content and the logically prior 'comprehension' of the speaker's meaning (her intended content).[5] That is, as well as assessing whether an interpretation attributed to the speaker (i.e. the output of the comprehension

[5] See Padilla-Cruz 2012 for a different proposal along the same line.

procedure) is allowed to enter the 'belief box' of the interpreter, epistemic vigilance mechanisms determine whether an interpretative hypothesis about the speaker's meaning ends up being the output of the comprehension module or not.

This suggestion is grounded on the well-established relationship between interpretation and trust, but it offers a new cognitively oriented perspective in which to frame such a relationship. Before exploring it, let us focus on the relevance-guided comprehension procedure, as presented in (3), and its stopping point (b).

(3) **Relevance-guided comprehension procedure**
  (a) Follow a path of least effort in computing cognitive effects: Test interpretative hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.
  (b) Stop when your expectations of relevance are satisfied.

Sperber (1994a) suggests that the relevance guided comprehension procedure comes in three different versions: 'naïve optimism', 'cautious optimism' and 'sophisticated understanding'. Interestingly, the difference between the three strategies relies on different assumptions about the communicator's competence and benevolence, which in turn raise different expectations of relevance (hence determine different stopping points in interpretation). A naïvely optimistic hearer takes for granted that the communicator is behaving both benevolently and competently: he takes the communicator to be competent enough to avoid misunderstanding, and benevolent enough not to lead him astray. Thus he expects 'actual optimal relevance'. In contrast, a cautiously optimistic interpreter assumes the communicator to be benevolent, but not necessarily competent. As a consequence, he looks for 'attempted optimal relevance'. Finally, a sophisticated interpreter drops not only the assumption that the communicator is behaving competently, but also that she is behaving benevolently. Then the expectations of relevance that guide the comprehension procedure and determine its stopping point are expectations of 'purported optimal relevance'. The following table illustrates the three different versions of the relevance-guided comprehension procedure (which differ with regard to clause (b)):

Three versions of the relevance-guided comprehension procedure:

	(a)	Follow a path of least effort in computing cognitive effects: Test interpretative hypotheses in order of accessibility
Naïve optimism	(b_1)	Stop when your expectations of *actual optimal relevance* are satisfied (i.e. stop at the first relevant enough interpretation)
Cautious optimism	(b_2)	Stop when your expectations of *attempted optimal relevance* are satisfied (i.e. stop at the first interpretation that the communicator *might have thought* would be relevant enough to you)
Sophisticated understanding	(b_3)	Stop when your expectations of *purported optimal relevance* are satisfied (i.e. stop at the first interpretation that the communicator *might have thought would seem* relevant enough to you)

To appreciate the difference between these interpretative strategies, consider again example (1). If Robyn were a naively optimistic interpreter, she would attribute to Sarah the first interpretative hypothesis that is relevant enough to

her. The first interpretation that comes to Robyn's mind is that $Neil_1$ has broken his own leg this morning. This interpretation is relevant enough to Robyn, in fact, it is highly relevant to her (it allows her to derive many contextual implications, e.g. that she should immediately go to the hospital, cancel her afternoon meeting, etc.). Thus, a naïve interpreter would retain it and mistakenly attribute it to the speaker. But what if Robyn adopted the cautiously optimistic version of the relevance-guided comprehension procedure? Robyn would not take for granted Sarah's competence and she would be vigilant to the possibility that Sarah may not know what Robyn knows (and may consequently fail in her attempt to make the relevant information that she intends to convey more accessible than any other possible interpretation). Robyn would realise that Sarah could not have intended the interpretative hypothesis '$Neil_1$ has broken his own leg this morning' to occur to her (precisely because she does not know that Robyn has a son). Sarah could not have thought that this interpretation would be relevant enough to her as she, Sarah, has no thoughts of any sort involving $Neil_1$. Thus, the comprehension procedure would go further and test the next most accessible interpretative hypothesis. For instance, it would access and assess the interpretation that $Neil_2$ has recently broken his own leg. Since Robyn takes it that Sarah might have thought this interpretation to be relevant enough to her (as in fact it is), the interpretation is selected and attributed to Sarah.

Utterance interpretation, thus, may depend on considerations about the speaker's competence (as in the example discussed above) and/or benevolence. The issue of what brings such considerations to bear on the interpretative process, however, has not been addressed within the literature. I suggest that epistemic vigilance mechanisms can modulate the hearer's expectations of relevance (i.e. from 'actual' to 'attempted' or 'purported' optimal relevance) and assess whether the interpretative hypothesis under construction satisfies these expectations. If the interpreter is vigilant towards the speaker's competence, for instance, he will expect 'attempted' optimal relevance'. As a consequence, he will stop at the first relevant interpretation that the speaker might have thought would be relevant to him (as described in the cautiously optimistic version of the relevance-guided comprehension procedure).

The issue to be addressed now is whether such extended interaction between the comprehension system and epistemic vigilance mechanisms is compatible with the thesis that pragmatics is a modular system. In what follows, I put forth a tentative suggestion to implement my proposal within a massively modular framework such as the one adopted by Relevance Theory.

From the perspective of the information flow through the architecture of the cognitive systems, the role of epistemic vigilance in the comprehension processes suggests that the epistemic vigilance module(s) does not receive its input only when the comprehension process is over. Rather, during the comprehension process, subparts of the interpretation are fed to the epistemic vigilance module(s) for its assessment. As a consequence, it may filter out interpretative hypotheses that are incompatible with the speaker's mental states (i.e. her beliefs and desires).

Consider an utterance of (1) in the context described above. Following a path of least effort, the interpreter starts "fleshing out" the propositional schema (delivered by the language decoding processor) '$Neil_x$ has broken his_y leg at some time *t*' by assigning to the proper name "Neil" the referent $Neil_1$ (this is because Robyn's concept of $Neil_1$ is more highly activated than her concept of

$Neil_2$). This part of the overall interpretative hypothesis is fed to the epistemic vigilance modules(s). It provides a hypothesised topic of conversation, that is, $Neil_1$, which in turn triggers epistemic vigilance mechanisms targeted at assessing the competence of the speaker (Sarah) on particular topics. These mechanisms access the piece of information that Sarah does not know that Robyn has a son called "Neil". Epistemic vigilance thus detects an incompatibility between the speaker's system of beliefs and the interpretative (referential) hypothesis under construction. As a consequence, it inhibits the comprehension procedure and prompts it to access (and assess) the next most accessible referential hypothesis.[6]

In general, epistemic vigilance mechanisms that monitor the speaker's competence and benevolence may restrict and direct the operations of the comprehension module. The role played by assumptions about the speaker's competence and benevolence in pragmatic interpretation has long been recognised (see Sperber 1994a) but its implications for the location of pragmatic abilities in the overall architecture of the mind have not been explored yet. I suggest that this role might be explained in terms of the interaction, as just discussed, between the comprehension module and epistemic vigilance mechanisms. I think this also opens up an interesting direction of research for developmental pragmatics, to which I now turn.

3.3 Developmental implications

Sperber (1994a) suggests that the three versions of the relevance guided comprehension procedure ('naïve optimism, 'cautious optimism' and 'sophisticated understanding') may correspond to different stages in pragmatic development. That is, children may start out as naïvely optimistic interpreters and progressively acquire the ability to monitor the speaker's competence and benevolence and to adapt their interpretative behaviours to these.

This gives rise to the interesting question of what allows the progression from naïve optimism to the further developmental stages. Sperber (1994a) claims that the three versions of the relevance-guided comprehension procedure require the interpreter to manipulate increasingly higher order representations of mental states. For this reason, Carston (1997) and Wilson (2005) have suggested that the development of more sophisticated interpretative strategies may correlate with the emergence of more complex mind-reading abilities: the move from naïve optimism to cautious optimism may correlate with the emergence of first-order mind-reading ability, the one from cautious optimism to sophisticated understanding with the emergence of second-order mind-reading abilities.

While this suggestion is certainly worth exploring further, the recent work on epistemic vigilance by Sperber *et al*. (2010) seems to open further interesting scenarios for developmental pragmatics. In light of the hypothesized interaction between the comprehension system and epistemic vigilance mechanisms, it seems plausible to assume that these three stages in the development of pragmat-

[6] A similar account is proposed by Sperber and Wilson (1986/1995: 186-87) with regard to the relationship between a decoding module and a central inferential system. Thanks to Deirdre Wilson for pointing this out to me and suggesting that the relation between decoding and inferential comprehension and between comprehension and epistemic vigilance mechanisms could be framed in the same way.

ic abilities may follow a similar developmental trajectory to that of epistemic vigilance capacities. Naïve optimism, cautious optimism and sophisticated understanding involve different assumptions about the communicator's competence and benevolence. As noted above, epistemic vigilance mechanisms focused on the source of information (*who* to believe) monitor the speaker's epistemic and moral reliability, that is, her competence and benevolence. Thus, it seems plausible to hypothesise that the emergence of epistemic vigilance mechanisms targeted at the assessment of the speaker's competence may correlate with (and prompt) the development of a cautiously optimistic interpretative strategy. Similarly, the move to sophisticated understanding may be triggered by the emergence of epistemic vigilance mechanisms monitoring the speaker's benevolence.

A very interesting and plausible picture emerges: the three interpretative strategies described above may be nothing but an *epiphenomenon* of the interaction between a *single* comprehension procedure and the epistemic vigilance mechanisms. On this construal, a 'cautiously optimistic interpreter' would be nothing but an interpreter who is vigilant towards the speaker's incompetence (i.e. her lack of knowledge or dependence on false belief), whereas vigilance towards the speaker's malevolence (i.e. her possibly deceptive intentions) would underpin the interpretative behaviour of a 'sophisticated interpreter'.

While the development of epistemic vigilance has been the subject of recent experimental investigation (e.g. Clèment, Koening and Harris 2004, Mascaro and Sperber 2009), an explicit comparison between the development of epistemic vigilance and of pragmatic competence remains to be carried out. Furthermore, there is a growing body of research on children's ability to track the communicator**'**s epistemic state and use this to infer what she intends to refer to (Carpenter, Call and Tomasello 2002, Southgate, Chevallier and Csibra 2010, i.a.). The implications of this literature for Sperber's (1994a) theoretical distinction between 'naïve optimism', 'cautious optimism' and 'sophisticated understanding' have not been assessed yet. Crucially, Southgate *et al.* (2010) show that 17-month-old infants can take account of the speaker's epistemic state (i.e. her false-belief) in reference resolution. In Southgate *et al.*'s study, the infants see the experimenter place two novel objects in different boxes and leave the room. An accomplice changes the position of the objects in her absence. When the experimenter returns, she points towards one of the boxes and says to the infant: "Do you know what's in here? Shall we play with it? Shall we play with it? Let**'**s play with it!" Finally she says, "Can you get it for me?" The issue here is which of the two objects the infant would take the word 'it' to refer to. The results showed that infants as young as 17-month-old were significantly more likely to choose the object in the box that the experimenter had not pointed to. What I would like to point out here is that the infant assigning the appropriate referent to the pronoun 'it' seems to require a cautiously optimistic interpretative strategy. That is, the infant should not stop at the first relevant referential interpretation (which corresponds to one where the referent is taken to be inside the pointed-to box), but rather at the first relevant referential interpretation that the experimenter could have thought would be relevant to him. This requires the infant to take account of the experimenter's epistemic state (i.e. her false belief that the intended object is in the pointed-to box) and reason that she could have not intended to refer to the object in the pointed-to box because she does not know that it has been swapped with the object in the non-pointed-to box. The

results raise the question of whether there is in fact any developmental stage corresponding to naïve optimism or whether naïve optimism is a theoretical construction without any empirical counterpart.

4. Conclusions

Relevance Theory has recently presented an account of pragmatics that is in line with an increasingly modular view of the mind. Pragmatics is conceived of as a sub-module of the mind-reading module, which exploits a dedicated inferential procedure that is attuned to the regularities of the domain of overt communication. As for any conceptual module, the comprehension module is part of a network of mental modules, which are highly interconnected with each other. This paper has focused on one of these connections, that is, the connection between the comprehension module and epistemic vigilance module(s). According to Sperber *et al*. (2010), epistemic vigilance mechanisms take as their proprietary input the output of the pragmatic module and assess its believability. In this paper, I make two further proposals: first, that epistemic vigilance mechanisms may affect not only the believability of the output of the relevance-guided comprehension procedure, but also the comprehension procedure itself; second, that the emergence of epistemic vigilance targeted at assessing the speaker's competence and benevolence may correlate with different developmental stages in pragmatics. These proposals suggest a programme of future research in cognitive and developmental pragmatics.

References

Carpenter, M., Call, J., and Tomasello, M. 2002, "A new false belief test for 36-month-olds", *British Journal of Developmental Psychology*, 20, 393-420.

Carston, R. 1997, "Relevance-theoretic pragmatics and modularity", *UCL Working Papers in Linguistics*, 9, 1-27.

Carston, R. 2002, *Thoughts and utterances: The pragmatics of explicit communication*, Oxford: Blackwell.

Carston, R. 2007, "How many pragmatic systems are there?", in Frápolli M.J. (ed.), S*aying, meaning, referring. Essays on the philosophy of François Recanati*, New York: Palgrave Macmillan, 18-48.

Clément, F., Koenig, M., and Harris, P. 2004, "The ontogeny of trust", *Mind & Language*, 19, 360-79.

Cosmides, L., and Tooby, J. 1992, "Cognitive adaptations for social exchange", in Barkow, J.H., Cosmides, L. and Tooby, J. (eds.), *The adapted mind: Evolutionary psychology and the generation of culture*, New York: Oxford University Press, 163-228.

Cosmides, L. and Tooby, J. 1994, "Origins of domain specificity: The evolution of functional organization", in Hirschfeld, L.A. and Gelman, S.A. (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge: Cambridge University Press, 85-116.

Fodor, J.A. 1983, *The Modularity of Mind*, Cambridge, MA: MIT Press.

Kasher, A. 1991, "On the pragmatic modules: A lecture", *Journal of Pragmatics*, 16, 381-97.

Mascaro, O. and Sperber, D. 2009, "The moral, epistemic and mindreading components of children's vigilance towards deception", *Cognition*, 112, 367-80.

Mazzarella, D. 2013, "'Optimal relevance' as a pragmatic criterion: the role of epistemic vigilance", *UCL Working Papers in Linguistics*, 25, 20-45.

Padilla Cruz, M. 2012, "Epistemic vigilance, cautious optimism and sophisticated understanding", *Research in Language*, 10 (4), 365-86.

Southgate, V. Chevallier, C. and Csibra, G. 2010, "Seventeen-month-olds appeal to false beliefs to interpret others' referential communication", *Developmental Science*, 13 (6), 907-12.

Sperber, D. 1994a, "Understanding verbal understanding", in Khalfa, J. (ed.), *What is Intelligence?*, Cambridge: Cambridge University Press,179-98.

Sperber, D. 1994b, "The modularity of thought and the epidemiology of representations", in Hirschfeld, L.A. and Gelman, S.A. (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge: Cambridge University Press, 39-67.

Sperber, D. 2001, "In defense of massive modularity", in Dupoux, E. (ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler*, Cambridge, MA: MIT Press, 47-57.

Sperber, D. 2005, "Modularity and relevance: How can a massively modular mind be flexible and context-sensitive?", in Carruthers, P., Laurence, S. and Stich, S. (eds.), *The Innate Mind: Structure and Content*, New York: Oxford University Press, 53-68.

Sperber, D. & Wilson, D. 1986/1995, *Relevance: Communication and Cognition*, Oxford: Blackwell.

Sperber, D. and Wilson, D. 2002, "Pragmatics, modularity and mind-reading", *Mind & Language*, 17 (1&2), 3-23.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. and Wilson, D. 2010, "Epistemic vigilance", *Mind & Language*, 24 (4), 359-93.

Willis, J. and Todorov, A. 2006, "First impression: making up your mind after a 100-ms exposure to a face", *Psychological Science*, 17, 592-98.

Wilson, D. 2005, "New directions for research on pragmatics and modularity", *Lingua*, 115, 1129-1146.

Wilson, D. and Sperber, D. 1986, "Pragmatics and modularity", in Farley, A.M., Farley, P.T. and McCullough, K.E. (eds.), *Chicago Linguistic Society Parasession on Pragmatics and Grammatical Theory*, 67-84, reprinted in Davies, S. (ed.), *Pragmatics: A Reader*, New York: Oxford University Press, 1991, 583-95.

Enactivism, Representations and Canonical Neurons

Gabriele Ferretti and Mario Alai***

* *Universiteit Antwerpen, Università di Urbino Carlo Bo*

** *Università di Urbino Carlo Bo*

Abstract

Enactivists often claim that since perception is one with action, it does not involve representations, hence perception is *direct*. Here we argue that empirical evidence on neural activity in the *ventral premotor cortex* confirms the enactivist intuitions about the unity of action and perception. But this very unity requires the detection of the action possibilities offered by the objects in the environment, which in turn involves certain representational processes at the neural level. Hence, the enactivist claim that perception is direct is wrong, or at least ambiguous and potentially misleading: in one important sense perception involves representations.

Keywords: Enactivism, Visuomotor Representations, Canonical Neurons, Affordances, Transparency of Perception, Sensorimotor Processes, Ventral Premotor Cortex, Direct Perception

Introduction

In current cognitive studies various different approaches are called enactivist, as they stress that our cognition is based on the unity of action and perception (henceforth: UAP): this is how we perceive and act upon the sensorimotor contingencies (i.e. affordances) found in the environment. On the basis of UAP enactivism claims that representations are not involved in perception, and perception is direct. There is wide empirical evidence in neuroscience confirming and explaining UAP and the role of affordances, but these explanations and confirmations crucially involve subpersonal representations: so we argue, against enactivism, that perception is indirect.

In § 1 we report the basic enactivist intuitions—exemplified by Noë 2004—on UAP, affordances, and the related claim that perception is direct. After focusing especially on the relation between enactivism and the concept of representation, we then turn to the two basic goals of this paper:

Argumenta 1, 2 (2016): 195-217
ISSN 2465-2334

© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.FER

A) In § 2 we show that empirical evidence on neural activity in the *ventral premotor cortex* confirms the enactivist intuitions about UAP and affordances. However, we also make clear that we can map the external objects thanks to the fact that the neural correlates of UAP perform representational processes.
B) Therefore in § 3 we claim that the enactivist doctrine of direct perception is potentially ambiguous, and might be wrong in an important sense: since UAP involves representations, perception is indirect.

This is not a survey of the various kinds of enactivism, each with its own philosophy concerning UAP, nor do we discuss the most recent debates on them[1]. For despite their differences, these various views share the intuition of UAP, and we stress that this intuition is empirically well confirmed. It is remarkable that the literature on enactivism never mentions these empirical findings, and above all, it does not notice the lurking contradiction between UAP and the claim that perception is direct.

1. Enactivism and UAP

According to Gibson (1986) we can explore reality through an indivisible movement-perception system, perfectly *attuned* with the visual invariants in the *umwelt*, that allows us to act: perception controls movements and movements are fundamental to get perceptual information. Gibson's primary concept is that of *affordance*[2]: the environmental opportunities for action that an object offers to an agent. In his view we can profit of affordances thanks to UAP: realizing what an object is (in the visual dimension) is realizing what we can do to and by it (in the motor dimension)[3]. Thus, against a common idea of classical cognitive sciences, he claims that, since our perceptual system *resonates* to the properties of objects, perception is direct.

Enactivism (also called *sensorimotor paradigm*) is heir to his view: it "question(s) the centrality of the notion that cognition is fundamentally representation" (Varela, Thompson and Rosch 1991: 9), and it was mainly developed in Clark 1998, O'Regan and Noë 2001, 2004. Many cognitive approaches sail under this banner, but they share at least these assumptions:

E1 UAP: i.e., perception consists in (and depends on) mastery of sensorimotor skills: it is based on the interdependent availability of perceptually conducted motor behaviour associated with the related sensory consequences, and it allows us to act upon the objects in accordance with their *sensorimo-*

[1] See Hutto and Myin 2013; Thompson 2007; Stewart, Gapenne and Di Paolo 2010; Gangopadhyay and Kiverstein 2009.
[2] Affordances are a visual process making us aware of the possibilities of action upon the object, not to be confused with the motor act we can perform on the (object itself on the basis of the) affordance. Which affordances we get in a given situation depends both on the characteristics of the environment (optic array, outlines, objects), and on the acting individual (its body, skills, etc.). Obviously, the relationship between motor acts and affordances can change with respect to different purposes.
[3] See § 2.

tor contingencies[4], i.e., affordances, the various ways objects afford our behaviour.

E2 It follows from **E1** that we have an immediate visualization of the sensorimotor contingencies and that perception is direct.

These basic assumptions are exemplified in Noë's account of enactivism, one of the most influential: perception "depends on the possession and exercise of a certain kind of practical knowledge (*know-how*)" (Noë 2004: 33). What we perceive is in function of the way we act, and the way we act is an aspect of perceptual processes. Perception (in particular visual perception) has been evolved to help motor control, it is part of a procedure aimed at achieving some purpose (Noë 2009). The experience of an object consists in the set of actions involved in perceiving the object (O'Regan 2011; Zipoli Caiani 2013b; for a complete review see Hutto and Myin 2013, Ferretti, forthcoming-b).

> […] perceiving is a way of acting. Perception is not something that happens to us, or in us. It is something we do. […] The world makes itself available to the perceiver through physical movement and interaction […] all perception is touch-like in this way: perceptual experience acquires content thanks to our possession of bodily skills. What we perceive is determined by what we do (or what we know how to do) (Noë 2004: 1).

Like Gibson, Noë (2004: 2) rejects the view of perception inspired by Marr (1982), whereby perceptual processes in the brain create detailed inner representations of the external environment. Instead, he claims that perception is direct. He rejects the "snapshot conception"[5], according to which the world is given to us as rich in details all at once: he denies that when we see we *represent* the whole scene in consciousness all at once in the way a photograph does (Noë 2004: 63, 72-73, 218-219). "There's no need to build up a detailed internal model of the world" (Noë 2004: 50). Our attention permits us to perceive just a portion of the scene and only a few objects. This is possible thanks to "the way in which objects structure and control our movements and our sensory stimulation" (ibi).

According to some authors enactivism nonetheless admits a form of representations, closely tied to bodily activities (or skills) (Wilson 2004: 186; Menary 2006: 3-5). In fact, at least in one passage Noë grants that "No doubt perception depends on what takes place in the brain, and very likely there are internal representations in the brain (Noë 2004: 2). But it is not clear what these representations are. We agree that there are representations, but in § 2, while supporting E1,[6] we offer a clearer analysis of what they are, and how they can explain bodily skills. In our opinion, the fact that we do not *represent* the entire scene does not mean that the objects we "select" on the scene are not represented; in fact, the perception of sensorimotor contingencies requires subpersonal visuomotor

[4] Noë treats 'sensorimotor contingencies' as a near synonymous of 'affordances': see Noë 2004: 105-106.

[5] See Noë 2002. He also uses experiments like those concerning *inattentional blindness*, *reverse lenses (inverting goggles)* and *change blindness*.

[6] Of course not all perception is aimed to action. But here we will focus on the precise portion of perception that is functional to action (see Milner and Goodale 2005; Jacob and Jeannerod 2003).

representations (henceforth: SV representations). Here we do not focus on the snapshot conception of representation. We just claim that rejecting the "snapshot conception" is not sufficient to show that perception is direct, i.e., that no representations are involved in perception. Therefore in § 3 we reject E2.

2. Empirical evidences

Studies on monkeys show that the brain's motor system is not involved only in executive functions. The motor system includes area 6, the premotor cortex, occupying the posterior portion of frontal lobe (the cortical region directly involved in voluntary movement), and area 4, the primary motor area (Fig. 1). Area 6 is not homogeneous, and it can be divided into dorsal premotor cortex (F7 and F2) and ventral premotor cortex (F4 and F5) (Fig. 2).

We will particularly focus on F5, which is directly connected with the primary motor area (area 4): it receives nervous signals from the parietal lobe, which for a long time was considered an associative area for sensory operations. Other studies demonstrated that the motor cortex influences the perceptual side of the parietal lobe, which is now considered a part of the motor brain (Mountcastle *et al.* 1975; Sakata *et al.* 1995).[7] Thus, the idea of a *motor system* exclusively involved in motor roles is now dismissed: the motor cortex is basic for sensory operations as well. As we shall see in § 2.2, these evidences already confirm UAP;[8] but more can be found by studying F5 (Matelli, Luppino, Rizzolatti *et al.* 1985).[9]

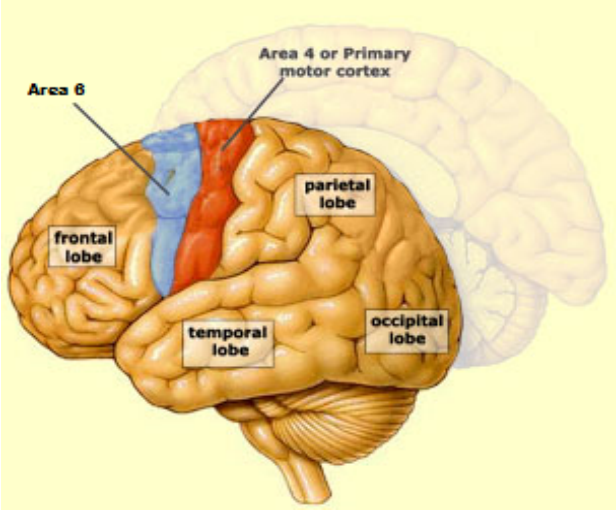

Fig. 1. Cortical division into lobes.[10]

[7] Cited in Gallese 2000: 27; see also Rizzolatti and Sinigaglia 2006.
[8] More evidences in § 2.2.
[9] Cited in Gallese 2003: 1235. Here we do not discuss the problem of the two cortical visual streams in relation with these points: see Noë and Thompson 2002; Rizzolatti and Sinigaglia 2006; Fridland 2012.
[10] Adapted from: http://www.docstoc.com/docs/122202947/9-Corteccia-motoria

Fig. 2. Premotor cortex (area 6): Dorsal (F7 and F2) Ventral (F4 and F5).[11]

2.1 F5: subpersonal visuomotor representations, goal relatedness and canonical neurons

The area called F5 occupies the most rostral part of the ventral premotor cortex. F5 neurons contain a distal hand and mouth movement representation (see Rizzolatti *et al.* 1981, 1988; Kurata and Tanji 1986).[12] F5 includes two large groups of neurons: the first is that of (A) *purely motor neurons*, whose activation is exclusively connected to actual movements. They constitute the overall majority of all F5 neurons, and belong to two kinds: (A1) neurons that fire whenever a movement is performed, and (A2) neurons that code only successful agent-object relationships, i.e., the achievement of a goal; for instance, grasping-related neurons fire whenever the monkey *successfully* grasps an object, regardless of the *effector* (i.e. the particular limb, or the particular organ employed) (see Rizzolatti *et al*. 1988; Rizzolatti, Fogassi, and Gallese 1999).[13] Therefore these (A2) neurons in F5 are called *goal-related* neurons.

But there is also a second very interesting large group of neurons in F5, that of (B) *visuomotor neurons*: they have not only motor properties indistinguishable from those of the purely motor neurons, but also peculiar "visual" properties: in experimental tests the purely motor neurons fired during the grasp, while visuomotor neurons fired significantly also during the visualization, regardless whether a grasp followed or not. Visuomotor neurons are also distinguished in two groups: (B1) the so called *canonical neurons*, which discharge when an objects is presented, even if no detectable action aimed at them is performed, either by the monkey or by the experimenter (Rizzolatti and Fadiga 1998; Rizzolatti, Fogassi, and Gallese 1999); (B2) the famous *mirror neurons*, which respond when the monkey observes an action performed by another individual, or when it performs the same or a similar action (Gallese *et al*. 1996; Rizzolatti *et al*. 1996).[14]

A fundamental functional property of area F5 is that most of its neurons do not encode *elementary movements* (like the mere extending of my arm), but *motor*

[11] Originally in Rizzolatti, Luppino and Matelli 1998: 285.
[12] Cited in Gallese and Metzinger 2003.
[13] Cited in Gallese 2003: 1235.
[14] Cited in Gallese 2003: 1236.

acts (or coordinated movements with specific purposes, like moving my arm in a specific direction to catch a glass) (Rizzolatti *et al.* 1988). The same elementary movement activating a neuron during a specific motor act (e.g., grasping) doesn't activate it during a different motor act (e.g., scratching). Thus, there are different groups of neurons in F5: grasping neurons, grasping-with-the-mouth neurons, hugging neurons, etc.

The activity of canonical neurons is characterized by "a strict congruence between their high selectivity for a particular type of prehension (executed grip) and the visual selectivity for objects that, although differing in shape, … require the same type of prehension in order to be grasped" (Gallese 2000; Murata *et al*. 1997; Rizzolatti *et al.* 1988; Rizzolatti, Fogassi, and Gallese 1999).[15] Imagine you have to grasp first a little box that can be contained inside your hand, and then a little stone: although their shape is different, these objects show the same affordance; therefore the motor acts satisfying this affordance (grasping the object with the whole hand) are the same.

In experiments with monkeys, just as the subject looks at the object its neurons fire, activating the motor program that *would be* involved were the observer actively interacting with the object. The evoked motor pattern remains just a potential act. Hence, the identification of an object is a preliminary form of action, a call to agency, characterized on the basis of its (visuo)motor opportunities, independently of whether an execution shall occur or not. This shows that in the recognition of objects agency and perception are two sides of the same coin: the sight guiding the hand is a kind of capacity to watch through the hand: the object that we perceive is encoded as a determined set of hypotheses of actions (Rizzolatti and Sinigaglia 2006: 44-47).

A very interesting result, correlated with the data described above is that mental action is a form of *neurophysiological simulation* of the physical action (Clark, Tremblay and Ste-Marie 2004). In particular, *motor imagery* and executed actions share similar physiological correlates (kinematic contents, dynamic changes in physiological parameters, functional anatomy). The (overt) execution of an action is necessarily preceded by its (covert) representation, but a (covert) representation is not necessarily followed by an (overt) execution of that action. This suggests that the representation "can actually be detached from execution and can exist on its own" (Jeannerod 2006: 2; see also Decety and Perani 1994; Decety and Ingvar 1990; Grafton *et al*. 1996; Johnson *et al*. 2002; Mühlau *et al*. 2005; Rumiati *et al*. 2004; Jeannerod 1994, 1997).

The activation of motor areas in representing an action provides the represented action with a 'motor' format, like the involvement of primary visual cortex in visual mental imagery restores the topographical layout of the image (see Kosslyn 1996, 2005). In order for a represented action to be felt as real, it needs to be framed within the constraints of a real action (Jeannerod 2001), so that it can be regarded by the motor system as the simulation[16] of a real action. Thus, *motor imagery* is a *prototypical form of action representation*, or a *representation of evoked motor responses*, even though neural commands for muscular contractions are effectively present during motor imagery, but simultaneously blocked at

[15] Cited in Gallese 2003: 1236; Gallese 2000: 31.

[16] According to Metzinger (2003: 49-50), simulations are internal representations of possible properties of the world *in general*, while mental representations are the *special case* in which *actual* properties are simulated. See also footnotes 18 and 19 below.

some level of the motor system by an active inhibitory mechanisms (Jeannerod 2006). In other words, the fact that these representations can be regarded as simulations of real actions, shows that they have a motor format, i.e. a format framed within the constraints of the real action.

In § 2.2 and § 2.3 we shall see that the interdependence of motor and visual selectivity in canonical neurons is linked to representational mechanisms. This is an interesting kind of abstraction: looking at objects is unconsciously "simulating" a potential action; the *representation* of an object (based on the visualization of its affordance) is *integrated* with the ongoing simulation of the precise *potential* action which could be performed upon the object (Gallese 2000: 31). Hence, seeing an object is getting at the same time the *subpersonal visuomotor* (SV) representation of its affordance, and the internal *simulation* of one of the actions we could perform upon it (i.e. the most suitable motor program required to interact with it). In fact, SV representations, which are representations of those visual aspects of a target that are relevant to the action to be performed, *translate* the geometrical features of the target object into opportunities for action, and the visuomotor transformation mechanism converts visual information into motor commands of arm and hand movement towards the object (Jacob 2005: 248). Thus, this transformation mechanism reads the affordances as motor acts that can achieve one's goal (see § 2.2 below).

Furthermore, these SV representations allow us, as we perceive an object, to automatically compute the most suitable motor act that could be performed on it for some purpose (say, the way I can grasp it: Butterfill and Sinigaglia 2012; Jacob and Jeannerod 2003; Jeannerod 2006). This idea of motor simulation[17] has been proposed as a general framework for motor cognition, as the

[17] For neurophysiological technical details and a more accurate philosophical distinction between the concept of motor simulation and that of motor representation, see Jeannerod 2006, Ch. 2 "Imagined Actions as a Prototypical Form of Action Representation", and Ch. 6 "The Simulation Hypothesis of Motor Cognition". His data show that motor activation is highly specific to the action that is represented. Both forming motor images and observing other people acting involve motor cortex, premotor cortex, SMA, the basal ganglia and the cerebellum, i.e. the main neural structures which are needed in executing actions (here however we cannot discuss simulation in the case of mirror mechanisms and mirror neurons processes). For a review see Jeannerod 1994, 1997, 2001, 2006; Decety and Ingvar 1990; Hommel *et al*. 2001: 860. What differentiates Jeannerod's proposal from the others is that according to him: "We do agree that actions are represented in terms of their goal: but we assume that the goal is only part of the content of the action representation. By representing the goal, we can answer the question of 'What the action is about', but not the question of 'How to do it'. The latter question requires motor simulation to be answered (and we saw how important it may be to answer it for mentally rehearsing an action or learning it by observation)" (Jeannerod 2006: 134). Furthermore, an interesting philosophical analysis concerning the concept of simulation is offered by Metzinger 2003, according to which mental representation is a *special case* of mental simulation: simulations are internal representations of possible properties of the world. Representations, instead, concern actual properties of the world (49-50). Here mental simulation is a form of internalized *motor behaviour* and can be compared to goal-representing states. Empirical frameworks very similar to this have been proposed, but they understand simulation as an *internal forward model of motor* consequences (Wolpert *et al.*, 1995; Wolpert 1997; Kawato 1997, 1999; Grush 2004), or as prediction (Friston 2009; Clark 2013). For further interesting discussions see Leopold, Logothetis 1999; Kukla 1992: 222.

basic mechanism for explaining the functioning of motor representations[18]. Simulation is what makes it possible to activate perceptual mechanisms in the absence of a stimulus, or to activate motor mechanisms without executing an action. "If motor cognition is based on simulation of our own actions, and if the mechanisms each individual uses to simulate his own actions is the same as that other individuals use, as we have good reason to believe, then we can develop the idea that perceiving and producing actions are the two faces of the same process […]. If the assumption that represented actions correspond to covert, quasi-executed actions is correct, then represented actions should involve a simulation of the mechanisms that normally participate in the various stages of action generation, including motor execution" (Jeannerod 2006: 130-31).

It should be noticed that what takes place in F5 can be easily described through the concept of representations, and "through the idea of a crystallization of motor codes as stable functional units within the brain. Since the movements codified in the abovementioned cortical areas have a somatotopic organization in the motor system[19], and the stimulation of the same cortical site always produces the same complex response, it is natural to conceive the cortical activity in those sites as representations of those evoked motor responses. In other words, each particular site includes a series of spatial and temporal directives. For instance, the area concerned with grasping and taking to the mouth includes directives about which muscles of the hand, wrist, arm and mouth must be contracted, and when" (Caruana and Borghi 2013). And all this is controlled by the *visuomotor transformations* happening in AIP-F5 (discussed in § 2.2 below).

Thus it seems to us that empirical evidence is best interpreted by models in which representations play a role in natural cognitive systems. This however is more a methodological or epistemological hypothesis on how to do cognitive science, than a metaphysical claim, which at least for the time being we cautiously prefer to avoid.[20]

[18] "This is not to say that activation of the motor system is the same during simulation (in its various forms) as during execution and various forms of simulation: simulating is not doing, and substantial differences are observed between simulation and execution. First, the activation of most of the areas of the motor system during action representation is consistently weaker than during execution. Secondly, it is coupled with an additional mechanism for suppressing motor output, a prerequisite for the off-line functioning of the representation. Thirdly, because the muscles do not contract and the limbs do not move, the sensory reafferences normally produced by a movement are lacking. These differences are sufficient to disentangle simulating from doing. However, the representations for executing and simulating do not completely overlap, which may allow this distinction to be made even in the absence of sensory reafferences" (Jeannerod 2006: 131).

[19] I.e., groups of neurons related to adjacent parts of the body are themselves adjacent, so that the control of the movement of different parts of the body is centered in specific regions of the cortex.

[20] Chemero (2000) calls respectively 'Nature hypothesis' and 'Knowledge hypothesis' what we have called here "epistemological hypothesis" and "metaphysical claim": "The main difference between the nature and knowledge hypotheses can be put as follows: the knowledge hypothesis is to a much greater extent a *(meta)scientific* hypothesis. That is, the knowledge hypothesis concerns how we ought to do cognitive science, whatever the mind is really like. The nature hypothesis, on the other hand, is to a much greater extent a *philosophical* hypothesis; it concerns what the same region of the world (cognitive agents) is really like, however that region is best explained scientifically" (Chemero 2000.

2.2 AIP-F5, Affordances, and Enactivism

Area F5 is highly connected and interacts with the anterior intraparietal area (AIP), whose neurons are activated during hand movements. We can call this unified system AIP-F5.[21] AIP-F5 involves the visuomotor transformations necessary to grasp an object (Sakata *et al*. 1995; Murata *et al*. 2000).[22] AIP-F5 neurons are selectively responsive to tridimensional stimuli, so supporting Gibson's intuition: the visualization of the object and the related affordances activate neural groups in AIP. Visual information is first elaborated in AIP, then it passes over to F5 visuomotor canonical neurons, which don't encode the individual affordances already visualized in AIP, but the potential motor acts congruent with them. Thus, F5 allows to act upon the object, selecting the best motor behaviour thanks to the previous visual information based on the affordances. The selected action does not depend only on the intrinsic properties of the object, but also on the use we make of it and its purpose (Rizzolatti and Sinigaglia 2006: 35-36). Visual information is then translated into motor information and sent into F1, the region involved in action execution. AIP-F5 is the cortical portion in which the visuomotor transformations occur. The link between AIP and F5 is much more complex than we describe here, but the details we offer here are sufficient in order to make our philosophical point. For a complete review of their activity in line with computational neuroscience and studies on single cell recordings see Chinellato and del Pobil 2015, Borghi and Riggio 2015; for an overview of the cortical areas which are very akin to the AIP-F5 circuit see Turella and Lignau 2014.

The set of actions and motor behaviours we can perform in the environment is thus inscribed into the cortex: every object offers several affordances, hence several possible ways of acting upon it. However, each time we exploit just a small set of these possibilities. Indeed, during our ontogenetic development, the pruning of our neural networks under the pressure of experience selects in F5 the few neural populations linked to the most effective motor acts. This learning mechanism is called "motor reinforcement". Thanks to this functional selection our cortex structures a sort of *motor vocabulary* (see below) that facilitates the combination between the motor acts encoded by F5 and the visual affordances abstracted in AIP (Rizzolatti and Sinigaglia 2006: 45).[23]

This is further confirmed by more recent empirical evidences:[24] concerning the motor functions of the parietal lobe, Fogassi and Luppino (2005) reported new data confirming the general consensus that the posterior parietal cortex is part of the motor system and plays a fundamental role in visuomotor transformations. Bonini *et al.* 2014 show the leading role of the supplementary motor area (SMA) in the capacity to evaluate the outcomes of our actions; this capacity is fundamental for adapting and optimizing behaviour and depends on an ac-

For this debate see Chemero 2000, 2009; Hutto 2013; van Gelder 1995, 1998; Dennett 1987; Brooks 1991, 1999; Clark and Toribio 1994).

[21] See also Castiello 2005.

[22] Cited in Rizzolatti and Sinigaglia 2006.

[23] Here we cannot deal with the so called *non-inferentiality* of perception, although a comparison might be fruitful: see Zipoli Caiani 2013a; Noë 2004; Rizzolatti and Sinigaglia 2006.

[24] We are grateful to an anonymous referee for calling our attention to this topic.

tion-monitoring system that assesses ongoing actions and detects errors and rapidly evaluates successful and erroneous actions.

Raos *et al.* (2006) further investigated the motor and visual properties of F5 grasping neurons, using a controlled paradigm that allows the study of the neuronal discharge during both observation and grasping of many different three-dimensional objects, with and without visual guidance. Hierarchical cluster analysis indicated that the selectivity of both the motor and the visual discharge of the F5 neurons is determined not by the shape of the object, but by the grip posture used to grasp the object. All neurons displayed a *preference* for grasping of a particular object or set of objects.[25] The same preference was maintained when grasping was performed in the dark without visual feedback. In addition to the motor-related discharge, about half of the neurons also responded to the presentation of an object or a set of objects, even when a grasping movement was not required. Often the object evoking the strongest activity during grasping also evoked optimal activity during its visual presentation. Because the same paradigm has been used to study the properties of hand-grasping neurons in the dorsal premotor area F2 and in the anterior intraparietal area (AIP), these authors have been the first to compare the functional properties of grasping-related neurons in the three cortical areas (F5, F2, AIP).

Baumann, Fluet, and Scherberger (2009) provide compelling evidence that while a macaque is planning to grasp a single object (a handle), neurons in its parietal area involved in hand preshaping simultaneously encode multiple potential grasp movements before one is chosen for action. Other recent studies on affordances and motor system also confirm that F5 and AIP form a fronto-parietal network for transforming visual signals into hand grasping instructions (Brochier and Umilta, 2007; Brochier *et al*. 2004). Besides, these areas represent upcoming hand movements at a conceptual or categorical level (Townsend, Subasi, and Scherberger 2011) well before their execution (Raos *et al.* 2006; Baumann, Fluet, and Scherberger 2009; Fluet, Baumann, and Scherberger 2010; Townsend, Subasi, and Scherberger 2011).

Despite the fact that a single object can afford multiple types of grip, depending on the intended goal, so far studies have examined AIP neural activity only in monkeys trained to perform a single type of grip on a particular object. While the simplicity of such tasks has largely revealed the functional importance of AIP in grasping, the context and circumstances of everyday situations demand more flexibility in the selection of types of grip, and the role of AIP in facilitating such flexibility has remained unexamined. Hand grasping requires the transformation of sensory signals to hand movements. Neurons in area F5 (ventral premotor cortex) represent specific grasp movements (e.g., precision grip) as well as object features like orientation, and are involved in movement preparation and execution.

Fluet, Baumann and Scherberger (2010) examined how F5 neurons represent context-dependent grasping actions in macaques. Their results reveal important differences in how grip type and object orientation are processed in F5, and suggest that anatomically and functionally separable cell classes collaborate to generate hand grasping commands. The same authors addressed this issue by investigating AIP neural activity during a delayed grasping task in which monkeys were cued to grasp a handle at one of five different orientations, using ei-

[25] I.e., they preferentially fire in front of that object or set of objects.

ther a precision or a power grasp (Baumann, Fluet and Scherberger 2009). When handle orientation and grip type information were concurrently presented, AIP neurons showed sensitivity to handle orientation, grip type, or both, and could be classified according to their tuning onset (planning vs. movement execution) (ibi, Fig. 2*A-C*; see also Gallivan and Wood 2009). In order to perform grasping movements, the hand is shaped according to the form of the target object and the intended manipulation, which in turn depends on the context of the action. The anterior intraparietal cortex (AIP) is strongly involved in the sensorimotor transformation of grasping movements, but the extent to which it encodes context-specific information for hand grasping is unclear. Baumann, Fluet, and Scherberger (2009) showed that, in a cue separation task, when the object was presented first, neurons representing power or precision grips were activated simultaneously until the actual grip type was instructed. In contrast, when the grasp type instruction was presented before the object, type information was only weakly represented in AIP, but was strongly encoded after the grasp target was revealed. We conclude that AIP encodes context-specific hand grasping movements toward perceived objects, but in the absence of a grasp target, the encoding of context information is weak.

Some studies examined in detail the tuning properties of single units in both AIP (Sakata *et al*. 1995; Murata *et al*. 2000; Baumann, Fluet and Scherberger 2009) and F5 (Rizzolatti *et al*. 1988; Fluet, Baumann and Scherberger 2010). Besides, it has been showed that neural activity in F5 is better suited for the decoding of the grip type, while in AIP it is more accurate in predicting object orientation (Baumann, Fluet and Scherberger 2009; Fluet, Baumann and Scherberger 2010). The experimental evidence we reported about the AIP-F5 circuit concerns both human and non-human primates, and the correspondences between the respective cortical areas have been studied in the literature (see Borghi and Riggio 2015: 3; Shikata 2003).

AIP-F5 neural activity represents more than just a confirmation of Gibsonian and enactivist intuitions about the UAP and the related sensorimotor activity at cognitive level: it also constitutes a neural correlate of the UAP and affordances in the cortex on which the cognitive level depends. Moreover, in literature there is no agreement on whether the epistemic portion of affordances, which clearly depends on the subject, is to be classified as personal or subpersonal, or both; but AIP-F5 activity shows that affordances certainly play a role at subpersonal level (whether or not it also emerges at personal level).

That visuomotor canonical neurons have both visual and motor selectivity shows precisely that sensory phenomena involve motor neural mechanisms, while the premotor cortex has an important visual component: object recognition is possible thanks to F5 motor "vocabulary", which is the basis of the cognitive functions commonly attributed to the sensory system. For instance, take the following sequence: (i) I am thirsty. So, when (ii) I see a glass of water, (iii) I realize that I can satisfy my thirst by catching it, hence, (iv) I get ready to act accordingly. Well, the remarkable thing is that, thanks to the visuomotor transformation mechanisms, (ii) and (iii) (the affordances) and (iv) (the motor act) are all encoded in the same circuit, that is, AIP-F5, and activated by the firing of the same neural populations.

Empirical evidences confirm UAP in two ways. First, as said, the motor cortex is basic for sensory and perceptual processes (§ 2). In particular, recognizing the affordability of an object crucially involves (beside the visual dorsal

pathway) also the ventral premotor cortex. This means that sensory information is directly mapped on motor areas, so allowing us to perceive the environment in terms of possible motor acts.

For instance, an automatic process of grip formation takes place during the transportation of the hand: as we mentioned, the fingers are preshaped much before the hand touches the object.[26] Transportation of the hand itself involves initially a progressive opening, and then a closure of the fingers. This movement is completely automatic and subpersonal, and at about 60% of its transportation the hand reaches its widest opening, or *maximum grip aperture* (henceforth: MGA). It has been found that the size of the finger-grip at MGA (though much larger than the object to be grasped) is linearly correlated with the size of the object. This shows that the calibration of the finger-grip aperture is made automatically on the basis of a SV representations of the geometrical properties of the object (Jeannerod 2006; Jacob and Jeannerod 2003). "Thus, motor commands are generated such that the corresponding arm, hand and finger movements match the geometrical properties of the object to be grasped and handled (its location, size, shape and orientation). Simply observing the grasping hand reveals that this process is largely anticipatory and pertains to an action representation, not to a mere on-line adaptation of the motor commands to the object" (Jeannerod 2006: 5).

Actually, a very interesting proposal to model affordance perception in a dynamical anti-representationalist approach has been offered by Chemero (2009, § 7.6). However, the account based on representations has proven epistemologically[21] more fruitful in understanding how affordances are perceived. Moreover, data on AIP-F5, which provide the best available explication of these processes, are never mentioned by the radicalist accounts. We think that challenging the representational stance would be challenging those data.

Moreover, this already suggests that canonical neurons have a representational nature: for, how can the hand take the right aperture before reaching the object, unless it is guided by a representation of the object codified in its motor system? Thus, even though apparently the empirical data in cognitive neuroscience can be interpreted both from a representational and a non-representational point of view (Gallagher 2008, Hutto 2005; Hutto and Myin 2013), in this precise empirical framework it seems they do not admit of an equally good interpretation from the two perspectives: for both affordances and canonical neurons (and in general the cognitive processes occurring thanks to the dorsal visual stream and the AIP-F5 area) are better and more fruitfully interpreted in a representational framework (Jacob and Jeannerod 2003). Furthermore, as mentioned in the Introduction, while the literature on enactivism usually cites Milner and Goodale's two-visual-systems model, it never mentions the empirical evidences concerning the AIP-F5 cortical circuit.[27]

[26] Baumann *et al*. 2009 provide compelling evidence that during grasp planning toward a single object (a handle), neurons in a macaque parietal area involved in hand preshaping simultaneously encode multiple potential grasp movements before one is chosen for action.

[27] Actually Noë has been criticized by Hutto (2005) for failing to take into account subpersonal representations. But Hutto's argument—explained in (Menary 2006)—is purely philosophical, and it does not report any empirical data. Also Gallese and Keysers claimed that "Positing the importance of sensorimotor contingencies for perception is by no means denying the presence and importance of [subpersonal] representations"

The second kind of empirical evidences confirming UAP concerns the functional linkage between perception and action discussed in the literature about disorders due to cortical lesions. As said, evidences showed that the dorsal visual pathway (the pragmatic pathway of vision for action)[28] links the primary visual cortex (V1) to the (posterior) parietal lobe, hence to the AIP-F5 circuit. The latter automatically transforms the visual information about the properties of objects into motor commands, responding to those 3D geometrical properties of objects that serve such visuomotor tasks as grasping them (Milner and Goodale 1995; Matelli, Luppino and Rizzolatti 1985; Castiello 2005; Jacob and Jeannerod 2003).

Certain lesions in the dorsal pathway damage the visuomotor transformation process, causing impairments in spatial orientation and in the ability to localize objects in space; in such cases patients can still perceive and recognize objects, but they fail to grasp them normally, even though no disease occurs in the motor cortex (Milner and Goodale 1995, 2004; Ungerleider and Mishkin 1982; Gangopadhyay, Madary and Spencer 2010; Fridland 2012).

Indeed, posterior parietal lesions can cause optic ataxia, a deep alteration of reaching movements directed towards a visual target, in the absence of any motor impairment (Jacob, Jeannerod 2003; Jeannerod 2006). Three phenomena are observed in such cases: first, the kinematics of the movements is altered, since they have a lower peak velocity and the duration of their deceleration phase is increased. Such a deficit cannot have a motor origin, since the same movements can be executed with a normal kinematic profile in non-visual conditions. Second, the movements are not properly directed towards the target—their directional coding is impaired—and large pointing and reaching errors are observed (e.g., Milner *et al*. 1999). Third, alteration of the movements is not limited to the reaching phase. Distal aspects of the movements are affected as well. During the action of grasping an object, the finger grip aperture is increased, and the usual correlation between MGA and object size is lost (Jeannerod 1986). Similarly, optic ataxic patients fail to orient their hand properly when they have to insert an object through an oriented slit (Perenin and Vighetto 1988).

Optic ataxia appears to be a disorder limited to transforming visual properties of objects into motor commands for a hand action directed towards these objects. It is not due to misperception of the shape, orientation or size of the objects (see also Jeannerod *et al.* 1994; Goodale *et al*. 1994). Moreover, patients with parietal lesions, with or without optic ataxia, often present visuospatial disorders.

(Gallese and Keysers 2001: 983). They used findings about mirror neurons (see § 2.1 and § 2.2) to show "the intrinsic relationship between action control and representation within the logic of forward models" (Gallese and Keysers 2001: 983). However, their data are different from ours, as they concern a different (although similar) kind of neurons. Moreover, they do not enter into the kind of considerations we develop in § 3.

[28] According to Milner and Goodale's (1995) famous two visual systems theory, the ventral stream of visual processing is responsible for conscious qualitative perceptual experience, while the dorsal stream is responsible for the fine-grained motor coordination required for action instantiation. Milner and Goodale claim that the two streams interact insofar as the ventral stream selects the goals for action and the dorsal stream carries out the movements required for satisfying these goals. As noticed in footnote 6, of course, not all visual perception is aimed to action.

2.3 Words and representations in the brain

It has been pointed out (Rizzolatti *et al*. 1988) that F5 represents a *vocabulary*, in which motor acts are inscribed, each "*word*" being constituted by a group of neurons, each of which refers to, or represents, one kind of *motor act* (rather than simple *movements*).

The referent of these "words" can be of different generality: for instance, the general goal of the action (e.g., in goal related neurons: grasping, holding, tearing), or the particular way in which it must be executed (e.g., in visuomotor neurons: *precision grip*, by the index finger and the thumb) (Gallese and Metzinger 2003: 367). So, these "words" work as general terms, referring to *kinds*, rather than particulars (Caruana and Borghi 2013): as far as F5 is concerned, the cortex ignores the single movements whose sequence makes up the act (like, e.g. flexing the thumb, extending the index, etc.): the motor act is represented as a whole, and directly selected by activating the related neural population (Gallese and Metzinger 2003: 367-68). There is a clear computational advantage, here, due to the "motor reinforcement" mentioned in § 2.2, which during ontogenetic development selects representations of the more effective motor acts.

Grasping acts are executed under visual guidance: there is a relationship between the 3D visual features of objects and the specific "words" of the vocabulary. For instance, seeing an object and wanting to grasp it evokes a command to grasp by a specifically suitable finger configuration. Thanks to the motor "vocabulary" the appearance of the graspable object in the visual space will immediately retrieve the description codifying the appropriate motor act. So the classification of the objects as to their visual aspects corresponds to the classification of the acts we can perform upon them recorded in the motor vocabulary: the cortical integration of vision and action generates and controls goal-related behaviours by producing internal copies of actions.

Thus subpersonal representations are at the basis of our motor skills to handle sensorimotor contingencies. Our sensorimotor behaviour does not result from the operation of two separate modules (vision and tact) which interact only at the cognitive level: they are already integrated at the level of cortical representations. So, subpersonal representations are fundamental for the intentional stance that characterizes our relation with the external objects.

As mentioned in § 1, according to Noë "We ought to reject the idea that the perceptual system constructs an internal representation of the world" (Noë 2004: 2). On the other hand, "No doubt perception depends on what takes place in the brain, and very likely there are internal representations in the brain" (ibi).

These quotes might seem contradictory, but they would be consistent if Noë were talking of representations in two different senses. Noë does not explain which kind of representations he is thinking about[29], but at least two reasonable hypotheses can be advanced. First, that Noë is concerned with rejecting "propositional" representations, in the sense in which cognitivists like Fodor have claimed that basic cognition relies on propositional attitudes; in other words, what he rejects are propositional representations which consist in a sort

[29] Noë and Thompson (2004: 4) focus on the question about the neural correlates of consciousness (NCC). However, they mainly discuss whether or not "neural states that have been shown experimentally to be correlated with conscious visual experiences match those experiences in content".

of amodal, disembodied *mental gymnastic* (Fodor 1975, 1978; for a review see Hutto 2005, Hutto and Myin 2013; Menary 2006; Noë 2002, 2004).[30]

The second hypothesis is that, as mentioned in § 1, he rejects only "snapshot representational pictures", that is, *pictorial* representations which "represent the whole scene in consciousness all at once in the way a photograph does", such as those involved in Marr's account of vision (Marr 1982).

But the representations we are concerned with are of neither kind: first, our SV representations are a particular form of representational mechanism arising from visuomotor transformation processes occurring in AIP-F5. So, it might be pointed out that since SV representations have a motor format,[31] they are non-conceptual and non-propositional (Evans 1982; Heck 2000; Peacocke 2001; McDowell 1994), contrary to classical representations (Fodor 1983, 1989, Fodor and Pylyshyn 1988). In fact, the "computation" by which their motor format is realized does not process propositions, or anything that can be true or false (like, e.g., the claim that an object exists), but just parameters: some particular spatial features translated as commands for a motor act.[32] For instance, it is the computation of the particular way in which an act can be executed with respect to, say, a small, tapered, thin object; that is, in this case, by a precision grip. As they put it in the literature, what is computed is the suitability of a potential motor act to operate upon what we find in the external world (Butterfill and Sinigaglia 2012; Jacob and Jeannerod 2003; Jeannerod 2006). The process of motor representation allows to functionally overcome the distinction between perception and action: in fact, through SV representations and the visuomotor transformation process by which they are subserved, the brain represents how the perceptual features of objects must be read as contents of a motor nature (precisely, of a sensorimotor nature: e.g., action goal, a precise grasp to perform). Indeed, the output of the motor processing of visual stimuli is 'motorically' encapsulated (Jacob and Jeannerod 2003: 177).[33] Hence, the motor representations can be defined as an internal state of the subject in which perception and action are not precisely delimited (Jeannerod 2006). Obviously, the vehicle of these representations is, in general terms, the entire pathway that goes from the primary visual cortex (V1), through the dorsal visual stream ending to the (posterior) parietal lobe, hence to the AIP-F5 circuit: i.e., the complex of the cortical sites described in this section.[34]

In particular, these representations are not inferential or propositional representations, such as those involved in the *language of thought hypothesis* (or *mentalese*) (Fodor 1975, 1978), nor they are amodal or disembodied, as they are developed through motor reinforcement (see § 2.2). Moreover, our representations are not "snapshot representational pictures", since they do not represent "the whole scene in consciousness all at once in the way a photograph does". On the

[30] We owe this suggestion to an anonymous referee.
[31] See § 2.1, § 2.2.
[32] Indeed, both hand preshaping and SV representations are automatic processes. In fact, the dorsal stream of visual processing grounding the perception of affordances is almost totally independent from conscious phenomena. For further technical details, see Jeannerod 2006; Jacob and Jeannerod 2003. For a complete account of how the AIP-F5 circuit constructs perceptual representations which are not structured in a propositional format, see Ferretti, forthcoming-a.
[33] See § 2.1, § 2.2, § 2.3 above.
[34] Further technical details have already been given at § 2.1, § 2.2.

contrary, since they are constituted by "words", they codify only some selected features of objects, like 3D structure, and some corresponding features of actions, like goals or types of prehension.

Therefore, there need be no contradiction between Noë's rejection of "propositional" and/or "snapshot" representations, and our claim that there are non-conceptual, non-propositional, SV representations at the cortical level.

3. Is perception direct in the light of UAP and Affordances?

As mentioned in the beginning, on the basis of UAP and the relative use of affordances (claim E1) enactivists claim that perception is direct (E2). In order to assess the latter claim, however, some conceptual clarification is needed.

There are at least 3 things one could mean by the claim that perception is direct:

1) That no representation at all is constructed or involved in perception, whether conscious or unconscious, at the personal or subpersonal level, pictorial or propositional.
2) That there is no "snapshot", no pictorial representation of the whole scene in all its details all at once as a photograph, at least at the personal level (as claimed by Noë), or even at the subpersonal level.
3) That perception *appears* direct, i.e. we have no conscience of any representation mediating between the object and our perception of it.

These claims are in order of decreasing strength: (3) might be true even if (2) is false, and (2) might be true even if (1) is false. In fact, there is no question that (3) is true, and we are ready to grant (2) as well. But the evidence described in § 2 abundantly shows that (1) is false: first-person experience stands on the basis of the SV representations, thanks to which we can be aware of objects and "catch" them (both cognitively and physically). Moreover, in spite of the truth of (2) and (3), any simple claim to the effect that "perception is direct" without further qualifications would be seriously misleading, for it would risk to conflate (2) and (3) with (1). Of course one may decide to understand 'direct' in sense (2) and/or (3), as apparently the enactivists tend to do; but this terminology would not be very useful; in fact, it might be misleading, given the crucial role played by neural representations: it is thank to them that objects *appear* as directly given at the phenomenal level.

As mentioned, enactivists hold that E1 (the doctrine of affordances and UAP, action-perception union) entails E2 (that perception is direct). In so doing, they cannot mean 'direct' but in senses (2) or (3). On the other hand, E1 could hardly be credible if we didn't understand how it is implemented at the neural level. But the empirical findings discussed here precisely show how this implementation occurs and what the neural correlates of UAP and affordances are. In particular, we have seen that the integration of perception and action does not happen at the cognitive level, but already at the cortical level. Just for this reason, however, perception is *not* direct in sense (1). So, experimental evidence shows that claim, E1, hence also E2, are not compatible with (1): in a very important sense, direct perception is *not* possible in the light of UAP and affordances.

Conceptual confusion might also derive from the claim that

(4) perception is transparent,[35]

and the ensuing argument that

(5) perception is direct *because* it is transparent.

In fact, (4) suffers the same kind of ambiguity as the claim that perception is direct: on the one hand, (4) might mean that we cannot perceive any representation between us and the object,[36] i.e., that perception *appears* direct; in this case (4) would be synonymous of (3), and (5) would be valid, but tautological.

On the other hand, we can properly call "*trans*-parent" only what is positioned between a subject and an object, but through which one can see, like a glass, a lens, the atmosphere, etc. One of these mediums is called transparent when (i) it cannot be noticed, and/or (ii) it does not prevent or significantly distort our perception of the object. In this particular case, we are not talking of a physical medium standing between our body and the external object, but of the neural visuomotor representations in F5 which mediate between the physical inputs from the object and our perception at personal level. Now, these representations are transparent precisely in both senses, since (i) the subject is unaware of them, for they do not surface at the personal level,[37] and (ii) they do not prevent or distort perception (in fact, they constitute its physical realization).

So, the most proper way of phrasing (4) would be saying that perception happens through transparent hidden representations. And this of course would entail both (3) and the negation of (1). In other words, perception involves subpersonal representations, so it is indirect in a relevant (perhaps the *most* relevant) sense of the term.

On the other hand, as explained by Metzinger, perception is no longer transparent in the case of hallucinations: "if an hallucination is occurring, I am no longer looking "through" (in the sense of thanks to) the state in my head onto the world, but only at the representational vehicle itself—without realizing *this* fact" (Metzinger 2003: 173).*

* We thank Alfredo Paternoster and two anonymous referees for many useful suggestions.

References

Baumann, M.A., Fluet, M.C. and Scherberger, H. 2009, "Context-Specific Grasp Movement Representation in the Macaque Anterior Intraparietal Area", *The Journal of Neuroscience*, 29, 6436-6448.

Bonini, F., Burle, B., Liégeois-Chauvel, C., Régis, J., Chauvel, P. and Vidal, F. 2014, "Action Monitoring and Medial Frontal Cortex: Leading Role of Supplementary Motor Area", *Science*, 343, 888.

Borghi, A.M. and Riggio, L. 2015, "Stable and Variable Affordances are both Au-

[35] E.g.: "The transparency, or diaphanousness, of experience poses a problem for the theory of perception. To describe sensory experience, to reflect on it, is to turn one's attention to the experienced world. The experience itself is transparent" (Noë 2004).

[36] E.g., "There's only encountering the world content as you experience it. It would seem, then, that we cannot reflect on experience itself" (Noë 2004).

[37] Gallese and Metzinger 2003: 370.

tomatic and Flexible", *Frontiers in Human Neuroscience*, 9, 351, DOI: 10.3389/fnhum.2015.00351. eCollection 2015.

Brooks, R. 1991, "Intelligence without Representation", *Artificial Intelligence*, 47, 139-59.

Brooks, R. 1999, *Cambrian Intelligence*, Cambridge, MA: MIT Press.

Brooks, A. and Akins, K. 2005, *Cognition and the Brain. The Philosophy and Neuroscience Movement*, Cambridge: Cambridge University Press.

Brochier, T., Spinks, R.L., Umilta, M.A. and Lemon, R.N. 2004, "Patterns of Muscle Activity Underlying Object-Specific Grasp by the Macaque Monkey", *Journal of Neurophysiology*, 92, 1770-1782.

Brochier, T. and Umilta, M.A. 2007, "Cortical Control of Grasp in Non-Human Primates", *Current opinions in Neurobiology*, 17, 637- 43.

Butterfill, S.A. and Sinigaglia, C. 2012, "Intention and Motor Representation in Purposive Action", 88, 1, 119-45.

Caruana, F. and Borghi, A.M. 2013, "Embodied Cognition: Una nuova psicologia", *Giornale Italiano di Psicologia*, 1, 23-48, DOI: 10.1421/73973. ISSN: 0390-5349.

Castiello, U. 2005, "The Neuroscience of Grasping", *Nature Reviews*, 6(9), 726-36, DOI: 10.1038/nrn1744.

Chemero, A. 2000, "Anti-Representationalism and the Dynamical Stance", *Philosophy of Science*, 67, 625-47.

Chemero, A. 2009, *Radical Embodied Cognitive Science*, Cambridge, MA: MIT Press.

Chinellato, E. and del Pobil, A.P. 2015, *The Visual Neuroscience of Robotic Grasping. Achieving Sensorimotor Skills through Dorsal-Ventral Stream Integration*, Cham: Springer.

Clark, S., Tremblay, F. and Ste-Marie, D. 2004, "Differential modulation of corticospinal excitability during observation, mental imagery and imitation of hand actions", *Neuropsychologia*, 42, 993-96.

Clark, A. 1998, *Being There: Putting Mind, Body, and World Together Again*, Cambridge, MA: MIT Press.

Clark, A. and Josefa Toribio, J. 1994, "Doing without Representing", *Synthese* 101 (3), 401-31.

Clark, A. 2013, "Whatever next? Predictive Brains, Situated Agents, and the Future of Cognitive Science", *Behavioral and Brain Sciences*, 36, 181-253.

Craighero, L. 2010, *Neuroni Specchio*, Bologna: Il Mulino.

Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., *et al.* 1994, "Mapping Motor Representations with PET", *Nature*, 371, 600-602.

Decety, J. and Ingvar, D.H. 1990, "Brain Structures Participating in Mental Simulation of Motor Behavior: a Neuropsychological Interpretation", *Acta Psychologica*, 73, 13-24.

Dennett, D. 1987, *The Intentional Stance*, Cambridge, MA: MIT Press.

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

Ferretti, G., forthcoming-a, "Neurophysiological States and Perceptual Representations: The Case of Action Properties detected by the Ventro-Dorsal Stream", in L. Magnani and C. Casadio (eds.), *Model-Based Reasoning in Science and Technology. Models and Inferences: Logical, Epistemological, and Cognitive Issues,* Cham: Springer.

Ferretti, G., forthcoming-b, "Enattivismo Sensomotorio" [Sensorimotor Enactivism], *Aphex. Giornale Italiano di Filosofia Analitica.*

Fluet, M.C., Baumann, M.A. and Scherberger, H. 2010, "Context-Specific Grasp Movement Representation in Macaque Ventral Premotor Cortex", *The Journal of Neuroscience*, 30, 15175-15184.

Fodor, J.A. 1975, *The Language of Thought*, Cambridge, MA: Harvard University Press.

Fodor, J.A. 1978, "Propositional Attitudes", in Fodor, J.A., *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, Cambridge, MA: MIT Press, 1981. (Originally in *The Monist* 64, 1978.)

Fodor, J.A. 1983, *The Modularity of Mind. An Essay on Faculty Psychology.* Cambridge, MA: MIT Press.

Fodor, J.A. 1989, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: Bradford Books.

Fodor, J.A. and Pylyshyn Z. 1988, "Connectionism and Cognitive Architecture: A Critical Analysis", *Cognition*, 28, 3-71.

Fogassi, L. and Luppino, G. 2005, "Motor Functions of the Parietal Lobe", *Current Opinion in Neurobiology*, 15, 626-31.

Fridland, E. 2012, "Perception, Action, and Consciousness: Sensorimotor Dynamics and the Two Visual Systems" (review of Gangopadhyay *et al.* 2010), *Phenomenology and the Cognitive Sciences*, 12 (4), 899-906.

Friston, K. 2009, "The Free-Energy Principle: A Rough Guide to the Brain?", *Trends in Cognitive Sciences*, 13 (7), 293-301.

Gallagher, S. 2008, "Are Minimal Representations Still Representations?", *International Journal of Philosophical Studies*, 16 (3), 351-69.

Gallese, V. 2000, "The Inner Sense of Action. Agency and Motor Representations", *Journal of Consciousness Studies*, 7 (10), 23-40.

Gallese, V. 2003, "A Neuroscientific Grasp of Concepts. From Control to Representations", *The Royal Society*, 358, 1231-1240.

Gallese, V., Fadiga, L., Fogassi, L. and Rizzolatti, G. 1996, "Action Recognition in the Premotor Cortex", *Brain*, 119, 593-609.

Gallese, V. and Keysers, C. 2001, "Mirror Neurons: A Sensorimotor Representation System" (Comments to O'Regan and Noë 2001), *Behavioral and the Brain Science*, 24, 983.

Gallese, V. and Metzinger, T. 2003, "Motor Ontology. The Representational Reality of Goals, Actions and Selves", *Philosophical Psychology*, 16 (3), 365-88.

Gallivan, J. P. and Wood, D.K. 2009, "Simultaneous Encoding of Potential Grasping Movements in Macaque Anterior Intraparietal Area" (review of Bauman *et al.* 2009), *The Journal of Neuroscience*, 29 (39), 12031-12032.

Gangopadhyay, N. and Kiverstein, J. 2009, "Enactivism and the Unity of Perception and Action", *Topoi*, 28 (1), 63-73.

Gangopadhyay, N., Madary, M. and Spencer, F. 2010 (eds.), *Perception, Action, and Consciousness: Sensorimotor Dynamics and the Two Visual Systems*, Oxford: Oxford University Press.

Gauthier, I., Rizzolatti, G., Fadiga, L., Gallese, V. and Fogassi, L. 1996, "Premotor Cortex and the Recognition of Motor Actions". *Cognitive Brain Research*, 3, 131-41.

Gazzaniga, M.S. (ed.), *The Cognitive Neurosciences*. Cambridge, MA: MIT Press.

Gibson, J.J. 1986, *The Ecological Approach to Visual Perception*, New York: Psychology Press.

Goodale, M. A., Jakobson, L.S., Milner, D.A., Perrett, D.I., Benson, P.J and Hietanen, J.K. 1994, "The Nature and Limits of Orientation and Pattern Processing Supporting Visuomotor Control in a Visual form Agnostic", *Journal of Cognitive Neuroscience*, 6, 46-56.

Grafton, S. T., Arbib, M.A., Fadiga, L. and Rizzolatti, G. 1996, "Localization of Grasp Representations in Humans by Positron Emission Tomography. 2. Observation Compared with Imagination", *Experimental Brain Research*, 112, 103-11.

Grush, R. 2004, "The Emulation Theory of Representation: Motor Control, Imagery, and Perception", *Behavioral and Brain Sciences*, 27, 377-442.

Heck, R.G. 2000, "Nonconceptual Content and the Space of Reasons", *Philosophical Review*, 109, 483-523.

Hommel, B., Müsseler, J., Aschersleben, G. and Prinz, W. 2001, "The Theory of Event Coding: a Framework for Perception and Action Planning", *Behavioral and Brain Sciences*, 24, 849-937.

Hutto, D.D. 2005, "Knowing What? Radical versus Conservative Enactivism", *Phenomenology and the Cognitive Sciences*, 4 (4), 389-405.

Hutto, D.D. and Myin, E. 2013, *Radicalizing Enactivism: Basic Minds without Content*, Cambridge, MA: MIT Press.

Ingle, D.J., Goodale, M.A. and Mansfield, R. J. W. (ed.). *Analysis of Visual Behavior*, Cambridge, MA: MIT Press.

Jacob, P. 2005, "Grasping and Perceiving Objects", in Brooks, A. and Akins, K. 2005, *Cognition and the Brain. The Philosophy and Neuroscience Movement*, Cambridge: Cambridge University Press.

Jacob, P. and Jeannerod, M. 2003, *Ways of Seeing. The Scope and Limits of Visual Cognition*, Oxford: Oxford University Press.

Jeannerod, M. 1986, "The Formation of Finger Grip During Prehension. A Cortically Mediated Visuomotor Pattern", *Behavioural Brain Research*, 19, 99-116.

Jeannerod, M. 1994, "The Representing Brain. Neural Correlates of Motor Intention and Imagery", *Behavioral and Brain Sciences*, 17, 187-245.

Jeannerod, M. 1997, *The Cognitive Neuroscience of Action*, Oxford: Blackwell.

Jeannerod, M. 2001, "Neural Simulation of Action: a Unifying Mechanism for Motor Cognition", *Neuroimage*, 14, 103-109.

Jeannerod, M. 2006, *Motor Cognition: What Actions Tell the Self*. Oxford: Oxford University Press.

Johnson, S.H., Rotte, M., Gafton, S.T., Hinrichs, H., Gazzaniga, M.S. and Heinze, H.J. 2002, "Selective Activation of a Parietofrontal Circuit During Implicitly Imagined Prehension", *Neuroimage*, 17, 1693-704.

Kawato, M. 1997, "Bidirectional Theory Approach to Consciousness", in Ito, M., Miyashita Y. and Rolls E.T. (eds.), *Cognition, Computation and Consciousness*, Oxford: Oxford University Press.

Kawato, M. 1999, "Internal Models for Motor Control and Trajectory Planning", *Neuroreport*, 9, 718-27.

Kosslyn, S.M. 1996, *Image and Brain. The Resolution of the Imagery Debate*, Cambridge, MA: MIT Press.

Kosslyn, S.M. 2005, "Mental Images and the Brain", *Cognitive Neuropsychology*, 22 (3-4), 333-47.

Kukla, R. 1992, "Cognitive Models and Representation", *British Journal of Philosophy of Science*, 43, 219-32.

Kurata, K., and Tanji J. 1986, "Premotor Cortex Neurons in Macaques. Activity before Distal and Proximal Forelimb Movements", *Journal of Neuroscience*, 6, 403-11.

Leopold, D.A. and Logothetis, N.K. 1999, "Multistable Phenomena: Changing Views in Perception", *Trends in Cognitive Sciences*, 3, 254-64.

Marr, D. 1982, *Vision*, San Francisco: Freeman.

Matelli, M., Luppino, G. and Rizzolatti, G. 1985, "Patterns of Cytochrome Oxidase Activity in the Frontal Agranular Cortex of the Macaque Monkey", *Behavioral Brain Research*, 18, 125-37.

McDowell, J. 1994, "The Content of Perceptual Experience", *Philosophical Quarterly*, 44, 190-205.

Menary, R. 2006, *Radical Enactivism. Intentionality, Phenomenology and Narrative Focus on the Philosophy of Daniel D. Hutto*, Amsterdam-Philadelphia: John Benjamins.

Metzinger, T. 2003, *Being No-One: the Self-Model Theory of Subjectivity*, Cambridge, MA: MIT Press.

Milner, D.A. and Goodale M.A. 1995, *The Visual Brain in Action*, Oxford: Oxford University Press.

Milner, D.A., Paulignan, Y., Dijkerman, H.C., Michel, F. and Jeannerod, M. 1999, "A Paradoxical Improvement of Misreaching in Optic Ataxia: New Evidence for Two Separate Neural Systems for Visual Localization, *Proceedings of the Royal Society*, 266, 2225-2229.

Milner, D.A., and Goodale, M.A. 2004, *Sight the Unseen. An Exploration of Conscious and Unconscious Vision*, Oxford: Oxford University Press.

Mountcastle, V.B., Lynch, J.C., Georgopulos, A., Sakata, H. and Akuna, C. 1975, "Posterior Parietal Association Cortex of Monkey. Command Functions for Operations within Extrapersonal Space", *Journal of Neurophysiology*, 38, 871-908.

Mühlau, M., Hemsdörfer, J., Goldenberg, G., Wohlschläger, A.M., Castrop, F., Stahl, R. *et al*. 2005, "Left Inferior Parietal Dominance in Gesture Imitation: An fMRI Study", *Neuropsychologia*, 43, 1086-1098.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V. and Rizzolatti, G. 1997, "Object Representation in the Ventral Premotor Cortex (area F5) of the Monkey". *Journal of Neurophysiology*, 78, 2226-2230.

Murata, A., Gallese, V., Luppino, G., Kaseda, M., and Sakata, H. 2000, "Selectivity for the Shape, Size and Orientation of Objects for Grasping in Neurons of Monkey Parietal Area AIP", *Journal of Neurophisiology*, 79, 2580-2601.

Noë, A. 2004, *Action in Perception*, Cambridge, MA: MIT Press.

Noë, A. 2002, "Is the Visual World a Grand Illusion?", Thorverton (Engl): *Imprint academic*, 9, 1-12.

Noë, A. 2009, *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*, New York: Hill & Wang.

Noë, A. and Thompson, E. 2002, *Vision and Mind. Selected Readings in the Philosophy of Perception*, Cambridge, MA: MIT Press.

Noë, A. and Thompson, E. 2004, "Are There Neural Correlates of Consciousness?", *Journal of Consciousness Studies*, 11 (1), 3-28.

O'Regan, K.J. and Noë, A. 2001, "A Sensorimotor Account of Vision and Visual Consciousness", *Behavioral and the Brain Sciences*, 24, 983-84.

O'Regan, K. J. 2011, *Why Red Doesn't Sound Like a Bell: Explaining the Feel of Consciousness*, Oxford: Oxford University Press.

Peacocke, C. 2001, "Does Perception have a Nonconceptual Content?", *Journal of Philosophy*, 98, 239-64.

Perenin, M.T. and Vighetto, A. 1988, "Optic Ataxia: A Specific Disruption in Visuomotor Mechanisms. I. Different Aspects of the Deficit in Reaching for Objects", *Brain*, 111, 643-74.

Raos, V., Umilta, M.A., Murata, A., Fogassi, L. and Gallese, V. 2006, "Functional Properties of Grasping-Related Neurons in the Ventral Premotor Area F5 of the Macaque Monkey", *Journal of Neurophysiology*, 95, 709-29.

Rizzolatti, G. and Sinigaglia, C. 2006, *So quel che fai. Il Cervello che agisce e i Neuroni Specchio*, Milano: Cortina.

Rizzolatti, G., Luppino, G. and Matelli, M. 1998, "The Organization of the Cortical Motor System: New Concepts", *Electroencephalography and Clinical Neurophisiology*, 106.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G. and Matelli, M. 1988, "Functional Organization of Inferior Area 6 in the Macaque Monkey. II. Area F5 and the Control of Distal Movements", *Experimental Brain Research*, 71, 491-507.

Rizzolatti, G. and Fadiga, L. 1998, "Grasping Objects and Grasping Action Meanings: the Dual Role of Monkey Rostroventral Premotor Cortex (area F5)", *Novartis Foundation Symposia*, 218, 81-95.

Rizzolatti, G., Fadiga, L., Gallese, V. and Fogassi, L. 1996, "Premotor Cortex and the Recognition of Motor Actions", *Cognitive Brain Research*, 3, 131-41.

Rizzolatti, G., Fogassi, L. and Gallese, V. 1999, "Cortical Mechanisms Subserving Object Grasping. A New View on the Cortical Motor Functions", in Gazzaniga, M.S. (ed.), *The Cognitive Neurosciences*, Cambridge, MA: MIT Press, 539-52.

Rizzolatti, G., Sandolara, C., Matelli, M. and Gentilucci, M. 1981, "Afferent Properties of Periarcuate Neurons in Macaque Monkeys. II. Visual Responses", *Behavioral Brain Research*, 2, 147-63.

Rumiati R. I., Weiss, P.H., Shallice, T., Ottoboni, G., Noth, J., Zilles, K. *et al.* 2004, "Neural Basis of Pantomiming the Use of Visually Presented Objects", *Neuroimage*, 21, 1224-1231.

Sakata, H., Taira, M., Murata, A. and Mine, S. 1995, "Neural Mechanisms of Visual Guidance of Hand Action in the Parietal Cortex of the Monkey", *Cerebral Cortex*, 5, 429-38.

Shikata, E., Hamzei, F., Glauche, V., Koch, M., Weiller, C., Binkofski, F. and Büchel, C. (2003), "Functional Properties and Interaction of the Anterior and Posterior Intraparietal Areas in Humans", *European Journal of Neuroscience*, 17 (5), 1105–1110.

Stewart, J., Gapenne, O. and Di Paolo, E.A. 2010, *Enaction. Toward a New Paradigm for Cognitive Science*, Cambridge (MA): MIT Press.

Thompson, E. 2007, *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*, Cambridge, MA: Harvard University Press.

Townsend, B.R., Subasi, E. and Scherberger, H. 2011, "Grasp Movement Decoding from Premotor and Parietal Cortex", *The Journal of Neuroscience*, 31(40), 14386-14398.

Turella, L. and Lignau, A. 2014, "Neural Correlates of Grasping", *Frontiers in Human Neuroscience*, 8 (686), DOI:10.3389/fnhum.2014. 00686.

Ungerleider, L. and Mishkin, M. 1982, "Two Cortical Visual Systems", in Ingle, D. J., Goodale, M.A. and Mansfield, R.J.W. (eds.), *Analysis of Visual Behavior*, Cambridge: MIT Press, 549-86.

Van Gelder, T. 1995, "What Might Cognition be if not Computation?", *Journal of Philosophy*, 91, 345-81.

Van Gelder, T. 1998, "The Dynamical Systems Hypothesis", *Behavioral and Brain Sciences*, 21, 1-14.

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

Wilson, R.A. 2004, *Boundaries of The Mind: The Individual in the Fragile Sciences*. Cambridge: Cambridge University Press.

Wolpert, D.M. 1997, "Computational Approaches to Motor Control", *Trends in Cognitive Sciences*, 1, 209-16.

Wolpert, D.M., Ghahramani, Z.M.I. and Jordan, M.I. 1995, "An Internal Model for Sensorimotor Integration", *Science*, 269, 1880-1882.

Zipoli Caiani, S. 2013a, "Extending the Notion of Affordance", *Phenomenology and the Cognitive Sciences*, 13, pp. 275-93, DOI: 10.1007/s11097-013-9295-1.

Zipoli Caiani, S. 2013b. "Cognizione incorporata", *Aphex. Giornale Italiano di Filosofia Analitica*, www.aphex.it/index.php?Temi=557D03012202740321070205777327.

Naïve Realism and the Explanatory Role of Visual Phenomenology

Takuya Niikawa

Hokkaido University

Abstract

This paper argues that naïve realism has an epistemic advantage over other rival views. The argument consists of two steps. First, I argue that the phenomenology of veridical visual experience plays an indispensable role in explaining how we can refer to the experience as a justificatory reason for a demonstrative judgment. Second, I argue that only naïve realism can coherently allow a veridical visual experience to be used as a factive reason.

Keywords: Naïve Realism, Factive Reason, Perceptual Experience, Visual Phenomenology, Introspection.

1. Naïve Realism

Naïve realism can be characterized as a conjunction of two claims, one explanatory and one metaphysical. The explanatory claim is that the phenomenology of veridical visual experience is *explained by* acquaintance (or perception), an irreducible mental relation between a subject and environmental objects. That is to say, a veridical experience has visual phenomenology in virtue of *its acquainting the subject with environmental objects (or the subject's perceiving environmental objects)*, rather than *its representing such objects* or *its acquainting the subject with private mental entities*.[1] I call the property *of acquainting the subject with environmental objects* the 'naïve realist property'.[2] This explanatory claim is insufficient by itself in that it does not address the metaphysics of visual phenomenology. Given this, naïve realists, such as Campbell and Logue, add the metaphysical claim that the visual phenomenology of veridical experience is wholly constituted by environ-

[1] According to this characterization, the following figures can be regarded as naïve realists: Bill Brewer (2011), John Campbell (2002a), Mike Martin (2004, 2006), Mark Johnston (2004, 2006), William Fish (2009), Matthew Kennedy (2013) and Logue (2012a, 2012b).

[2] Throughout this paper, I use "see" and "perceive" in the factive sense that when we see or perceive X, then X exists.

Argumenta 1, 2 (2016): 219-231
ISSN 2465-2334

© 2016 University of Sassari

DOI 10.14275/2465-2334/20162.NII

mental objects and their properties, with which the subject is acquainted.[3] Hence, I characterize naïve realism as the conjunction of these explanatory and metaphysical claims.

Given that the phenomenology of experience is usually regarded as *a property* of experience, one may wonder if it is a category-mistake to think that the phenomenology of veridical experience is constituted by environmental objects and their properties. There is, however, a conception of visual phenomenology that does not presuppose that it is a property of experience. William Fish introduced a notion of presentational character, which is "the array of features that we are presented with [in having an experience] and that characterize what it is like to have that experience".[4] In his characterization, it is left open what metaphysical status the presentational character has. That is, the features in question can be of environmental object, private mental entity, intentional object or experience itself. I characterize the phenomenology of visual experience in terms of its presentational character. Hence, this characterization opens the possibility that the phenomenology of veridical experience is constituted by environmental objects and their properties.

A large number of studies have been made on the viability of naïve realism, but as Logue pointed out, little attention has been given to the motivations of naïve realism.[5] Why should we adopt naïve realism? Among others, one popular idea is that naïve realism best captures the explanatory roles of the phenomenology of veridical experience. The aim of this paper is to develop and defend this idea.

2. The Explanatory Role of Visual Phenomenology

The first question that needs to be asked is, what explanatory roles does the visual phenomenology of veridical experience have? The following is a list of potential explanatory roles.

(1) Demonstrative Thought: the phenomenology of veridical experience plays a role in explaining how it is possible to demonstratively think about an environmental object.
(2) Concept of Perceptible Properties: the phenomenology of veridical experience plays a role in explaining how we acquire certain concepts of perceptible properties such as redness and roundness.
(3) Knowledge about the Nature of Perceptible Properties: the phenomenology of veridical experience plays a role in explaining how we acquire knowledge about the nature of perceptible properties such as redness and roundness.
(4) Knowledge of our Surrounding Environment: the phenomenology of veridical experience plays a role in explaining how we acquire knowledge of our surrounding environment such as the knowledge that there is a bottle of whiskey in front of me.

Naïve realists maintain that naïve realism best captures all or some of these explanatory roles. The opponents of naïve realism can object to this claim in two different ways. One way is to argue that a certain theory other than naïve

[3] Campbell 2002a; Logue 2012b.
[4] Fish 2009: 15.
[5] Logue 2012b.

realism can also sufficiently capture the explanatory roles. The other way is to argue that the visual phenomenology of veridical experience does not play any such explanatory roles. Naïve realists provide detailed arguments against the first type of objection.[6] In this paper, I do not discuss this type of objection. What I focus on here is the second type of objection. Indeed, naïve realists tend to simply assume, without sufficient arguments, that the visual phenomenology of veridical experience plays such explanatory roles. For instance, Heather Logue and Thomas Raleigh respectively claim:

> Why should we think that the phenomenal character of veridical experience ever puts us in [a position to know something of what things are like independently of experience]? My response is to turn the tables—why shouldn't we think this?[7]

> I assume that a subject's conscious experience of O can play a role in allowing the subject to demonstratively refer to O.[8]

Such assumptions seem plausible from the first-person perspective. When reflecting on my cognitive life, it seems to be the case that I actually employ the visual phenomenology of veridical experience in order to form demonstrative thoughts and acquire various kinds of knowledge and certain concepts. First, consider a case in which I see a red apple and then form a demonstrative judgment that this is red (and further suppose that the judgment counts as knowledge). It seems to me that what I would do in that case is, roughly speaking, to pay attention to the red-apple-phenomenology of the experience and then apply the demonstrative concept *this* and the colour concept *red* to the phenomenology. Second, consider a case in which I am seeing a red apple but it is my first time seeing a red object. It seems to me that I would acquire the concept of *red* by attending to the red-apple-phenomenology if I am in an appropriate context. Likewise, it seems to me that I would come to know an essential aspect of redness by attending to the red-apple-phenomenology if I am in an appropriate context. If these considerations are correct, the phenomenology of veridical experience plays an important role in making demonstrative judgments and acquiring relevant knowledge and concepts. It is this first-person reflection that makes it apparently reasonable to assume that visual phenomenology has such explanatory roles.

Nevertheless, the fact that such an assumption seems plausible from the first-person perspective does not mean that it is indeed true. Although the first-person reflection may give a positive evidence for such an assumption, it is not decisive. To deny such an assumption is to endorse the view that the phenomenology of veridical experience is explanatorily impotent. I call it the "impotence view".

In order to justify the impotence view, its advocates may claim that introspection is not a reliable tool to know the nature of our cognitive activities. It is standard in cognitive science to think that "[common sense and introspection]

[6] For the first and second explanatory roles, see Campbell 2002a, 2002b, 2011. Thomas Raleigh (2011) also provides a plausible argument implicitly suggesting that naïve realism alone can capture the first explanatory role. For the third explanatory role, see Logue 2012b. For the fourth one, see Johnston 2006, 2011.

[7] Logue 2012b: 231.

[8] Raleigh 2011: 174.

can give a misleading picture of mental operations".[9] This idea is supported by the fact that we are surprised by various mental phenomena such as change blindness or inattentional blindness.[10] Taking this into account, it seems inappropriate to swallow everything that first-person reflection tells us as to how we perform cognitive activities on faith. This consideration in itself does not provide a positive evidence for the impotence view, but it casts a doubt on the first-person reflection suggesting that we actually utilize the phenomenology of veridical experience to perform various cognitive activities. Unless there is some additional support for the reliability of first-person reflection in the given cases, advocates of impotence view claim, the suggestion is not justified. If we should not trust the first-person reflection, then there seems to be no negative evidence for the impotence view.

Perhaps, advocates of the impotence view may further claim that there is a positive evidence for it. One strategy is to appeal to the undoubted fact that cognitive psychology as a scientific approach to human mind has largely succeeded.[11] It is generally accepted that cognitive psychology has successfully provided elaborate accounts of various mental phenomena such as memory, object-recognition and attention. Given this, it seems reasonable to posit a working hypothesis that cognitive psychology can, in principle, provide plausible accounts of relevant cognitive and epistemic activities such as forming demonstrative judgments and acquiring various types of concepts and knowledge. First, it seems possible to explain how we form a demonstrative thought in terms of perceptual-recognition systems and its causal connection to an environmental object. Second, in developmental psychology, there are some theories of concept acquisition. Third, cognitive psychological theories as to how we are able to visually recognize and identify a colour may produce a plausible answer to the question of how we can know what a colour is. Here, it is important to note that the phenomenal aspect of perceptual experience would not be mentioned in such scientific explanations. That is to say, it seems plausible to think that we can adequately explain the aforementioned cognitive activities without reference to the phenomenology of perceptual experience. Based on this, advocates of the impotence view may claim that the phenomenology of perceptual experience has no role in explaining such cognitive activities.

However, this argument fails. Even if cognitive psychology can provide an account of cognitive activity without reference to the phenomenology of perceptual experience, it does not mean that the phenomenology has no role in our performing the cognitive activity. The first-person descriptions of how we perform a cognitive activity can be regarded as a sort of account of the cognitive activity from the first-person perspective. Such an account seems to be compatible with any cognitive psychological explanation of the same activity. Suppose that we can provide accounts of digestion from both biological perspective and physiological perspective. Although these accounts are about the same digestive process and different in many respects, they are not incompatible. Rather, it seems that they complement each other. Likewise, the first-person account of a cognitive activity and the cognitive psychological account of it are different, at least,

[9] Thagard 2010: Section 2.
[10] For change blindness and inattentional blindness, see the Simons Lab Website (http://www.simonslab.com/videos.html).
[11] For cognitive psychology in general, see Gazzaniga, Ivry and Mangun 2013.

in the presence or absence of reference to the phenomenology of perceptual experience, but they can be compatible. Unless some argument against this compatibility is provided, it is reasonable to think that the cognitive psychological account of a cognitive activity does not exclude the phenomenological account of it. Rather, they seem to complement each other. Hence, the idea that cognitive psychology can provide an account of the aforementioned cognitive activities does not support the impotence view.

Instead of appealing to cognitive psychology, advocates of the impotence view may appeal to the possibility of the philosophical zombie. Philosophical zombies, who are cognitively the same as us but lack phenomenal consciousness, are regarded as conceivable. If philosophical zombies can perform a cognitive activity, the explanation as to how they perform it must not involve any reference to phenomenology. This suggests that it is unnecessary to mention the phenomenology of experience in order to explain how we, not philosophical zombies but conscious subjects, perform the same cognitive activity. Because philosophical zombies are completely the same as us in all cognitive aspects, the phenomenology of veridical experience would be unnecessary to explain the aforementioned cognitive activities. Therefore, if we accept that the conceivability of philosophical zombie leads to its metaphysical possibility, then we should accept the impotence view.

However, this argument is ineffective. The disagreement between naïve realists and advocates of the impotence view consists in whether the phenomenal aspect of experience is independent of its cognitive aspect. Naïve realists answer in the negative; the advocates of the impotence view answer in the affirmative. Since the possibility of philosophical zombie depends on the idea that the phenomenal and cognitive aspects of experience are independent, naïve realists can object to this argument by contending that it is question-begging to assume that philosophical zombie is metaphysically possible.

These considerations demonstrate that there is neither positive nor negative evidence for the impotence view. On the one hand, Naïve realists draw upon the first-person reflection by which it is shown that the phenomenology of veridical experience plays a role in explaining certain cognitive activities; advocates of impotence view plausibly argue that such a reflection is unreliable. On the other hand, advocates of the impotence view may claim that the general success of cognitive psychology and the conceivability of philosophical zombie support the impotence view; naïve realists can plausibly deny the claim. This dispute seems to lead to the philosophical stalemate that is characterized as follows. Naïve realists and advocates of the impotence view have opposite intuitions. On the one hand, naïve realists have the intuition that first-person reflection must be respected, but the advocates of the impotence view do not have it. On the other hand, advocates of the impotence view have the intuition that the phenomenal aspect of a perceptual experience is metaphysically independent of its cognitive aspect, but naïve realists do not have it. Since their starting points are different, it is impossible to meaningfully develop this discussion. In order to avoid such a philosophical stalemate, naïve realists need to construct an argument against the impotence view without appealing to first-person reflection. In section 3, I will argue, without drawing on first-person reflection, that there is at least one explanatory role of the phenomenology of veridical experience we have a good reason to accept. Then, in section 4, I will discuss whether naïve realism can be promoted by appealing to the explanatory role.

3. Introspection and Phenomenology

It is generally accepted that we can refer to a perceptual experience as a justificatory reason for a demonstrative judgment. Let us consider the case in which I see a red apple in front of me in good daylight and then form the judgment that this apple is red. In this case, I can introspect on the veridical visual experience and then refer to it as a justificatory reason for the judgment. However, it is not the case that all visual states can be accessed via introspection. Note that there are various sub-personal visual states. Consider the case in which I am in a sub-personal visual state of *processing low level visual information*. Such a sub-personal visual state cannot be accessed by introspection.

What makes a difference in introspective accessibility? One plausible idea is that the difference consists in the presence or absence of distinctive phenomenology.[12] A perceptual experience is introspectively accessible because it has distinctive phenomenology. A sub-personal visual state is introspectively inaccessible because it does not have distinctive phenomenology. Nevertheless, it seems implausible to think that if a *mental state* is introspectively accessible, then it has distinct phenomenology. Beliefs may be a counterexample to this conditional statement, since they are usually regarded as introspectively accessible but seemingly do not have distinctive phenomenology.[13] Given this, I restrict the scope of discussion to perceptual state. Thus, my claim is that a *perceptual experience* is introspectively accessible in virtue of having distinctive phenomenology. I call this the "phenomenal introspective principle". It follows from this principle that the phenomenology of veridical experience plays an indispensable role in explaining how we can justify a demonstrative judgment by introspecting on a perceptual experience. I call this the "phenomenal justification view".

The phenomenal introspective principle is supported by first-person reflection. Let us consider a case in which I refer to a veridical visual experience of a red object, via introspection, as a reason for the demonstrative judgment that this apple is red. When reflecting on what I would do in such a case, it seems that I first attend to the red-apple-phenomenology of the experience, and then come to know that the experience can be utilized as a justificatory reason for the judgment. Moreover, in the conceivable case in which the experience is deprived of its phenomenology, I have no idea how I can know, via introspection, that the experience is useful for justifying the judgment. In light of this first-person reflection, the phenomenal introspective principle seems plausible.

As we have seen, however, advocates of the impotence view have objected to the reliability of first-person reflection. If naïve realists cannot produce an additional support for the phenomenal introspective principle, the dialectical situation does not change. In the rest of this section, I construct an argument for the principle without appealing to first-person reflection.

First, I will illuminate the important difference between the cognitive activities directed towards the external world and the ones directed towards perceptual experience. The cognitive activities to form a demonstrative thought and to

[12] A similar idea is proposed by Smithies (2012, 2014). But his argument for the idea is different from mine.

[13] This is controversial. Advocates of cognitive phenomenology may claim that beliefs are introspectively accessible in virtue of being related to distinctive phenomenology (Smithies 2014). For the disputes over cognitive phenomenology, see Bayne and Montague 2011.

acquire knowledge of our surrounding environment are directed towards *the external world*. Likewise, the cognitive abilities to acquire concepts of perceptible properties and knowledge about the nature of perceptible properties are directed towards properties that are instantiated in the *external world*. Even if perceptual experience is involved in the processes of these world-directed cognitive activities, they are not directed towards perceptual experience itself. In contrast, the cognitive activity to introspectively access one's own perceptual experience is obviously directed towards *perceptual experience itself*. It is implausible that the external world consists of the phenomenology of perceptual experience. It is also very doubtful that perceptible properties which can be instantiated in the external world consist of the phenomenology of perceptual experience.[14] Then again, it is plausible that a perceptual experience partly consists of its phenomenology. Although perceptual experience has not only phenomenal aspect but also other various aspects such as psychological and functional aspects, it is standard to think that the distinctive and essential feature of perceptual experience is having visual phenomenology. Hence, we can plausibly assume that our perceptual experience partly consists of its phenomenology. I call it the "phenomenal constitution principle". On one hand, the world-directed cognitive activities are directed toward the entities that do not consist of the phenomenology of perceptual experience. On the other hand, introspection is directed toward a perceptual experience, the components of which includes visual phenomenology. This difference is crucial for my argument.

Second, I shall point out that there is a conceptual connection between phenomenal experience and introspective access to it. It is unintelligible that I, as a lucid and attentive subject, am undergoing a phenomenal experience, but I am not in a position to know what it is like to have the experience.[15] If this were the case, we would have no idea what the concept of *phenomenology* and its cognates refer(s) to. Assuming that we actually have the concept of phenomenology, therefore, it is reasonable to accept the following principle: if one has a perceptual experience, then he/she is in a position to introspectively know about its phenomenology. I call this the "accessible phenomenology principle".[16] In cases where it is unintelligible that X occurs but Y does not occur, it is plausible to think that X conceptually involves Y. Hence, it is plausible to think that the reason why the accessible phenomenology principle is true lies in the conceptual connection between phenomenal experience and introspective access to it. Hence, I plausibly assume that the accessible phenomenology principle is a conceptual truth.

The conjunction of the phenomenal constitution principle and the accessible phenomenology principle leads to the view that we can get access to a perceptual experience via the introspective access to its phenomenology. Assuming that the justificatory aspect and phenomenal aspect of perceptual experience are connected in such a way that the justificatory aspect is made available via the introspective access to the phenomenal aspect, it follows from that view that we

[14] However, it is not implausible that the phenomenology of perceptual experience consists of perceptible properties. The relation *consisting of* is asymmetrical.

[15] Horgan and Kriegel 2007 and Butler 2013 (section 4) pushed the same point.

[16] This does not mean that the introspective access to a phenomenal experience is infallible. The accessible phenomenal principle is different from the principle that if one has a phenomenal experience, then he/she *knows* what it is like to have it.

can refer to a perceptual experience as a justificatory reason via the introspective access to its phenomenology. This entails the phenomenal justification view that the phenomenology of veridical experience plays an indispensable role in explaining how we can justify a demonstrative judgment by introspecting on a perceptual experience.

How to conceive of the connection between the phenomenal and justificatory aspects of perceptual experience depends on each theory of perceptual experience. In the next section, I will give a naïve realist account of this connection. Here, it might be claimed that it is impossible for any view that those aspects are connected in a suitable manner. However, the burden of proof for this extremely strong claim lies with such opponents. To the best of my knowledge, there is no theory-neutral argument showing the impossibility of the justificatory-phenomenal connection. Furthermore, the opponents need to clarify how we can refer to a perceptual experience as a justificatory reason for a demonstrative judgment without appealing to the introspective access to its phenomenology. At the very least, this is a difficult task. Given these considerations, it is reasonable to conclude that the conjunction of the phenomenal constitution principle and the accessible phenomenology principle leads to the phenomenal justification view.

What is important is that this argument does not rely on the first-person reflection on how we refer to a perceptual experience as a justificatory reason. The phenomenal constitution principle is derived from the plausible conception of the metaphysical relation between perceptual experience and its phenomenology; the accessible phenomenology principle is derived from the conceptual connection between phenomenal experience and introspective access to it. Thus, even if first-person reflection were to be generally unreliable, it would have no effect on this argument. It is also important to note that the same line of argumentation cannot be used to show that the phenomenology of perceptual experience plays a role in explaining other *world-directed* cognitive activities, such as forming demonstrative thought and acquiring knowledge of our surrounding environment. This is because the phenomenal constitution principle does not hold for such cognitive activities.

In this section, I have defended the phenomenal justification view. In the next section, I will discuss whether or not this view promotes naïve realism.

4. Introspectively Accessible Factive Reason

Recall that the metaphysical claim of naïve realism is that the visual phenomenology of veridical experience is wholly constituted by environmental objects and their properties, with which the subject is acquainted. As Johnston and Fish emphasized, the structured couples of environmental objects and perceptible properties are plausibly regarded as *truth makers* of demonstrative judgments.[17] Suppose that I see a red apple and form the demonstrative judgment that this apple is red. What makes this judgment true is the structured couple of the perceived apple and its redness, such as the apple being red or the apple having the red trope. It is plausible that we can justify a judgment by referring to its truth maker. Thus, naïve realists can reasonably claim that we can justify a demonstrative judgment by referring to the phenomenology of a veridical experience.

[17] Johnston 2006, Fish 2009.

However, some theories other than naïve realism arguably allow a perceptual experience to be used as a justificatory reason for a demonstrative judgment. According to Kriegel and Horgan and Tienson, the phenomenology of perceptual experience is representational in the sense that perceptual experience has representational content in virtue of having phenomenology.[18] This view is called "phenomenology-first intentionalism".[19] Moreover, some representationalists, such as Chalmers and Brogaard, claim that the phenomenology of perceptual experience is identical to its content which is represented in a certain manner.[20] Following Chalmers, I call this view "impure representationalism".[21] On these views, it is plausible to think that the introspective access to the phenomenology of perceptual experience makes it possible for us to use its representational content as a justificatory reason. Hence, these views can arguably accept the phenomenal justification view.[22]

With respect to perceptual justification, the distinctive feature of naïve realism is that it can allow a perceptual experience to be used as a *factive* reason. A reason whose content is that O is P is factive if and only if the fact that we have the reason entails that O is P. For instance, seeing the structured couple of O and P entails that O is P. Hence, when we use the visual state of *seeing the structured couple of O and P* as a justificatory reason, the reason is factive. Naïve realists can maintain that we can use a perceptual experience as a factive justificatory reason for a demonstrative judgment. This is because if the phenomenology of a veridical experience is constituted by the structured couple of O and P, then the fact that we are undergoing the phenomenology entails that O is P.

In contrast, the phenomenology-first intentionalists and the impure representationalists cannot allow a perceptual experience to be used as a factive reason. This is because such theorists usually accept the common factor principle that the phenomenal aspects of a veridical visual experience and an introspectively indiscriminable hallucinatory experience are the same. Consider a case in which I have a veridical visual experience whose representational content is that O is P. Going by phenomenology-first intentionalism, this case is characterized as follows: I am undergoing the phenomenology that grounds the representational content that O is P. According to impure representationalism, this case is, roughly speaking, characterized as follows: I am undergoing the phenomenology that is identical to the representational content that O is P. Given the common factor principle, however, such theorists would accept that a corresponding hallucinatory experience can have the same phenomenology even though it is not the case that O is P. This means that the fact that I am undergoing the said phenomenology does not entail that O is P. According to phenomenology-first

[18] Kriegel 2007, Horgan and Tienson 2002.
[19] Fish 2010: 67-68.
[20] Chalmers 2004, Broggard 2010.
[21] Chalmers 2004.
[22] It might be disputable whether the representational content of perceptual experience, which is based on or identical to its phenomenology, can be adequately justificatory of *demonstrative* judgments. This is because it might be suspected that such representational content cannot contain demonstrative elements which can pick out external objects. I set this issue aside and assume that such representational content can be somehow justificatory of demonstrative judgments. Given that the aim of this section is to show the advantage of naïve realism over other views, this assumption accords with the principle of charity.

intentionalism and impure representationalism, hence, we cannot use the phenomenology of veridical experience as a factive justificatory reason for the judgment that O is P.[23]

The question to be asked here is, why do we need a factive reason? Whether naïve realism is promoted by adopting the phenomenal justification view depends on the answer to this question. If there is a good reason for allowing a perceptual experience to be used as a factive reason, then we can motivate the adoption of naïve realism.

According to Duncan Pritchard[24], we should embrace introspectively accessible factive reason for two reasons. The first reason is that we can best capture our commonsense about everyday justificatory practices by appealing to introspectively accessible factive reason.[25] The second reason is that we can effectively block radical skepticism by appealing to it.[26] However, these reasons have been criticized by Smithies.[27] His objection to the first reason is, roughly speaking, that our commonsense on such a topic is so complicated that it is difficult to draw a clear evidence for a particular epistemological theory from it. His objection to the second reason is that there are many theories that can block radical skepticism without appealing to factive reason as effectively as naïve realists do. It might be possible to develop and refine Pritchard's argument so as to avoid Smithies objections. Since I agree with Smithies objections, however, I would like to look for another reason.

William Alston maintains that what epistemic activities aim for is nothing less than *truth*.[28] That is to say, the aim of our epistemic activities is to form *true* judgments or beliefs. As Alston points out, there are various kinds of epistemic desiderata.[29] Where X is a feature of a judgment, X is epistemically desirable if X is the sign showing that the judgment is true or likely to be true. For example, it is epistemically desirable that a judgment is formed by a sufficiently reliable judgment-forming process, and that a subject has adequate evidence for a judgment.[30] Though Alston does not explicitly mention, it is unquestionable that it is epistemically desirable to have a factive reason for a judgment. My suggestion is that having a factive reason is *in itself* a distinct epistemic desideratum. Whether we need such a reason in order to effectively block radical skepticism is not essential for its epistemic value. Whether we need such a reason in order to acquire knowledge is not essential as well. In my opinion, the reason why we

[23] Even representationalists, such as Dretske and Tye, who claim that veridical visual experience has de-re or singular content, accept the common factor principle (Dretske 1995: 101-102, Tye 2009: 115-16). This means that they are committed to the idea that the *part* of representational content of perceptual experience which is regarded as identical to its phenomenology is common in veridical and hallucinatory experiences. If *the very part* of representational content can be justificatory of demonstrative judgments, they can accept the phenomenal justification view. Nevertheless, as long as the part of representational content is common in veridical and hallucinatory experiences, they cannot allow a veridical visual experience to be used as a factive reason.
[24] Pritchard 2012.
[25] Ibid: 17-18.
[26] Ibid: 110-52.
[27] Smithies 2013.
[28] Alston 2005: 29.
[29] Ibid: 39-57.
[30] Ibid: 43.

should embrace the idea of factive reason is that it is *in itself* epistemically desirable to have a factive reason. Since having a factive reason for a judgment entails the truth of the judgment, it is one of the most desirable things from the epistemic point of view to have a factive reason. This indicates that it is epistemically much better to admit that a perceptual experience can be used as a factive reason than otherwise.

However, there is a potential objection to this line of argument for naïve realism. Suppose that a veridical experience is introspectively indiscriminable from a hallucinatory experience. According to naïve realism, these experiences provide us with different kinds of reasons for our judgments. The introspective access to the phenomenology of veridical experience makes it possible for us to use the experience as a factive reason for a demonstrative judgment. In contrast, introspection does not make it possible for us to use a hallucinatory experience as such a factive reason. At best, it can be used as a non-factive reason for a demonstrative judgment. Nevertheless, we cannot be introspectively aware of the rational difference between these experiences, for these are introspectively indiscriminable. However, it seems intuitively plausible that if we have a reason in one case and a different reason in the other, then we should be introspectively aware of the rational difference between the two cases. Hence, even if it is epistemically better to admit that a veridical experience can be used as a factive reason, it costs too much because it leads to the implausible consequence that we cannot be introspectively aware of the rational difference between a veridical experience and the corresponding hallucinatory experience.

Due to the space constraints, I cannot discuss this objection to the necessary extent. I just sketch my basic idea of how naïve realists can avoid it. The objection has been described by using three notions: *introspective indiscriminability, introspective access and introspective awareness*. It is obscure what exactly these notions respectively mean. Perhaps, the word "introspective" has different senses in each notion; the objection may rely on the equivocality of "introspective". If this is correct, naïve realists would be able to dissolve the objection by analysing these notions in detail. In my opinion, if we would like to promote naïve realism by appealing to introspectively accessible factive reason, what we should engage in is the attempt to analyse various notions containing the word "introspective" or its cognates.

5. Conclusion

I have discussed how we can promote naïve realism by appealing to the explanatory roles of the phenomenology of veridical visual experience. I have argued for the phenomenal justification view (without appealing to first-person reflection). Given the phenomenal justification view, I have then argued that only naïve realism can coherently allow a veridical visual experience to be used as a *factive* reason.[31] This can be regarded as one theoretical advantage of naïve realism.

[31] Thanks to anonymous referee, and Ichiro Nishida, Joel Smith, Koichi Nakashima, Tomoyuki Yamada, Yu Yoshii. I gratefully acknowledge the support of JSPS.

References

Alston, W. 2005, *Beyond Justification: Dimensions of Epistemic Evaluation*, Ithaca: Cornell University Press.

Bayne, T. and Montague, M. 2011, "Cognitive Phenomenology: An Introduction", in Bayne, T. and Montague, M. (eds.), *Cognitive Phenomenology*, Oxford University Press, 1-34.

Brewer, B. 2011, *Perception and Its Objects*,Oxford: Oxford University Press.

Broggard, B. 2010, "Strong Representationalism and Centered Content", *Philosophical Studies*, 151, 373-92.

Butler, J. 2013, *Rethinking Introspection: A Pluralist Approach to the First-Person Perspective*, London: Palgrave Macmillan.

Campbell, J. 2002a, *Reference and Consciousness*, Oxford: Oxford University Press.

Campbell, J. 2002b, "Berkeley's Puzzle", in Gendler, T. and Hawthorne, J. (eds.), *Conceivability and Possibility*, Cambridge, MA: MIT Press, 127-43.

Campbell, J. 2011, "Relational *vs*. Kantian Responses to Berkeley's Puzzle", in Roessler, J., Lerman, H. and Eilan, N. (eds.), *Perception, Causation, and Objectivity*, Oxford: Oxford University Press, 35-50.

Chalmers, D.J. 2004, "The Representational Character of Experience", in Leiter B. (ed.), *The Future for Philosophy*, Oxford: Oxford University Press, 153-81.

Dretske, F. 1995, *Naturalizing the Mind*, Cambridge, MA: MIT Press.

Fish, W. 2009, *Perception, Hallucination, and Illusion*, Oxford: Oxford University Press.

Fish, W. 2010, *Philosophy of Perception: A Contemporary Introduction*, New York: Routledge.

Gazzaniga, M.S., Ivry, R. and Mangun, G. 2013, *Cognitive Neuroscience: The Biology of the Mind (Fourth Edition)*, New York: Norton & Company.

Horgan, T. and Kriegel, U. 2007, "Phenomenal Epistemology: What is Consciousness That We May Know It so Well?", *Philosophical Issues*, 17 (1), 123-44.

Horgan, T. and Tienson, J. 2002, "The Intentionality of Phenomenology and the Phenomenology of Intentionality", in Chalmers, D. (ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford: Oxford University Press.

Johnston, M. 2004, "The Obscure Object of Hallucination", *Philosophical Studies*, 120 (1-3), 113-83.

Johnston, M. 2006, "Better Than Mere Knowledge? The Function of Sensory Awareness", in Gendler, T. and Hawthorne, J. (eds.), *Perceptual Experience*, Oxford: Oxford University Press, 260-90.

Johnston, M. 2011, "On a Neglected Epistemic Virtue", *Philosophical Issues*, 21 (1), 165-218.

Kennedy, M. 2013, "Explanation in Good and Bad Experiential Cases", in Macpherson, F. and Platchias, D. (eds.), *Hallucination*, Cambridge, MA: MIT Press, 221-54.

Kriegel, U. 2007, "Intentional Inexistence and Phenomenal Intentionality", *Philosophical Perspectives*, 21 (1), 307-40.

Logue, H. 2012a, "What Should the Naïve Realist Say About Total Hallucinations?", *Philosophical Perspectives*, 26 (1), 173-99.

Logue, H. 2012b, "Why Naive Realism?", *Proceedings of the Aristotelian Society*, 112, 211-37.

Martin, M.G.F. 2004, "The Limits of Self-Awareness", *Philosophical Studies*, 120 (1-3), 37-89.

Martin, M.G.F. 2006, "On Being Alienated", in Gendler, T. and Hawthorne, J. (eds.), *Perceptual Experience*, Oxford: Oxford University Press, 354-410.

Pautz, A. 2010, "Why Explain Visual Experience in Terms of Content?", in Nanay, B. (ed.), *Perceiving the World*, Oxford: Oxford University Press, 254-309.

Pritchard, D. 2012, *Epistemological Disjunctivism*, Oxford: Oxford University Press.

Raleigh, T. 2011, "Visual Experience and Demonstrative Thought", *Disputatio*, 4 (30), 171-93.

Smithies, D. 2012, "A Simple Theory of Introspection", in Smithies, D. and Stoljar, D. (eds.), *Introspection and Consciousness*, Oxford: Oxford University Press, 259-93.

Smithies, D. 2013, "Review of Duncan Pritchard, Epistemological Disjunctivism", *Notre Dame Philosophical Reviews*, URL: http://ndpr.nd.edu/news/36590-epistemological-disjunctivism/

Smithies, D. 2014, "The Phenomenal Basis of Epistemic Justification", in Kallestrup, J. and Sprevak, M. (eds.), *New Waves in Philosophy of Mind*, New York: Palgrave MacMillan.

Siegel, S. 2011, *The Contents of Visual Experience*, Oxford: Oxford University Press.

Thagard, P. 2010, "Cognitive Science", *Stanford Encyclopedia of Philosophy*, URL: http://plato.stanford.edu/entries/cognitive-science/.

Tye, M. 2009, *Consciousness Revisited: Materialism Without Phenomenal Concepts*, Cambridge, MA: MIT Press.

Conscious Experiences as Ultimate Seemings: Renewing the Phenomenal Concept Strategy

François Kammerer

Université Paris-Sorbonne

Abstract

The Phenomenal Concept Strategy is a popular strategy used to support physicalism in the realm of conscious experience. This Strategy accounts for dualist intuitions but uses the ways in which we think about our experiences to explain these intuitions in a physicalist framework, without any appeal to ontological dualism.

In this paper, I will raise two issues related to the currently available versions of the Phenomenal Concept Strategy. First, most of the theories belonging to the Phenomenal Concept Strategy posit that phenomenal concepts are exceptional and *sui generis* concepts, and these theories can be shown to be largely *ad hoc*. Second, these theories may explain the *existence* of anti-physicalist intuitions, but they do not explain their *persistence*.

My aim is to put forward a new theory of phenomenal concepts that can rise up to these challenges to the Phenomenal Concept Strategy. In my view, phenomenal concepts are not independent and *sui generis* concepts. On the contrary, they are closely related to our other epistemic concepts, especially our concepts of "justification". Thinking about an experience means thinking about a specific kind of justification – an unjustified justification, or, in other words, an "ultimate seeming". I will show why this explains the existence and the persistence of anti-physicalist intuitions in a non-*ad hoc* way.

Keywords: conscious experience, phenomenal consciousness, hard problem of consciousness, physicalism, dualism, phenomenal concept, cognitive phenomenology

Some philosophers have tried to show that conscious experience does not threaten ontological physicalism, by arguing that anti-physicalist intuitions concerning consciousness (and notably the intuition of conceivability), which sustain the well-known anti-physicalist arguments (Chalmers 1996; Chalmers 2010; Jackson 1982; Kripke 1980), are nothing but a by-product of certain epistemological features of phenomenal concepts (the concepts we use to think about phenomenal experiences notably, but not only, through introspection). In con-

Argumenta 1, 2 (2016): 233-243
ISSN 2465-2334

© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.KAM

temporary philosophy, this line of thought has been labelled "the Phenomenal Concept Strategy" (Loar 1997; Papineau 2002; Tye 2003). The Strategy has been the subject of numerous objections (Ball 2009; Chalmers 2007; Goff 2011; Levine 2007); it constitutes nevertheless one of the most accepted physicalist answers to the anti-physicalist arguments concerning consciousness.

The purpose of this paper is twofold. First, I want to raise two issues against the currently available versions of the Phenomenal Concept Strategy. One concerns the kind of explanations of anti-physicalist intuitions provided by these versions of the Strategy. I will show that they can be construed as typical *ad hoc* explanations, because they have to posit exceptional and *sui generis* entities (that is to say, here, concepts) in order to reach their explanatory goal and to defend a central thesis (physicalism) against refutation. The other issue concerns what is explained by these versions of the Strategy. I want to show that they do not explain the persistence of anti-physicalist intuitions (even if they may explain their existence), which is nevertheless part of the real *explanandum*. For this reason, these versions of the Strategy offer merely incomplete explanations.

My second aim is to develop a conception of phenomenal concepts able to meet these challenges. In this conception, phenomenal concepts are not *sui generis* concepts, but are specific concepts of justifications. Self-ascriptions of conscious experiences using phenomenal concepts amount to self-ascriptions of unjustified justifications (what I will call "ultimate seemings"); and self-ascriptions of these special kinds of justifications are themselves rendered necessary by the norms governing our practices of justification. I think that such a conception can explain the existence and the persistence of anti-physicalist intuitions in a non-*ad hoc* way, and therefore constitute a satisfying version of the Phenomenal Concept Strategy. This conception also has the advantage to maintain the traditional (Cartesian) link between the (hypothetical) metaphysical specificity of conscious experiences and their peculiar justificatory powers—even if this link is reversed compared to what happens in the traditional picture.

I will proceed as follows: First, I will briefly present the current versions of the Phenomenal Concept Strategy and raise the two aforementioned issues against them. Second, I will present my conception, according to which phenomenal concepts are concepts of unjustified justifications. Third, I will show how this explains anti-physicalist intuitions (focusing on the intuition of conceivability). Fourth, I will explain why my conception meets the two challenges. Finally, I will detail various other virtues of my conception.

1. The Phenomenal Concept Strategy and its Problems

Some philosophers have tried to show that conscious experience does not threaten ontological physicalism, by arguing that anti-physicalist intuitions (and notably the well-known "intuition of conceivability") are nothing but a by-product of certain epistemological features of phenomenal concepts (the concepts we use to think about phenomenal experiences through introspection). In contemporary philosophy, this line of thought has been labeled "the Phenomenal Concept Strategy" (Loar 1997; Papineau 2002; Tye 2003; see Stoljar 2005 for the expression).

However, most of the theories developed under this label have in common two flaws: First, they are largely *ad hoc*. Second, they explain the *existence* of anti-physicalist intuitions, but not their *persistence*.

Why can we say that most of these theories are *ad hoc*? First, let us note that these theories often posit that phenomenal concepts have extremely specific features, which make them fundamentally different from other concepts: according to these theories, phenomenal concepts are *sui generis*, exceptional concepts. For example, some theories say that phenomenal concepts are pure recognitional concepts (or pure demonstrative concepts), without any associated mode of presentation (Tye 2003; Levin 2007): they are some kind of "blind demonstratives", conceptually completely independent from other concepts. Some other theories (sometimes labeled "constitutionalists") state that phenomenal concepts include their referents as parts of their own modes of presentation (Loar 1997), and that any use of a phenomenal concept requires the instantiation of the phenomenal property it refers to (Papineau 2002; Balog 2012) (or of a resembling phenomenal property).

The problem is that these features are "exotic" features, shared by no other concepts. In the case of the demonstrative account, we can note that all demonstrative concepts (except phenomenal concepts) seem to have an associated descriptive mode of presentation which serves to fix their referent in a given context (for "this", it would be something like "the thing in front of me"). Concerning the constitutionalist account, we can say that a systematic physical inclusion of the referent within the concept cannot be found in the case of any of the other "normal" concepts (and also that it is hard to see why this feature should be conceptually or cognitively relevant (Levine 2007)).

This means that if phenomenal concepts really did happen to have these features, they would be *sui generis* concepts. But positing a new class of *sui generis* concepts represents a huge theoretical cost; and we have few theoretical reasons to posit that such "strange" concepts exist, except the need to defend physicalism. Indeed, these features seem tailored to explain why anti-physicalist intuitions arise, and there is little or no independent support for their existence. Current theories of phenomenal concepts can therefore be considered *ad hoc*.

Second, and perhaps more importantly, current theories of phenomenal concepts explain the existence of anti-physicalist intuitions, but not their persistence. Indeed, let us admit that phenomenal concepts, given their specificity, give rise to anti-physicalist intuitions. The problem is the following: why can't we just abandon these concepts, and use new concepts instead—new concepts that would probably be quite similar to phenomenal concepts but that would not create such fallacious intuitions?

For example, let us assume that phenomenal concepts are indeed pure demonstrative concepts, without any modes of presentation. Because of this peculiarity, these concepts are supposed to give rise to anti-physicalist intuitions. But if this is true, why should not we just abandon them, and replace them with unproblematic demonstrative concepts (like "this brain state") in order to refer to the states phenomenal concepts referred to?[1] It would not be the first time (see, for example, the history of science) that fallacious concepts are replaced by new and better ones. For instance, to take a well-known example, we no longer use the Aristotelian concepts of "motion" and "speed", because these concepts

[1] This kind of demonstrative could still be used through introspection, if we stipulate that what we reach through phenomenal introspection is necessarily a brain state—that is, if we accommodate our introspective and conceptual practices to our acceptance of ontological physicalism.

generate insuperable theoretical problems. It seems as if we should do the same with phenomenal concepts: get rid of them.

However, it looks like we cannot help using phenomenal concepts. They seem central and indispensable to our mental lives. Now this is what we have to explain: not only why anti-physicalist intuitions arise, but also why these intuitions persist. Currently available theories of phenomenal consciousness do not seem able to cope with this problem. Indeed, they do not explain the "necessity" of phenomenal concepts, nor why it seems we cannot live without them.

2. Phenomenal Concepts as Concepts of Ultimate Seemings

I wish to defend a conception that could avoid these pitfalls. My starting point is the notion of "seeming", which I define as follows: It *seems* to me that P when I have a reason to believe that P (or a justification for believing that P). In other words, it seems to me that P when I am in a situation such as, *ceteris paribus*, I *should* believe that P (in a sense of "should" which implies that I can be held responsible for the fulfilment of this duty).[2] If I am rational, the fact that I believe that P implies that it seems to me that P, but not conversely; I can both believe that it seems to me that P and that not-P. Such a conception does not account for all the ordinary or philosophical uses of "seem", but this is not my purpose.

Rationality commands that I be able to justify my belief that P by self-ascribing a reason to believe that P, and in the same way rationality also commands that I be able to justify my belief that I have a reason to believe that P. That is to say: if I self-ascribe a reason to believe that P, I should be able to justify this self-ascription by making the following mental acts:

(1) *I assert that Q* (a fact distinct from the fact that P and from the fact that it seems to me that P)
(2) *I assert that the fact that Q is a reason to believe that P*
(3) *I assert that (1) and (2) are the reason why it seems to me that P.*

These assertions are not actually made every time it seems to me that P. However, rationality commands that such assertions be available every time it seems to me that P, even if I am not disposed to produce them explicitly; and in various situations I indeed would have to produce them explicitly. It will be the case, for example, when it seems to me that P while not believing that P (if I want to rationalize such situation); also when, in a conversational context, I disagree with another subject, and I want to convince her by expounding and justifying my reasons to believe what I believe; or when I enter into a process of examination and evaluation of my beliefs.

We sometimes self-ascribe some very peculiar kind of "seemings" I call "ultimate seemings". They happen to be what we call conscious experiences. What

[2] Three precisions here: first, one can see that this notion of "seeming" is fundamentally a *normative* notion. Second, this notion of "seeming" expresses an "internal" notion of justification, that is to say, a justification which is necessarily accessible to a subject to which something seems to be the case. This is made necessary by the idea according to which the subject can be held responsible for the epistemic duty conveyed by the seeming; but this doesn't mean that I am committed to an internalist account of epistemic justification—as I will explain later. Third, it is important to keep in mind that the notion of seeming I am now expounding is purely epistemic, and in no way phenomenal (yet).

are these "ultimate seemings"? They are what I self-ascribe when it seems to me that P while, even if I judge that it could be the case that not-P (that is to say, I judge that "P" is a contingent statement), I judge that there is no fact Q which could be used to fulfil the conditions (1), (2) and (3); that is to say, I judge that I cannot justify the fact that it seems to me that P by appealing to another fact.

In other words, when I self-ascribe an "ultimate seeming" that P, I judge that I am in a situation such as, *ceteris paribus*, I *should* believe that P, and in which there is no fact (different from the fact that I should believe that P) in virtue of which I should believe that I should believe that P.

To sum it up, when I self-ascribe introspectively a conscious experience that P:

(A) *I self-ascribe a reason to believe that P (where "P" is a contingent statement)*
(B) *I judge that this reason is not itself justified* (that I have no reason to believe that I have a reason to believe that P)
(C) *I nevertheless maintain, in spite of (B), that I have a reason to believe that P* (maybe because of the nature of our informational, and particularly perceptual, devices: sometimes we do not dismiss an unjustified reason to believe that P, because we cannot help believing that we have a reason to believe that P).

I think that this conception accounts for our introspective self-ascriptions of experiences. To show this would require an extensive argumentation, which I cannot produce here for reasons of space, but I will give an example in order to show how it is supposed to work.

Doris is reading a book next to Frank. Frank believes that Doris is a philosopher. Why does he believe that (that is to say, how can he justify his belief)? Because Doris is reading a book by Kant and because someone reading a book by Kant is a reason to believe that the person is a philosopher. Why does he believe that Doris is reading a book by Kant? Because the words "Kant" are printed on the cover of the book (and because that is a reason to believe that the book is indeed by Kant). Why does he believe that these words are printed on the book cover? Because he sees that they are so printed (and the fact that he sees them so is a reason to believe that they are indeed so). Why does he believe that he sees that? Because it seems to him that it is so; that is to say, it *simply seems to him* that the words are so and so. And at this point, he cannot go further to justify his beliefs; at this point, he will call his last reason to believe that the words are so and so a *visual experience of the words being so and so*. Indeed, he can only self-ascribe a last reason to believe something, which is also an unjustified reason to believe something, an ultimate seeming: a conscious experience.

Even if this example does not replace an argumentation, I think it shows why it is plausible that the self-ascription of a conscious experience does indeed amount to the self-ascription of an unjustified justification to believe something (or, in other words, of an ultimate seeming). We talk and think about our conscious experiences when we are unable to go further in the chain of justification; when the only thing we can say to justify a belief is that "it just seems to me that it is so", and nothing more. And I think that the fact that we are sensitive to the demand for justification of our beliefs (that is to say, that we think that each of our beliefs can be asked to be justified) renders necessary some self-ascriptions of ultimate seemings (on pain of infinite regress).

3. Explaining the Intuition of Conceivability

I believe that this conception of what it means to self-ascribe a conscious experience is plausible. But here is the important point: if this conception is correct, it explains the main anti-physicalist intuition, that is to say, the intuition of conceivability.[3] Indeed, this intuition is based on the fact that it always seems possible to have a conscious experience (say, pain) without the brain state which is supposed to be identical with it (say, C-fiber activation), and conversely. And this appearance of possibility cannot be explained away by saying that in such cases, we do not conceive pain without C-fiber activation, but merely a thing that *appears* to be pain without C-fiber activation. Indeed, when it comes to conscious experiences, "appearing to be pain" and "being pain" are the same thing (Kripke 1980).

This last fact is explained by my conception. Indeed, a thinker cannot think "it seems to me that I have an experience that P but I do not" because such a thought would imply that the subject self-ascribe a reason to believe that she has an experience without having the experience; that is to say, that she self-ascribe a reason to believe that she has an ultimate reason to believe something without having this ultimate reason to believe something. But such a self-ascription is conceptually impossible, because conscious experiences are precisely conceived of as reasons to believe which cannot be justified by anything else. In other words, if experiences, when thought about introspectively, are conceived of as "ultimate seemings" (reasons to believe which themselves are not justified by any reason), *it is by conceptual necessity impossible that I have a reason to believe that I have an experience without having this experience*. Indeed, I implicitly asserted, when introspectively self-ascribing a conscious experience, that such a reason (that I could have without having the experience itself) does not exist. This explains why, when it comes to conscious experiences, there is no appearance/reality distinction (and that itself explains the intuition of conceivability).

This conception of phenomenal concepts is very traditional in a way, even if it consists in a reversal of this tradition. Indeed, according to the Cartesian tradition, it is because conscious experiences have a special, purely mental, non-physical nature, that they can be known in a better and indubitable way, and therefore can constitute the basis for the justification and the foundation of our knowledge.

In my theory, the link between metaphysics and epistemology is maintained, but goes the other way. It is because our concept of justification requires us to ascribe some ultimate justifications (on pain of infinite regress) that we are led to self-ascribe such ultimate seemings, that it to say, unjustified justifications (what we call conscious experiences). And it is because these justifications are conceived as fundamentally unjustified that we cannot make any appearance/reality distinction about them, which makes us tend to think (falsely) of them as having a non-physical nature.

[3] I will not consider here other anti-physicalist intuitions, because it would make my paper too long. Rather, I will do as if these intuitions all amount to the intuition of conceivability. At the end of my paper, I will quickly say a word about other dualist intuitions.

4. Answering the two challenges

A. This theory is not ad hoc

Compared to other theories of phenomenal concepts, this conception has a first advantage: it escapes the objection of being *ad hoc*.

First, in my view, phenomenal concepts are not *sui generis* concepts. They belong to the family of epistemic concepts; more precisely, they are concepts of justification. Of course, they are very specific concepts of justification: they are concepts of *unjustified* justifications. But this specificity itself can be understood and analyzed as being composed of concepts that are already available: an unjustified justification is nothing but a justification which, *de jure*, is not itself justified.

Second, I think that my view is *prima facie* plausible, even if we are not considering the fact that we do have anti-physicalist intuitions. Indeed, if one grants that humans are engaged in practices of justification of their beliefs and that they are sensitive to the demand of justification of their beliefs, she understands why they are led to self-ascribe some unjustified justification to believe things (on pain of infinite regresses in the chains of justification). Therefore, I think that the core of my view concerning phenomenal concepts is plausible even if we set aside the fact that we actually do have anti-physicalist intuitions.

For these reasons, my account escapes the objection of being *ad hoc*.

B. This theory explains the persistence of anti-physicalist intuitions

My view implies that, inasmuch as humans are engaged in practices of justification and are sensitive to the demand for justification concerning their beliefs, they will be led to self-ascribe some ultimate seemings (understood as unjustified justifications). Then, anti-physicalist intuitions concerning the metaphysical nature of these seemings will naturally arise.

Besides, getting rid of these anti-physicalist intuitions would require to get rid of phenomenal concepts (understood as concepts of unjustified justifications). But it does not seem possible as long as we are engaged in practices of justification, given that the use of phenomenal concepts is made necessary by the very functioning of these practices of justification.

So, my account shows not only why anti-physicalist intuitions arise; it also explains why these intuitions persist, and why we cannot get rid of phenomenal concepts. Indeed, anti-physicalist intuitions are a consequence of some deep and fundamental features of our nature as epistemic subjects: the fact that we are located creatures, of whom it is required that they give reasons for their beliefs.

C. A few more precisions

One could think that my account is committed to epistemic foundationalism, or to epistemic internalism, which are both contested views. However, I am only committed to the thesis that our folk concepts of justifications are internal concepts of justifications, and that in our folk practices of justifications we are sensitive to the demand for justification of our beliefs (sensitivity which leads, when philosophically systematized, to epistemic foundationalism and epistemic internalism). These points seem supported by the fact that epistemic internalism and

epistemic foundationalism were the mainstream during the vast majority of history of philosophy.

It is also important to note that my thesis does not concern the nature of phenomenal states, but only the nature of our *concepts* of phenomenal states. It states that, when we self-ascribe a certain phenomenal experience through introspection (using phenomenal concepts), we self-ascribe an unjustified justification to believe something (what I called an ultimate seeming). It does not state that phenomenal experiences really *are* ultimate seemings. Indeed, I have shown that ultimate seemings are conceived in a way which precludes us to understand them as physical. That explains why anti-physicalist intuitions arise concerning phenomenal consciousness; that also means that, if physicalism is true (as I think it is), there cannot *be* ultimate seemings in the real world (and phenomenal experiences, inasmuch as they exist, cannot be ultimate seemings). However, this fact does not preclude us to conceive (falsely) of them, through introspection, as being that way.

On another vein, I want to make clear that my account can retain some advantages of the classical theories of phenomenal concepts. For example, one advantage of the demonstrative conception is that it explains the "fineness of grain" of phenomenal concepts (the fact that, for example, we can introspectively discriminate many more colour hues, on the basis of a current visual experience, than what we are able to conceptualize when we do not experience them). However, my own conception, being partly compatible with the demonstrative theory of phenomenal concepts, can also explain this phenomenon. Indeed, if the self-ascription of a phenomenal experience amounts to the self-ascription of an ultimate seeming that P, nothing precludes "P" to be a thought constituted by concepts which are themselves demonstrative (and fine-grained) concepts.

5. Other virtues of this approach

I will now detail three other virtues of my approach. First, note that most of the theories of phenomenal concepts explain anti-physicalist intuitions as being the result of brute psychological facts (for example, of some hard-wired features of human brains). These theories offer a mere causal explanation of these intuitions. On the contrary, my account offers a real understanding of why these intuitions arise and persist, by locating them in a comprehensive picture of our conceptual practices of justification. In other words, my account offers a meaningful explanation of anti-physicalist intuitions. I think this to be an advantage of my account.

Second, I think that my view is not only able to explain the intuition of conceivability, but also other important dualist intuitions. For example, my view can account for what we could call the "intuition of subjectivity". It is often said that conscious experiences are problematic for physicalism because they appear to be fundamentally subjective, in a way that no physical object is. They are always *for* a subject (Nagel 1974); one can also say that there is an intrinsic "for-me-ness" in every conscious experience (Levine 2001; Kriegel 2005). But if we indeed think about our conscious experiences as "seemings" (understood as reasons to believe), this is explained in quite a natural way. Indeed, a reason to believe always concerns a subject which is responsible for the epistemic duties which constitute this reason to believe. Without a subject who has beliefs and can be held responsible for them, there can be no reasons to believe understood

in this sense. This explains why conscious experiences always concern a subject and are *for* a subject—because it is the subject herself who is concerned by the normative dimension conveyed by a reason. The case here is quite similar to the case of an order: an order always concerns someone who is ordered to do something by someone else. There can be no order that is not addressed to someone who is *concerned* by the order (because, if the order creates a duty, this duty must be fulfilled by the person who received the order and no one else). In the same way, there can be no conscious experience without an epistemic subject concerned by the epistemic duty; any experience will be *for* such a subject.

My view also has a third advantage: indeed, it not only addresses the metaphysical problem of consciousness. It also gives a framework able to solve some of the current debates concerning the existence of a "cognitive phenomenology" (Bayne and Montague 2011).

There is an on-going debate concerning the existence of a cognitive phenomenology. Some philosophers assert that there is a phenomenology specific to thoughts, while others do not accept any phenomenology above and outside the sensory and the emotional. The following dialectic often takes place: the defenders of the existence of a cognitive phenomenology put forward examples of phenomenal contrasts, where two mental states differ from a cognitive and a phenomenal point of view, but not from a sensory point of view. For example, it is often said that two subjects who hear the same sentence pronounced, say, in French, would have two different experiences if one of them understands French and the other does not, whereas they would have the same sensory phenomenology (Strawson 1994: 5-6). On the other hand, the opponents of cognitive phenomenology highlight cases where two states differ from a cognitive point of view, without any phenomenal difference. For example, it seems that judging that 17 is a prime number does not *feel* any different from judging that 19 is a prime number (Nichols and Stich 2003: 196).

I think that my view offers a framework which accounts for these cases and these distinctions, because it predicts which thoughts are likely to be considered as presenting a phenomenal aspect and which are not.

Indeed, in my view, the self-ascription of a phenomenal experience amounts to the self-ascription of an ultimate seeming. Such an ultimate seeming is what we self-ascribe when we think that we have a reason to believe something even if we think that there is no justification for this reason itself; it is an unjustified justification.

Now let us think about the way new information is made available for a cognitive system like ours. Sometimes this information is the result of a conscious inference; at other times this information is the result of a sub-personal and unconscious process. On my view, we have to expect that the availability of a new piece of information would have a phenomenal aspect in the second case, but not in the first case. Indeed, in the first case, this information can be justified by the subject, which has access to the evidential basis of this information (that is to say, to the premises of the conscious inference). In the second case, the subject is unable to justify this new information which just happens to "come" to her. This information, which can be considered by the subject as a reason to believe something about the world,[4] is itself unjustified: when thinking about this

[4] Namely, to believe that the situation represented by the information is indeed the case.

information, the subject will conceive it of as an "ultimate seeming", an unjustified reason to believe—it will be thought to be phenomenally given.

Therefore, we have to expect that any information which is made available to the subject without the evidential basis for this information being made available itself would be thought by the subject to be phenomenal. This is precisely the case of sensory, perceptual and hedonic information, but this is also the case of some types of "cognitive" information produced by encapsulated and unconscious devices, such as the identification of certain objects, of faces, the understanding of an oral or written sentence by an expert speaker/reader, etc. So, my view predicts that some cognitive information will be considered as phenomenal by the subject, and these kinds of information precisely match the typical cases of experiential thoughts put forth by the defenders of the existence of a "cognitive phenomenology". I take it to be a confirmation of my view. It is also a theoretical virtue, because it shows that my conception of phenomenal concepts could set the basis for a framework able to solve some problems belonging to the "cognitive phenomenology" debate, by explaining which thoughts are considered by subjects as phenomenal, which are not, and why it is the case.[5]

References

Ball, D. 2009, "There Are No Phenomenal Concepts", *Mind*, 118 (472), 935-62.

Balog, K. 2012, "Acquaintance and the Mind-Body problem", in Hill, C. and Gozzano, S. (eds.), *New Perspectives on Type Identity: The Mental and the Physical*, Cambridge: Cambridge University Press, 16-42.

Bayne, T. and Montague, M. (eds.) 2011, *Cognitive Phenomenology*, Oxford: Oxford University Press.

Chalmers, D. 1996, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.

Chalmers, D. 2007, "Phenomenal concepts and the explanatory gap", in Alter, T. and Walter, S. (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford: Oxford University Press, 167-94.

Chalmers, D. 2010, "The Singularity: A Philosophical Analysis", *Journal of Consciousness Studies*, 17 (9), 7-65.

Goff, P. 2011, "A Posteriori Physicalists Get Our Phenomenal Concepts Wrong", *Australasian Journal of Philosophy*, 89 (2), 191-209.

Jackson, F. 1982, "Epiphenomenal Qualia", *Philosophical Quarterly*, 32, 127-36.

Kriegel, U. 2005, "Naturalizing Subjective Character", *Philosophy and Phenomenological Research*, 71, 23-56.

Kripke, S. 1980, *Naming and Necessity*, Cambridge, MA: Harvard University Press.

[5] I am very grateful to audiences at the SIFA Conference in Cagliari (September 2013) and the Graduate Conference in Urbana-Champaign (March 2014), for the useful feedback. I also would like to thank Joseph Levine, Daniel Andler, Pascal Ludwig, Fabien Mikol and Isabelle Montin for their help. Finally, I want to thank the two anonymous reviewers of *Argumenta* for their remarks.

Levin, J. 2007, "What is a Phenomenal Concept?", in Alter, T. and Walter, S. (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford: Oxford University Press, 87-110.

Levine, J. 2001, *Purple Haze: The Puzzle of Consciousness*, Oxford: Oxford University Press.

Levine, J. 2007, "Phenomenal Concepts and the Materialist Constraint", in Alter, T. and Walter, S. (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford: Oxford University Press, 145-66.

Loar, B. 1997, "Phenomenal States (Revised Version)", in Block, N., Flanagan, O. and Güzeldere, G. (eds.), *The Nature of Consciousness*, Cambridge, MA: MIT Press, 597-616.

Nagel, T. 1974, "What Is It Like to Be a Bat?", *Philosophical Review*, 83, 435-50.

Nichols, S. and Stich, S. 2003, "How to Read Your Own Mind: A Cognitive Theory of Self-Consciousness", in Smith, Q. and Jokic, A. (eds.), *Consciousness: New Philosophical Perspectives*, Oxford: Oxford University Press, 157-200.

Papineau, D. 2002, *Thinking about Consciousness*. Oxford: Oxford University Press.

Stoljar, D. 2005, "Physicalism and phenomenal concepts", *Mind and Language*, 20 (2), 296-302.

Strawson, G. 1994, *Mental Reality*, Cambridge, MA: MIT Press.

Tye, M. 2003, "A Theory of Phenomenal Concepts", in O'Hear A. (ed.), *Minds and Persons*, Cambridge: Cambridge University Press, 91-112.

The Ancient Master Argument and Some Examples of Tense Logic

Fabio Corpina

University of Cagliari

Abstract

The Master Argument of Diodorus Cronus has been long debated by logicians and philosophers. During the Hellenistic period it was so famous that doxographers and commentators took for granted its notoriety and none of them gave us a detailed report. The first section presents a philosophical account of the ancient Master Argument, by trying to retrace its meaning, originated from the Megarian context, and so halfway between ancient logic and metaphysics. The second section introduces a logical analysis of the Master Argument against the backdrop of the Jarmużek-Pietruszczak semantics for the tense logic K_t4P; but the main aim of the section is to deal with one of the most fascinating attempts to peruse the Master Argument, i.e. A. Prior's reconstruction. Prior stays true to the Diodorean philosophical stance even if he uses modern logical tools. The significance of the work by Prior marks the beginning of tense logic. The third section expounds an argument by Øhrstrøm-Hasle. Danish logicians do not consider additional premises for the Master Argument. They give, *in primis*, a sentential example for the third premise, proving its inconsistency with the first two. The deterministic conclusion is the implicit result of this stratagem. Finally, in the fourth section, we compare the strategies by Prior and Øhrstrøm-Hasle.

Keywords: Arthur Prior, Diodorean logic, modalities, semantics for tense logic, time and tenses

1. The ancient Master Argument

The debate about the doctrine of potency by Aristotle is a *vexata quaestio* in the Ancient context. It had wide appeal among the contemporaries of the Stagirite and the topic was dealt with great interest by the Megarian philosophers, the strongest opponents of Aristotle.

In a first period, the Megarian thesis seemed to vouch for the position to which the *incipit* of Arist., *Metaph*. IX, 3, alludes.[1] But the more articulate Megarian thesis involves such a use of the temporal notions within the modal no-

[1] The Megarian thesis implied in this passage is made too trivial. The critics appear do not object to the official version by Aristotle. However, an alternative view which does justice to the Megarians is in Makin 1996.

Argumenta 1, 2 (2016): 245-258
ISSN 2465-2334

© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.COR

tions: the craftiest adversary of Aristotle is Diodorus Cronus with his *kurieuon logos*, i.e. the Master Argument. Several logicians evaluate the ancient Master Argument as the best rejoinder to Arist. *Int.* IX. For instance, Jarmużek 2009 introduces to a linear model within a semantics for linear future; Gaskin 1995 deals with a metaphysics of the future. It reviews the texts by Aristotle and Diodorus in light of the formalism and tools of modern logic.

The most complete report of the Master Argument is in Epict., II 19, 1. However, the report restores only the three premises of the argument and its conclusion. About the strategy we know no more than that Diodorus ruled out the third premise to obtain the conclusion. Vuillemin 1996 suggests an interesting reconstruction. It is an account in which the Diodorean argument is closely related to Arist. *Cael.* I, 283b 6-17. However we are completely unaware of what was the deduction process to obtain the main thesis: Nothing is possible which is neither true now nor ever will be. The conclusion is the consequence of the alternative view of the potency, by the Megarian Diodorus. The perspective of Diodorus links the possibility to the actuality, and it does so by the interdefinition of modal and temporal notions. In Boeth. *in Int. sec. ed., 234* is referred that Diodorus defines the possible as what is now or will be in some future instant, and the necessary as what is now and will always be hereafter.

Here is the Master Argument as it is in Epict., II 19, 1:

> The Master Argument appears to have been propounded on the strength of some such principles as the following. Since there is mutual contradiction between these three propositions, to wit:
>
> - Everything true as an event in the past is necessary
> - The impossible does not follow from the possible
> - What is not true now and never will be, is nevertheless possible
>
> Diodorus, realizing this contradiction, used the plausibility of the first two proposition to establish the principle
>
> - Nothing is possible which is neither true now nor ever will be.

In general we note that the propositions are temporally definite statements.

The first sentence means the irrevocability of the past: what has occurred in the past cannot be differently from what was the case. Therefore, it is and will always be true as a given occurrence in the past. We can state that: "If Christopher Columbus arrived in the Americas, then it is necessary that he did it"; i.e. Columbus discovered the Americas, entails that it is and will always be true that there exists a given instant in the past in which it is true that he did it.

The second sentence has often created some problems. In fact, the first step is to understand the meaning of the verb "to follow" in the context; the second step is to establish a correct interpretation of the modal notions in the tricky endeavour of avoiding a vicious circle with the conclusion.

"To follow" translates the Greek verb *akolouthein*. It has different meanings: "to occur subsequently in time", "to imply", and "to be in accordance with", are the most plausible. However the range of these meanings is very wide, and the term has a considerable importance in order to interpret the second premise. To interpret the verb *akolouthein* as "to follow in time /after" (cf. e.g. Zeller 1882, Rescher 1966), is out of place when it is used by a crafty dialectician as Diodo-

rus.[2] From a logical point of view the most accurate translation of it seems "to infer", "to entail", maybe in a Diodorean sense (cf. e.g., Mates 1973, Denyer 1981: 40-41). But on the other hand, it would be a mistake to underrate the third solution: "to be in accordance with" hints to a kind of modal principle of non-contradiction in relation to possibility (cf. e.g., Becker 1956, Mignucci 1966: 11-15), i.e. if a proposition is possible, at the same time its impossibility is ruled out. Better yet, this formulation appears at least suggested by the wide sense of the second proposition of the Master Argument.

Another question is what "possible" and "necessary" mean according to Diodorus. The definition of the possible is obviously temporal, in the sense of "what is already now or will be in a given future". That of necessary has been usually interpreted temporally as "what is and will be hereafter", but we could also interpret it differently: e.g. "what, being true, will not be (at the same time) false"—in this case, the necessity would lose its *strictu sensu* modal value.

The fact that Christopher Columbus arrives in the Americas is not in accordance with the fact that he arrives in India. More clearly, if it is not the case that possibly Christopher Columbus arrives in the Americas and at the same time he is in India, but he is in India, then possibly he is in India. The incompatibility expressed in the previous argument is evident, because it locates Christopher Columbus in two different places. We might also say that, if it is impossible for Christopher Columbus to be in India, if the fact that he is in the Americas necessarily implies that he is in India, then it is impossible for him to be in the Americas. Therefore, if Christopher Columbus is not in India, and neither will be there in the future, this *(not-)*occurrence leads to impossibility.

2. An introduction to Prior's investigation of the Diodorean frame and his perspective on time

To look for an adequate formalisation and to guess the strategy for the ancient argument of Diodorus Cronus was an important step in Arthur N. Prior investigation of the Diodorean frame.[3] The interest for ancient logic, the debate about indeterminism and determinism, a particular attention to C.I. Lewis' modal systems, are decisive factors that stimulated Prior in formalising the Master Argument and discussing Diodorean modalities (see Prior 1955, 1967, and the critics Denyer 2009, Ciuni 2009). These researches led to the birth of tense logic.

What is tense logic? The name is in use since the Sixties of the last century,

[2] In S.E. *M.* VIII 112, the explanation for the Diodorean implication, and therefore also for the verb *akolouthein*, excludes that "to follow" may be meant on a temporal sequence.

[3] Arthur Norman Prior, born in New Zealand in 1914. He was an eclectic scholar with various interests in logic and philosophy. Prior's studies in logic are in non-classical logic area, and they have been developed in the Fifties and Sixties of the last century. But Prior was very keen on several philosophical fields (metaphysics, theology, ethics, history of logic). Prior obtained his major results in logic during his stays at Manchester University and at the Balliol College in Oxford. References are Prior 1957, Prior 1967, the posthumous collection Prior 2003. He died unexpectedly in 1969. Today a critical study on his unpublished works is underway, namely, *the Virtual Lab for Prior Studies* (http://research.prior.aau.dk/login_user.php). The *Arthur Prior Centenary Conference*, August 20-22, 2014, Balliol College, Oxford (http://conference.prior.aau.dk/) grouped together scholars from all the world.

when Prior was looking for an intermediate system between S4 and S5. His aim was to discover an adequate system that formalises Diodorean modalities.

In fact, Prior was a logician, and an expert in history of logic too. He was interested in translating modal notions into temporal ones, as it used to be in the Hellenistic period (see e.g., Boeth. *in Int. sec. ed., 234*; Cic. *Fat.* VII 13, IX 17; Alex.Aphr. *in A.Pr.* I 183,34-184,10 *etc.*), whit modern tools.

Relevant in Prior research was the investigation of the correlations between tenses and sentences.

Let us notice that Prior seems to attribute a kind of ontological supremacy to the present tense.

Prior supported a version of *presentism* (see Orilia 2012: 107-36, and Dorato 2013 for recent examinations about the theme), according to which what is present is what is real (Prior 1972: 320). Tense operators do not form propositions out of propositions: by prefixing to a sentence p a temporal operator, we specify a property in a given time. However, the sentence p alone, is already an English present progressive sentence. Using the same strategy as the ancient philosophers, Prior need not create any explicit temporal-index-link for sentences. Prior's account refers to statements which already correspond to propositional functions, and the truth-value of a proposition can vary from time to time.[4]

So, Prior's ideas are the starting point of contemporary temporal logics.

In this section we present Prior's formalisation of the Master Argument and his hypothesis about its conclusion. First, we mention Jarmużek-Pietruszczak semantics for the tense logic K_t4 + Prior's formula $p \wedge Gp \rightarrow PGp$, namely the semantics for the K_t4P system. Actually, the declared aim of Jarmużek and Pietruszczak 2009 (86) is not to express a minimal logic for Prior's Master Argument, but to study K_t4P. K_t4P, in fact, allows for a selection of characteristic properties. Following Jarmużek and Pietruszczak 2009, we hope to disclose the logical power involved in the Master Argument. A *strictu sensu* algebraic semantics for the Diodorean modal systems was proposed in the memorable study of (Bull 1965).

Let us briefly recap Prior's formalisation of temporal and modal operators:

Fp: "It will be the case that p" (Weak future operator)
Gp: "It will always be the case that p" (Strong future operator)
Pp: "It has been the case that p" (Weak past operator)
Hp: "It has always been the case that p" (Strong past operator)
$\Diamond p$: "Possibly p", i.e. $p \vee Fp$
$\Box p$: "Necessarily p", i.e. $p \wedge Gp$

Observe that a sentence may be true at a given time, and false at another.

[4] For instance, "It was the case that Columbus is discovering the Americas", as a sentential case for Pp, is false before he did it, and it is always true from his coming (in 1492); (2) "It will be the case that I am attending an advanced Logic course and Barack Obama is the President of the United States", as a sentential case for F($p \wedge q$), is true from the beginning of my Ph.d career two days a week, with the proviso that Obama does not resign from his position and I will continue to work in logic. On the other hand, F($p \wedge q$) is always false before I started my Ph.d career, and false five days a week from the beginning of my doctoral studies. Further, F($p \wedge q$) is false both, in the case in which Obama or myself decide to leave the respective employment, and definitively false after the end of Obama's term of office.

We should think in terms of *propositional functions*: sentences are the arguments of the operators. Further, if a formula is a law, then for every substitution, we obtain a proposition true at all times.

A minimal logic for Prior's Master Argument consists of the following set of axioms and rules:

1. $G((p \rightarrow q) \wedge Gp) \rightarrow Gq$
2. $Gp \rightarrow GGp$
3. $(p \wedge Gp) \rightarrow PGp$
4. $PGp \rightarrow p$
5. $Gp \rightarrow Fp$
6. $\Box((p \rightarrow q) \wedge \Box p) \rightarrow \Box q$
7. $\Box p \rightarrow p$
8. $\Box p \rightarrow \Box\Box p$

MP. $\vdash (p \rightarrow q) \wedge p \Rightarrow \vdash q$
RG. $\vdash p \Rightarrow \vdash Gp$
R$\Box$. $\vdash p \Rightarrow \vdash \Box p$

2.1 State of the art and outline of a recent temporal semantics

Prior's study of the Master Argument focuses on the inter-definability between modal and temporal notions. Prior explicitly cites Boethius and other ancient and medieval authors, in the discussion of a *logic of futurity*.[5]

Boeth. *in Int. sec. ed., 234.* gave the following account of Diodorus Cronus' modal notions:

> Diodorus establishes to be possible, what is or will be; to be impossible, what being false, it will be not true; to be necessary, what being true, it will be not false; to be non-necessary, what is or will be false.

One of the purposes of this paper is to take advantage of the tools of contemporary logic. We aim to express some fundamental notions about time and modal categories. Boethius is the first to suggest the notions above. Boethius develops the discussion both on a philosophical and linguistic level. On the other hand, we hope to trace some developments proposed to examine a modern Diodorean system.

First, how many Diodorean systems have been examined by logicians?

Before dealing with Prior's strategy of the Master Argument—the main topic of the paper—we summarise the best attempts of building a Diodorean logic.

In fact, the Master Argument should be consistent with a Diodorean logic. Many logicians analysed different schemas for time features, both on a syntactic and semantical level.

Second, which Diodorean properties are relevant for a modal or temporal system?

Third, which class of frames does satisfy the Diodorean properties?

To begin with, it is useful to define a state of art. The next step will clarify a

[5] By admitting the rule defined like *mirror image* by C.L. Hamblin, i.e. the replacing between specular time operators, we can theorise from a "logic of futurity" by Prior, a "logic of pastness". It is sufficient to substitute P to F, H to G.

basic temporal semantics, to interpret some logics for Prior's Master Argument.

In spite of the different languages and approaches, it is possible to scientifically explain the temporal meaning of modal notions: we will show the power of a pure tense logic linked to the Master Argument. In particular, we will focus on some frames by running over the semantics for K_t4P system.[6]

It is not easy to identify the exact number of Diodorean systems. By a historical analysis, we understand that it is more a sequence of results, rather than a collection of systems.

The search for a Diodorean frame (see Ciuni 2009) starts from Prior's studies on an intermediate logical system between Lewis' modal S4 and S5.

Prior conjectured that the Diodorean Frame was an analogue of S4. In (Prior 1957) we find the author's reasons, the outstanding one was its *reflexivity* and *transitivity*. Since at the time Parry's S4.5 was believed the only intermediate system between S4 and S5 (cf. Parry 1939), later discovered to be equivalent to S5, the Diodorean frame should have been S4.[7]

According to Hintikka 1958 and Dummett and Lemmon 1959, the Diodorean frame does not correspond to S4. In fact, the Diodorean system should include the modal $\Diamond p \wedge \Diamond q \rightarrow (p \wedge q) \vee \Diamond(p \wedge \Diamond q) \vee \Diamond(q \wedge \Diamond p)$—or some analogue—to preserve a *transitive and linear* accessibility relation on the frame. However, there are some transitive frames that falsify the previous formula. So, an intermediate modal system including the axiom for linearity was gathered: S4.3.[8]

Nevertheless, Dummett and Lemmon (1959) pointed out that S4.3 does not include *discreteness*, e.g. $\Box(\Box(p \rightarrow \Box p) \rightarrow \Box p) \rightarrow (\Diamond \Box p \rightarrow \Box p)$, while we know that an adequate Diodorean system has an atomistic notion of time. So, Bull (1965) proved the Diodorean frame as *discrete*, *reflexive*, *transitive* and *linear*, and Zeman (1968) identified this logic in S4.3.1.[9]

For brevity sake, I mentioned only a schema of the most relevant results, while Ciuni (2009) provides a detailed account of the search for the Diodorean frame in a modal logic analogue system.

Many interpretations of the Diodorean system have been proposed from the Eighties of the last century. Much has been done on a semantical level (*e.g.*, White 1984, Trzęsicki 1987, or Zanardo 2009), and the Diodorean system has been interpreted in very different fields, for instance, the physical Minkowski spacetime account (see Goldblatt 1980).

I wish to take stock of the situation about some semantics of the Diodorean system in order to discuss the logic of the Master Argument. I proceed by looking into the class of frames which satisfies the Diodorean properties before comparing Prior's Master Argument to what I name Danish Master Argument. In particular, I will examine a semantics for a pure tense logical system, in the spirit of Boethius' translation from a modal to a temporal notion.

Following Jarmużek and Pietruszczak 2009, I will analyse the characteristic

[6] We obtain K_t4P system from K_t4 tense logic + (P): $p \wedge Gp \rightarrow PGp$. Further, we know that: $K_t4 = K_t$ [1. $G(p \rightarrow q) \rightarrow (Gp \rightarrow Gq)$; 2. $H(p \rightarrow q) \rightarrow (Hp \rightarrow Hq)$; 3. $p \rightarrow HFp$; 4. $p \rightarrow GPp$] + ax. $PPp \rightarrow Pp = K_t$ + ax. $Hp \rightarrow HHp = K_t$ + ax. $Pp \rightarrow GPp = K_t$ + ax. $FHp \rightarrow Hp$.

[7] *Reflexivity* and *transitivity* are characteristic properties for S4. The above-named properties are respectively described by the following axioms:

T. $\Box p \rightarrow p$; 4. $\Box p \rightarrow \Box\Box p$.

[8] S4.3 = T + 4 + ax. $\Box(\Box p \rightarrow \Box q) \vee \Box(\Box q \rightarrow \Box p)$.

[9] S4.3.1 = S4.3 + ax. $\Box(\Box(p \rightarrow \Box p) \rightarrow \Box p) \rightarrow (\neg\Box\neg\Box p \rightarrow \Box p)$.

formula (P) for the tense logic K_t4P. Jarmużek and Pietruszczak 2009 maintains K_t4P as the pure tense logic analogue of the Diodorean system, therefore, including the fundamental premises of the Master Argument.

The characteristic formula (P), namely $p \wedge Gp \rightarrow PGp$, is the equivalent formula of what I will call (+d) in the next section, namely the second additional premise of Prior's Master Argument.

Let me first recapitulate some ideas from Jarmużek and Pietruszczak 2009.

A frame F is a (P)-frame iff the formula (P) is valid in F.
F = <*T, R*> is defined from the relation ' ‹ ' of *immediate-precedence/succession*;
F is a LIP-frame[10] *iff*

$\forall_{x \in T}(x \text{ not-}R\ x \Rightarrow \exists_{y \in T}\ y \lessdot x)$;

F is a BC-frame[11] *iff*

$\forall_{x, y, z \in T}(x \lessdot y\ \&\ x\ R\ z\ \&\ y \neq z \Rightarrow y\ R\ z)$.

LIP-BC-frames are the class of frames satisfying LIP and BC properties. Therefore in Jarmużek and Pietruszczak 2009 (98) we find the following theorem:

THEOREM: (i) F *is a* (P)*-frame iff* F *is a* LIP-BC*-frame*.
(ii) F *is an irreflexive* (P)*-frame iff* F *is an* IP-BC*-frame*.[12]

1) Some treelike IP-frames are not BC-frames, therefore they are not (P)-frames.
2) Some linear and BC-frames are not IP-frames, therefore they are not (P)-frames.
3) Some irreflexive, transitive, right-total (P)-frames (so also IP-BC-frames) are not treelike frames.
4) There is a frame F = <*T, R*> such that F is a treelike (P)-frame, namely a IP-BC-frame, but it is not right-total, i.e.: $\exists_{x, y, z \in T}(z\ R\ x\ \&\ z\ R\ y\ \&\ x \neq z\ \&\ x \text{ not-}R\ y\ \&\ y \text{ not-}R x)$. Therefore, the *branching condition* is weaker than *linearity*. There are some branching but not-linear frames.

It may be useful to see Jarmużek and Pietruszczak 2009, in order to study and compare the resulting tree-graphics. Clearly, frames associated to *Diodorean conditions* guarantee several interpretations of the Diodorean temporal account.[13]

So far, we proposed some hints on the power of tense logics, in particular for K_t4P system. In the next step will confine our discussion to a formal strategy to explain the ancient Master Argument in the modern language of tense logics.

[10] Every LIP-frame is characterised by the relation of *limited immediate precedence* between two ordered temporal points. (i) All reflexive frames are LIP-frames; (ii) All irreflexive LIP-frames are IP-frames and conversely; (iii) All IP-frames are left-discrete and cannot have a minimum.
[11] Every BC-frame is characterised by the *branching condition*. (i) All reflexive frames are BC-frames; (ii) All right-total frames are BC-frames.
[12] Proof of the theorem is in Jarmużek and Pietruszczak 2009: 98.
[13] White (1984) considers semantical assumptions for discreteness secondary: only the assumption on irreflexivity is necessary. Differently, Trzęsicki (1987) needs a tense-logical semantics satisfying the condition of discreteness. Therefore, the author conclusion is that: even if we introduce irreflexivity, this property is not sufficient to infer the Master Argument conclusion.

2.2 Prior's strategy

Let us start with Prior formalisation of the Master Argument (see Prior 1955: 209-13).

(a) When anything has been the case, it cannot not have been the case:
$Pp \rightarrow \neg\Diamond\neg Pp$

(b) If anything is impossible, then anything that necessarily implies it is impossible:
$\neg\Diamond q \rightarrow (\Box(p \rightarrow q) \rightarrow \neg\Diamond p)$

(+c) When anything is the case, it has always been the case that it will be the case:
$p \rightarrow HFp$ / or directly $\Box(p \rightarrow HFp)$[14]

(+d) When anything neither is nor will be the case, it has been the case that it will not be the case:
$(\neg p \wedge \neg Fp) \rightarrow P\neg Fp$

(z) What neither is nor will be true, is not possible.
$(\neg p \wedge \neg Fp) \rightarrow \neg\Diamond p$

The argument is a modern formulation of the ancient *kurieuon logos*. But (a) and (b) are modern translations of Diodorus' first and second premises, while (+c) and (+d) are premises from Prior. However, both (+c) and (+d) would denote two theses of Diodorus, even if omitted by his *kurieuon logos*. Then, (+c) and (+d) are introduced since they can be considered impliedly accepted by Diodorus.

In particular, (+c) is connected to the position expressed e.g. in Cic. *Fat.,* XII 27, which means: if p is true now, then at any instant in the past it was the case to say that p will be true. For, the actual now was a time in the future, seen from the past.

Further, (+d) alone does not allow for determinism. In fact, we are able to obtain, e.g., an IP-BC-frame which is branching but not-linear. Of course, we cannot say that an IP-BC-frame was in Diodorus' mind, and surely (+d) is necessary to infer the deterministic conclusion of the Master Argument.

In any case, we assume time as a discrete sequence, in order to respect a historically faithfully Diodorean interpretation.[15]

Here is Prior's strategy to prove the conclusion (z):

1. $(p \rightarrow q) \rightarrow ((q \rightarrow r) \rightarrow (p \rightarrow r))$ [*Instance of the law of transitivity*]
2. $(p \rightarrow (q \rightarrow r)) \rightarrow (q \rightarrow (p \rightarrow r))$ [*Instance of the law of exchange*]
3. $P\neg Fp \rightarrow \neg\Diamond\neg P\neg Fp$
   (a) $p/\neg Fp$ [*Substitution in (a)*]
4. $P\neg Fp \rightarrow \neg\Diamond HFp$
   by df. $H = \neg P\neg$ [*3 defined by H*]

[14] Prior considered the string $\Box(p \rightarrow HFp)$ as (+c) in (Prior 1967); while in Prior 1955 (211) (+c) is not prefixed by the box ($\Box$), although the previous formulation is deduced at a later stage. In this paper we note the passage at line 10.

[15] Cf., both, *S.E. M. 10, 119-120 and previous*, for an historical view on Diodorus Hellenistic atomistic account; and Zeman 1968 for a contemporary system, namely S4.3.1, as the adequate atomistic outline for Diodorus account.

5. $((\neg p \wedge \neg F p) \rightarrow P\neg F p) \rightarrow (((P\neg F p \rightarrow \neg\Diamond HF p) \rightarrow ((\neg p \wedge \neg F p) \rightarrow \neg\Diamond HF p))$
   (1) $p/\neg p \wedge \neg F p$; $q/P\neg F p$; $r/\neg\Diamond HF p$
   [*Substitutions in (1) in order to obtain some instance of the law of transitivity composed by* $(+d) \rightarrow (4 \rightarrow ((\neg p \wedge \neg Fp) \rightarrow \neg\Diamond HFp))$]

6. $(\neg p \wedge \neg F p) \rightarrow \neg\Diamond HF p$
   $(+d) \rightarrow (4 \rightarrow 6)$
   [*(+d) is supposed true like a premise of the Master Argument, 4 is proved, and since we are considering some instance of a law like in 5, then it is impossible for 6 to be false. Therefore 6 is proved.*]

7. $((\neg p \wedge \neg F p) \rightarrow \neg\Diamond HF p) \rightarrow ((\neg\Diamond HF p \rightarrow (\Box(p \rightarrow HF p) \rightarrow \neg\Diamond p)) \rightarrow ((\neg p \wedge \neg F p) \rightarrow$
   $\rightarrow (\Box(p \rightarrow HF p) \rightarrow \neg\Diamond p)))$
   (1) $p/\neg p \wedge \neg F p$; $q/\neg\Diamond HF p$; $r/\Box(p \rightarrow HF p) \rightarrow \neg\Diamond p$
   [*Substitutions in (1) by obtaining some instance of the law of transitivity*]

8. $\neg\Diamond HF p \rightarrow (\Box(p \rightarrow HF p) \rightarrow \neg\Diamond p)$
   (b) $q/HF p$ [*Substitution in (b)*]

9. $(\neg p \wedge \neg F p) \rightarrow (\Box(p \rightarrow HF p) \rightarrow \neg\Diamond p)$
   $6 \rightarrow (8 \rightarrow 9)$
   [*6 is proved, 8 is proved, and since we are considering an instance of a law like in 7, then it is impossible for 9 to be false. Therefore 9 is proved.*]

10. $\Box(p \rightarrow HF p)$
    (+c) by RL [*By applying the necessitation rule to (+c)*]

11. $(\neg p \wedge \neg F p) \rightarrow (\Box(p \rightarrow HF p) \rightarrow \neg\Diamond p)) \rightarrow (\Box(p \rightarrow HF p) \rightarrow ((\neg p \wedge \neg F p) \rightarrow \neg\Diamond p))$
    (2) $p/\neg p \wedge \neg F p$; $q/\Box(p \rightarrow HF p)$; $r/\neg\Diamond p$
    [*Substitutions in (2), to obtain an instance of the law of exchange composed by* $9 \rightarrow (10 \rightarrow (z))$]

(z) $(\neg p \wedge \neg F p) \rightarrow \neg\Diamond p$
   [*9 is proved, 10 is proved, and since we are considering some instance of a law like in 11, then it is impossible for (z) to be false. Therefore (z), i.e. the conclusion of Prior's Master Argument, is proved.*]

Prior's proof of (z) uses (a) and (b) of the ancient Master Argument, and by adding (+c) and (+d) attains Diodorus' conclusion.

Nevertheless, Prior accepts the validity of Diodorus argument, but objects to its soundness. In fact, Prior is afraid of validating the determinism.

Prior criticizes the truth of (+d). He supposes ½ as the truth value for future propositions that are not true from now. In fact, Prior shows the conclusion of the argument, but the truth value of (z) is ½.[16]

However, Diodorus supported determinism. He would not admit a third value. Moreover, it is possible to obtain (+d) from the fourth axiom of Hamblin's system, namely $p \vee Pp \leftrightarrow \neg F\neg Pp$.[17] This is relevant since even if Hamblin

[16] During a first period, swayed by Łukasiewicz, Prior was inclined to think that the only chance to attain an indeterministic tense logic was via a three valued system. Prior (1966) seems to accept other solutions also.

[17] H4. $p \vee Pp \leftrightarrow \neg F\neg Pp$
$\neg F\neg Pp \rightarrow p \vee Pp$
$\neg F\neg Pp \rightarrow \neg(\neg p \wedge \neg Pp)$ [by *De Morgan*]
$\neg p \wedge \neg Pp \rightarrow F\neg Pp$ [by *contrapositon*]
i.e. (+d): $\neg p \wedge \neg F p \rightarrow P\neg F p$ [by *mirror image*]

claimed that time was dense, Prior noted that the previous axiom supports the discreteness of time, and Diodorus' account was atomistic (Denyer 1981: 49).

Further, if we contrapose (+d), we get the formula $HFp \rightarrow p \vee Fp$, that seems to codify both discreteness and determinism.

3. The Danish Master Argument

A fascinating reconstruction of the Master Argument is by P. Øhrstrøm and P. Hasle.[18]

They propose do not add any additional premise to Diodorus argument. Nevertheless, the authors require some background assumptions:

(a) time is discrete;[19]
(b) the relation $\boldsymbol{T}(t, p)$ means "p is true at t". Further, the verb *akolouthein* in the second premise refers to Diodorean implication, defined by $(p \Rightarrow q)$ iff $(\forall t)(\boldsymbol{T}(t, p) \rightarrow \boldsymbol{T}(t, q))$;
(i) the Master Argument refers to statements which correspond to propositional functions.

These assumptions should be considered along with the following definitions of possibility and necessity from Boeth., *in Int. sec. ed., 234*: $\Diamond p \leftrightarrow p \vee Fp$, $\Box p \leftrightarrow p \wedge Gp$.

The first premise of the Master Argument is as in Prior: $Pp \rightarrow \Box Pp$.

The second premise entails the concept of Diodorean implication, which is formalised as: $((p \Rightarrow q) \wedge \Diamond p) \rightarrow \Diamond q$.

Finally, the third premise is $\neg q \wedge \neg Fq \wedge \Diamond q$.

Øhrstrøm and Hasle use semantical methods to show the contradiction between the third premise and the previous two.

As a first step, they assume as a hypothesis about the meaning of q, allowing that $\neg q \wedge \neg Fq \wedge \Diamond q$, i.e. the excluded juncture by Diodorus.

"Dion is here" is q. Further, let w be the statement "The prophet says: Dion will never be here", that is supposed to be true only in the atomic instant immediately before the present instant.

Hence, Pw is false at any past time, and it is true from now on.

Prior gives an analysis of Hamblin's system in (Prior 1967: 45-50); Hamblin deals with the theme in the correspondence preserved in *Prior's Nachlass*.

[18] Peter Øhrstrøm (Aalborg University) deals with the concept of time, philosophical logic and ethics. Per Hasle (Copenhagen University) is an expert in temporal logics and computer science. They are leading the research about Arthur N. Prior and the Foundations of Temporal Logic. Øhrstrøm and Hasle 1995 (23-28) is the relevant publication to understand their philosophical background and to define what we name Danish Master Argument.

[19] In the case (a) is brought into question—but we believe it is not the Diodorean case—Øhrstrøm and Hasle (1995) suggests to substitute (a) by (A). Namely, no proposition has a first instant of truth. If a proposition is true, it has already been true for some time (Arist. *Phys.,* 236a 12-14): it is true over intervals with last but without first instant of time.

From the first premise $Pp \rightarrow \Box Pp$, we are able to get for the present time the formula $Pw \rightarrow \Box Pw$, and write the consequent as $\neg\Diamond\neg Pw$.

Then, we are also able to get the following matrix, by (a), where t_0 stands for the present time, t with positive n for the future, and t with negative n for the past:

t_{-3}	t_{-2}	t_{-1}	t_0 (now)	t_1	t_2	t_3
$?q$	$?q$	$?q$	$\neg q$	$\neg q$	$\neg q$	$\neg q$
$\neg w$	$\neg w$	w	$\neg w$	$\neg w$	$\neg w$	$\neg w$
$\neg Pw$	$\neg Pw$	$\neg Pw$	Pw	Pw	Pw	Pw

We deduce the *Diodorean implication* between q and $\neg Pw$, that is $q \Rightarrow \neg Pw$. In fact, it is evident that $(\forall t) \neg (\boldsymbol{T}(t, q) \wedge \boldsymbol{T}(t, Pw))$, therefore $(\forall t) (\boldsymbol{T}(t, q) \rightarrow \boldsymbol{F}(t, Pw))$.

From the Master Argument, and assumption (i), we get the second premise $(p \Rightarrow q) \wedge \Diamond p \rightarrow \Diamond q$, from substitutions p/q, $q/\neg Pw$.

Therefore we obtain $(q \Rightarrow \neg Pw) \wedge \Diamond q \rightarrow \Diamond\neg Pw$.

But by the substitutions in the first premise, we already get $\neg\Diamond\neg Pw$. A contradiction with the last sentence and the consequent $\Diamond\neg Pw$. Moreover, we also obtain the negation of the second sentence of the Master Argument, i.e. the impossible does not follow from the possible, therefore Øhrstrøm and Hasle rule out the third proposition $\neg q \wedge \neg Fq \wedge \Diamond q$.

Conclusion

In the second and in the third section we presented Prior's Master Argument and Danish Master Argument, respectively.

We close the paper with a comparison between these accounts; finally, we introduce a "philosophical provocation" about temporal schemas in *computer science*.

We should notice that Prior's Master Argument includes the formalisation of the original *kurieuon logos* and from four premises deduces the conclusion. On the other hand, the Danish Master Argument formalises the Hellenistic argument, but does not propose any decisive strategy to infer the conclusion. In fact, Øhrstrøm and Hasle assume the third premise, which contradicts the first two.

Moreover, Prior has used four premises, two of them are Diodorean, while the other two are supposed to be consistent with Diodorus' doctrine.

Øhrstrøm and Hasle' Master Argument achievement consists in avoiding new premises.

However, to reach their goal they require some assumptions, namely (a), (b), (i).[20]

Let's see how the premises are used in these different accounts.

[20] (i) guarantees the opportunity to substitute sentences or constants to the variables in the premise.

- The Danish first premise is (a) in Prior. It suffices to define the box (□) by the diamond (◊) and *vice versa*.
- The second premise of the Master Argumet is different in Prior's argument and in the Danish one. In Prior it is $\neg\Diamond q \rightarrow (\Box(p \rightarrow q) \rightarrow \neg\Diamond p)$, in Øhrstrøm and Hasle's version is $((p \Rightarrow q) \wedge \Diamond p) \rightarrow \Diamond q$. The second premise of the Danish Master Argument requires some interpretation for the big arrow. That is a semantics for the Diodorean connective ⇒ : $(p \Rightarrow q)$ iff $(\forall t)(\boldsymbol{T}(t, p) \rightarrow \boldsymbol{T}(t, q))$. And yet, it is provable that the Prior second premise is equivalent to the Danish Master Argument second premise.[21]
  In any case the formalisation of Øhrstrøm and Hasle extends the system from propositional logic to first order logic.
- If (A) is, in some way, valid in Prior's account also, (i) is properly present. If we substitute an actual sentence to a variable in a law we still have a tautology.
- Contrarily to Øhrstrøm and Hasle, even if Prior uses some additional premises, he refuses the assumption of the modal definitions. In fact, the same Alex. Aphr. *in APr.* I, 184, 5 mentioned that the *kurieuon logos* was proposed by Diodorus to obtain the modal definitions, in particular for the possible.

In general, Prior strategy achieves his goal, step by step, on the syntactic side via a Hilbert style proof. On the other hand, Danish Master Argument seems more perspicuous on the semantics side, by exemplifying or considering explicit counterexamples.

In both proofs, we are trying to define time and modality, the metaphysical topic of Diodorus Cronus (cf. Denyer 1999), using the tools of modern tense logic.

Let us conclude by observing that temporal logical tools can prove successful in fields as diverse as the analysis of an ancient metaphysical text and algorithm design, in particular *artificial intelligence*, *software engineering* (Galton 1987), and *model checking* (Clarke and Glundberg 1999).[22]

References

Becker, O. 1956, "Ueber den *kurieuon logos* des Diodoros Kronos", *Reinisches Müseum für Philologie*, 99, 289-304.

Bull, R.A. 1965, "An Algebraic Study of Diodorean Modal Systems", *The Journal of Symbolic Logic*, 30 (1), 58-64.

Ciuni, R. 2009, "The Search for the Diodorean Frame", *HumanaMente*, 8, Ciuni, R.

[21] We prove the equivalence from Prior's formula to the Øhrstrøm and Hasle's one:

1. $\neg\Diamond q \rightarrow (\Box(p \rightarrow q) \rightarrow \neg\Diamond p)$
2. $\neg (\Box(p \rightarrow q) \rightarrow \neg\Diamond p) \rightarrow \neg\neg\Diamond q$ [by *contraposition*]
3. $\neg (\Box(p \rightarrow q) \rightarrow \neg\Diamond p) \rightarrow \Diamond q$ [by eliminating the *double negation*]
4. $(\Box(p \rightarrow q) \wedge \Diamond p) \rightarrow \Diamond q$ [inferred by *Chrysippus C11*]

Since $(p \Rightarrow q)$ iff $(\forall t)(\boldsymbol{T}(t, p) \rightarrow \boldsymbol{T}(t, q))$, we get $\Box(p \rightarrow q) \equiv \neg\Diamond(p \wedge \neg q) \equiv (p \Rightarrow q)$. Therefore, 1 (Prior's second premise) is equal to 4 (Øhrstrøm and Hasle's second premise).

[22] I am grateful to the anonymous referees. Their feedbacks and suggestions have enriched my paper, expanding my knowledge on the topic. And last but certainly not least, I would like to thank my supervisors at the University of Cagliari, Francesco Paoli and Antonio Ledda, for their patience and expertise.

(ed.), *Models of Time*, 47-65: www.humanamente.eu/Issues/Issue8.html

Clark, E. and Grumberg, O. 1999, *Model Checking*, Cambridge, MA: MIT Press.

Denyer, N. 1981, "Time and Modality in Diodorus Cronus", *Theoria*, 47, 31-53.

Denyer, N. 1999, "The Master Argument of Diodorus Cronus. A Near Miss", *Philosophiegeschichte und Logische Analyse*, 2, 239-52.

Denyer, N. 2009, "Diodorus Cronus: Modality, the Master Argument and Formalisation", *HumanaMente*, 8, Ciuni, R. (ed.), *Models of Time*, 33-46: www.humanamente.eu/Issues/Issue8.html.

Dorato, M. 1994, *Modalità e temporalità. Un raffronto tra le logiche modali e le logiche temporali*, Roma: Bagatto.

Dorato, M. 2013, *Che cos'è il Tempo*, Roma: Carocci.

Dummett, M.A.E. and Lemmon, E.J. 1959, "Modal Logic between S4 and S5", *Zeitschrift für Mathemathische Logik und Grundlagen der Mathematik*, 5, 250-64.

Galton, A. 1987, *Temporal Logics and their applications*, London: Academic Press.

Gaskin, R. 1995, *The Sea Battle and the Master Argument*, Berlin: De Gruyter.

Goldblatt, R. 1980, "Diodorean Modality in Minkowski Spacetime", *Studia Logica*, 39, 219-36.

Hintikka, J. 1958, Review of "Time and Modality" by A.N. Prior, *The Philosophical Review*, 67 (3), 401-404.

Jarmużek, T. and Pietruszczak A. 2009, "The Tense Logic for Master Argument in Prior's Reconstruction", *Studia Logica*, 92, 85-108.

Jarmużek, T. 2009, "Master Argument VS. Sea-Fight Tomorrow", *Bulletin of the Section of Logic*, 38 (3), 205-14.

Makin, S. 1996, "Megarian Possibilities", *Philosophical Studies*, 83, 253-76.

Mates, B. 1973, *Stoic Logic*, Berkeley: University of California Press.

Mignucci, M. 1966, "L'Argomento Dominatore e la Teoria dell'Implicazione in Diodoro Crono", *Vichiana*, 1, 3-28.

Orilia, F. 2012, *Filosofia del Tempo*, Roma: Carocci.

Parry, W.T. 1939, "Modalities in the Survey System of Strict Implication", *The Journal of Symbolic Logic*, 4 (4), 137-54.

Prior, A.N. 1955, "Diodoran Modalities", *The Philosophical Quarterly*, 5, 205-13.

Prior, A.N. 1957, *Time and Modality*, Oxford: Oxford University Press.

Prior, A.N. 1966, "Postulates for Tense-Logic", *American Philosophical Quarterly*, 3 (2), 153-61.

Prior, A.N. 1967, *Past, Present and Future*, Oxford: Clarendon Press.

Prior, A.N. 1972, "The Notion of the Present", in Fraser, J.T. *et al*. (eds.), *The Study of Time*, Berlin: Springer, 320-23.

Prior, A.N. 2003, *Papers on Time and Tense*, 2nd expanded edition, Braüner, T. *et al*. (eds.), Oxford: Oxford University Press.

The Prior's Nachlass, at the *Bodleian Library*, Oxford.

Rescher, N. 1966, "A Version of the 'Master Argument' of Diodorus Cronus", *Journal of Philosophy*, 36, 438-45.

Trzęsicki, K. 1987, "Is Discreteness of Time Necessary for Diodorean Master Argument?", *Bulletin of the Section of Logic*, 16 (3), 125-31.

Vuillemin, J. 1996, *Necessity or Contingency*, Stanford: CSLI.

White, M.J. 1984, "The Necessity of the Past and Modal-Tense Logic Incompleteness", *Notre Dame Journal of Formal Logic*, 25 (1), 59-71.

Zanardo, A. 2009, "Modalities in Temporal Logic", *HumanaMente*, 8, Ciuni, R. (ed.), *Models of Time*, 47-65: http://www.humanamente.eu/Issues/Issue8.html.

Zeller, E. 1882, "Ueber den *kurieuon* des Megarikers Diodorus", *Sitzungsberichte der Kgl. Akademie der Wissenschaften*, 1, 151-59.

Zeman, J.J. 1968, "The Propositional Calculus *MC* and its Modal Analog", *Notre Dame Journal of Formal Logic*, 9 (4), 294-98.

Øhrstrøm, P. and Hasle, P. 1995, *Temporal Logic. From Ancient Ideas to Artificial Intelligence*, Dordrecht: Kluwer.

Williamson on the psychological view

Serena Maria Nicoli

University of Torino

Abstract

What is the nature of the evidence provided by thinking about hypothetical cases, such as those presented in the thought experiments (TE)? Is it psychological, as those who speak about intuitions seem to think, or not? This problem is closely related to that of the nature of the subject matter of philosophy, that most philosophers tend to conceive as non-psychological.

Williamson's position on the matter (Williamson 2007) consists in rejecting the psychological view on intuitions: if we want this method—the armchair method—to provide us with evidence in favour or contra theses or theories concerning the non-psychological subject matter of our inquiry, then we must understand the evidence, collected "by thinking about the cases", as non-psychological as well.

Unlike Williamson, Brown (Brown 2011) thinks that the psychological view on intuitions can be maintained: an indirect approach to the object of our inquiry is feasible; the gap between these data and the non-psychological object they are supposed to provide evidence for can be closed.

The main aim of this paper is to argue (against Williamson and with Brown) that a revision of the classical view on intuitions is not required. My strategy consists in adopting a Wittgensteinian perspective on the nature of the aims and results of the philosophical inquiry; in showing how this can help us provide an easy solution to the gap problem; and in arguing—against Williamson—that conceiving the subject matter of philosophy as conceptual does not necessarily amount to conceive it as psychological.

Keywords: intuitions, thought experiments, psychological view, conceptual analysis.

What is the nature of the evidence provided by thinking about hypothetical cases, such as those presented in the thought experiments (TE)? Is it psychological, as those who speak about intuitions seem to think, or not? This problem is closely related to that of the nature of the subject matter of philosophy, that most philosophers tend to conceive as non-psychological. Williamson's position on the matter consists in rejecting the psychological view on intuitions: if we want this method—the actual method, i.e. the armchair method—to provide us with evidence in favour or contra theses or theories concerning the non-

Argumenta 1, 2 (2016): 259-269
ISSN 2465-2334

© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.NIC

psychological subject matter of our inquiry, then we must understand the evidence, collected "by thinking about the cases", as non-psychological as well.

In the next section, Williamson's argument against the psychological view on intuitions and his rejection of the intuition-talk are examined.

In section 2, I object that, by arguing that the evidence must be conceived as non-psychological in order to make a certain kind of project plausible, Williamson is addressing the wrong problem: the problem does not consist in deciding how the evidence provided by the armchair methodology should be understood in order to avoid the sceptical challenge. On the contrary what we should ascertain is whether the data collected through assessing actual and hypothetical cases could in fact provide evidence for the relevant object of inquiry.

In section 3, an alternative solution to the gap problem is presented: it is Brown's reliabilist proposal. Brown's solution has the advantage of avoiding Williamson's move. However, apart from proposing an analogy with the question of perceptual evidence and roughing in the reliabilist solution, Brown doesn't explain how and whether this can work for the problem at issue.

Therefore, in section 4, I introduce a third proposal. This proposal maintains the most obvious and acknowledged view on intuitions, i.e., that intuitive judgements are the expression of what we would say about X in such and such circumstances (real or counterfactual) in the light of our linguistic or conceptual competence. In addition, the third solution conceives the request of a theory on X (i.e., the non-psychological object of the inquiry) as a demand for explicit norms for the use of the term "X"—for the application of the concept X.

Section 5 expounds the criticisms that Williamson advances in the seventh chapter of *The Philosophy of Philosophy* against the renegotiation of the nature of the aims and results of philosophical inquiry: Williamson believes that conceiving the object of the philosophical inquiry as conceptual amounts to conceiving it as psychological, and that a research on concepts cannot satisfy the expectations that philosophers have when they ask for a theory on X.

The last part of the paper consists in a defence of the third proposal from these two criticisms. In section 6, I present Wittgenstein's position on the nature of the results of the philosophical inquiry. In section 7, I briefly illustrate how an inquiry starting from our intuitions can indeed lead to results satisfying philosophers' expectations. Apropos, I suggest that two aspects should be taken into account: the structure of the method leading from intuitions to the theory (it's reflective equilibrium method, a method that is both descriptive and revisionary) and the normative aspect of the intuitive judgements.

1. The argument in *The Philosophy of Philosophy*

In the seventh chapter of *The Philosophy of Philosophy* (2007), Williamson argues against a position that he dubs the *psychological view*. He takes it to be the mainstream stance on the nature of the verdicts expressed at the end of the thought experiments (TE) and, in general, of all the verdicts that philosophers treat as the evidential basis for their inquiries. His presentation of the view and argument against it can be briefly reconstructed as follows.

Many contemporary analytic philosophers

> think that, in philosophy, ultimately our evidence consists only of intuitions (to use their term for the sake of argument). Under pressure, they take that to mean not that our evidence consists of the mainly non-psychological putative facts which are the contents of those intuitions, but that it consists of the psychological facts to the effect that we have intuitions with those contents, true or false. On such a view, our evidence in philosophy amounts only to psychological facts about ourselves (Williamson 2007: 235).

What philosophers are then supposed to do is

> to infer to the philosophical theory that best explains the evidence. But since it is allowed that the philosophical questions are typically not psychological questions, the link between the philosophical theory of a non-psychological subject matter and the psychological evidence that is supposed to explain becomes problematic (Williamson 2007: 5).

In particular, as Williamson argues, the psychological view ends up encouraging scepticism, since "psychological evidence has no obvious bearing on many philosophical issues" (Williamson 2007: 234), which mainly concern non-psychological matters. This view, and the intuitions-talk altogether, should then be abandoned. Indeed, philosophers should recognize the non-psychological nature of the evidence they have access to: "our evidence in philosophy consists of facts, most of them non-psychological, to which we have appropriate epistemic access" (Williamson 2007: 241).

Let us take, for instance, the Gettier cases (GC) and the theory these cases are supposed to provide evidence against, i.e., the justified true belief (JTB) theory of knowledge. A genuine counterexample to the JTB theory of knowledge, Williamson argues, would be a case of a justified true belief without knowledge, not, as the tradition presents it, the fact that it seems to one that this is the case. This would in fact raise

> the challenge of arguing from a psychological premise, that I believe or we are inclined to believe the Gettier proposition [the proposition that the Gettier subject has a non-knowledge justified true belief] to the epistemological conclusion, the Gettier proposition itself. The gap is not easily bridged (Williamson 2007: 211).

So the psychological proposition that it seems to one as if the subject in the GC has a non-knowledge justified true belief is not and cannot be a counterexample to the JTB theory of knowledge. What is rather needed, for the argument to work, is the fact itself: the fact that the subject does not know.

Therefore, Williamson concludes, if we want to keep the idea that TE can in fact provide us with evidence pro or contra certain generalizations, we have then to acknowledge that, thinking about the scenarios described in the TE, we have access to the relevant facts *themselves*.

2. An objection to Williamson's move

Briefly, this is how Williamson's argument can be summarized: given the fact that the subject matter of philosophy is non-psychological and that the way to investigate it is widely based on TE, how should we conceive the nature of the

evaluations we express at the end of TE? Not as psychological—we would in fact expose ourselves to the sceptical challenge—but as directly concerning the object of our inquiry.

This argument is straightforward, but nevertheless suspicious. One could in fact object that, by arguing that the evidence must be conceived as non-psychological in order to make a certain kind of project plausible, Williamson is addressing the wrong problem: the problem does not consist in deciding how the evidence provided by the armchair methodology should be understood in order to avoid the sceptical challenge. On the contrary, what we should ascertain is whether the data collected through thinking about actual and hypothetical cases could in fact provide evidence for the relevant object of inquiry. Namely, what we are trying to understand is whether the actual methodology is suitable for a certain kind of project, or not. We could end up answering yes, that given the nature of the evidence provided by thinking about what we would say in actual and counterfactual situations, one could in fact engage in that kind of project; or no, that one could not. If the latter was the case, we would have two choices: we could conclude that the methodology is inappropriate, and that philosophy, as it is pursued, is a hopeless enterprise; or we could conclude that the methodology is appropriate, but we have characterized the object of philosophical inquiry in the wrong way.

So the nature of the TE evaluations should be established *preliminarily*, rather than *a posteriori* as Williamson does, on the basis of a methodology, the current one, that he wants to keep hold, and on the basis of a certain way to characterize the non-psychological nature of the aims and results of philosophy, that he conceives as the only possible.

3. Brown's proposal

An attempt to avoid the move Williamson makes is that proposed by Brown in "Thought Experiments, Intuitions and Philosophical evidence" (2011). Brown agrees with Williamson on the two points we have just considered: the nature of the subject matter of philosophy and the characterization of its method. However, Brown denies that arguing in favour of the possibility for this method to provide evidence for the object, so conceived, forces us to review the classical position about the nature of the evidence provided by TE (and by armchair methodology in general): if the only problem with the classical view on intuitions were just its vulnerability to the sceptical challenge, then, in order to go on supporting it, it would be sufficient to find a good argument against the sceptics.

Let us see how Brown argues. First, she proposes an analogy with the question of the nature of perceptual evidence. Then, she examines the different strategies supporters of the psychological view for perception adopt against scepticism and consider their effectiveness for the matter at issue. We will not take into account the detailed report of the internalist solutions that Brown makes and the reasons she gives for rejecting them, but we will move directly to the approach she favours: reliabilism.

Pursuing the parallel between perception and intuition, she argues in this way: suppose that the method of forming beliefs about the external world—about the non-psychological facts of philosophy—on the basis of perceptual experiences—on the basis of psychological/intuited propositions—is reliable, then

beliefs formed in that way—perceiving or intuiting—are correct. But how can we establish if a certain belief-forming method is generally reliable? Usually, she answers, by evaluating whether the appropriate external relations hold. Here is an example:

> suppose that, when one has the experience as of a large barking dog in front of one, one forms the belief that there is a large barking dog in front of one. On the externalist approach to justification, such as reliabilism, as long as the appropriate external relations hold, the beliefs so formed are justified (Brown 2011: 513).

Applying this remark to the case we are interested in—the GC—gives the following answer: suppose that, when one has the intuition that the subject in the GC does not know, one forms the belief that subject does not know. According to reliabilism, as long as the appropriate external relations hold, the beliefs so formed are justified. Therefore, in order to affirm (not just to suppose) that a belief-forming method is reliable and the particular belief so formed is correct, the challenge is to check whether the external relations, i.e., the relations in the world corresponding to those expressed by the majority of the beliefs formed by that beliefs-forming method, generally hold.

Unfortunately Brown does not go further: she does not explain neither how these relations are conceived, nor how one is supposed to check if they hold.

4. A third solution to the gap problem

In the following paragraph, I introduce a third solution to the gap problem, which is alternative to Williamson's and Brown's. At the beginning, however, I present this solution as an attempt to implement Brown's suggestion.

Let us go back to GC. The argument is traditionally presented in the following way: if the JTB theory of knowledge were true, it would follow that the subject in the GC would know that p, but it is clear, Gettier says, that he does not, or, better, it is clear that we would not say that he does. Here the idea is that, given the way we use "knowledge" (or given our concept of knowledge), this belief would not be called "knowledge" (would not be categorized as *knowledge*).

Let us then face the question at issue: what does entitle us to say that Gettier's conclusion is not wrong? How do we know that the judgement Gettier or, in general, epistemologists express at the end of the case is correct? One can answer that we can legitimately believe that the judgement they express is correct—can in fact provide evidence for the object under inquiry, where the object under inquiry is the norm governing the use of the word "knowledge" (or the norm governing the application of the concept *knowledge*) in our community—on the basis of their semantic (conceptual) competence, that is to say their ability to use the word "knowledge" (to apply the concept *knowledge*) in a manner that tends to reflect the way that word is used (that concept is applied) in the community.

So, let us generalize this proposal and see how it can be applied to Brown's suggestion: the particular judgement on X the philosopher expresses at the end of the TE is correct—it gives inquirer legitimate evidence about X, where X (the non-psychological object of our inquiry) is the norm that governs the use of

the term "X" in our community—under the *proviso* that the way the inquirer himself or herself forms judgements on X is generally reliable. But how do we know that the way the inquirer forms himself or herself judgements on X is generally reliable? Usually, by considering whether he or she is semantically (conceptually) competent, i.e., capable of using "X" in inferences that connect "X" with other words (that connect the concept *X* with other concepts) and able to apply the word "X" (the concept *X*) in correspondence to real and imaginary situations, in a way that tends to reflect the way that same word is used in the community. Hence, if the way he or she usually uses "X" actually matches the way the same word is used in the community—if these external relations hold—then we could say that he/she is competent. Furthermore, being the particular judgement he or she expresses at the end of the TE yielded by his or her competence, we can conclude that this judgement is correct. Namely, competence, the ability that guides one in everyday life, enabling one to be successful, is the same ability that guides one in the imaginary case.

At this point, two remarks are needed: the former concerns competence and the idea that the capacities involved in the evaluation of the cases described in the TE are ordinary capacities. The latter concerns the way in which the nature of the non-psychological object, the psychological data at our disposal are supposed to provide evidence for, can be conceived.

First, saying that the ability we use to evaluate a case like Gettier's is an ordinary ability (such as competence), means recognizing—as Williamson does—that no special faculty is involved in the evaluation of TE. Under this respect my position is very closed to his. Williamson, however, would refuse to describe the ordinary capacity (or better, one of the ordinary capacities) involved in the evaluation of TE in terms of linguistic or conceptual competence: namely, all the first part of *The Philosophy of Philosophy* is devoted to argue against the idea that philosophy is committed with something peculiarly conceptual or linguistic.

Second, I see my solution as a development of Brown's idea but in no way I mean to suggest that she would endorse it. I said that the judgement on X stated at the end of the TE is correct as long as a person is semantically/conceptually competent, i.e., capable to use the word "X" in a manner that tends to reflect the way the word is used in the community he or she belong to. But at this point, one could point out the following: this solution works under the assumption that what we are in fact looking for are the norms for the use of the word "X" (norms for the application of the concept *X*); but this maybe does not work under Williamson's assumption, i.e., that what we are looking for are the necessary true propositions about the entity X. In other words, it is pretty clear—although not trivial—how competence can provide mostly reliable data about the object under investigation (e.g., knowledge) when the query of it is understood in terms of a query of rules for the usage of the term in question. By contrast, it is not clear how competence could provide any evidence if the object of our inquiry were substantial, were the entity itself, as Williamson puts it.

Later on I say more about how an inquiry starting from our competence is supposed to bring us to the norms for the use of "X". On time being, it is important to remark that Brown would disagree with the way in which I tried to substantiate her proposal. She in fact believes, as Williamson does, that, when we ask "What is X?", we are asking about the necessary truths about the *entity*

X. Brown does not mention the possibility to conceive the aims and results of philosophical inquiry differently than so. However Williamson does.

5. Williamson on the possibility to renegotiate the nature of the aims and results of the philosophical inquiry

In the seventh chapter of *The Philosophy of Philosophy*, Williamson asks himself: could a reinterpretation of the nature of the aims and results of the philosophical inquiry lead to an alternative resolution of the gap problem? Here is his answer:

> Attempts have been made to close the gap by psychologizing the subject matter of philosophy. If we are investigating our own concepts, our application of them must be relevant evidence. But this proposal makes large sacrifice for small gains. As seen in early chapters, the subject matter of philosophy is not distinctive in any sense. Many epistemologists study knowledge, not just the ordinary concept of knowledge. Metaphysicians studying the nature of identity over time ask how things persist, not how we think or say they persist (Williamson 2007: 211).

This consideration is made in the wake of the criticisms to conceptual analysis Williamson made in the first part of *The Philosophy of Philosophy.* There he argued against the idea of philosophy as a conceptual inquiry on the basis of two points: (1) the incapacity to discriminate between allegedly conceptual questions and obviously non-conceptual ones, and (2) the non-viability of the notion of analytic truth, that he sees as the best explication of the notion of a conceptual truth. In the seventh chapter, however, the objection Williamson raises is different from (1) and (2).

Here the idea is that conceptual analysis is not viable because what philosophers are looking for, when they ask what knowledge, identity, reference, causation, etc. are, is something non-psychological. Conceptual analysts look rather for concepts, so, Williamson says, something mental and therefore uninteresting for the concerns of philosophy, when they are properly understood. This objection is different from (1) and (2) and can, therefore, be dealt with separately.

Hence, in what follows I am not going to take a stand against (1) and (2), but I am just going to argue against the specific problem he raises in this quotation.[1] So, against (3) the idea that conceiving the object of the philosophical inquiry as conceptual necessarily amounts to conceiving it as psychological.

6. Wittgenstein and Williamson on conceptual analysis

First of all, claiming that the aims and results of the philosophical inquiry are conceptual does not necessarily amount to saying that they are psychological: indeed, some rules are not. In particular, rules which are outlined and adjusted

[1] That said, the specific strategy I adopt against (3) has interesting consequences on (2) as well: namely, embracing the idea that the results of conceptual inquiry amount to norms/grammatical propositions means refusing the idea that a conceptual investigation amounts to the search for conceptual truths and that conceptual truths are to be identified with analytic truths.

on the basis of GC-like judgements (i.e., judgements stating what we would say in specific circumstances) are not.

At this point, a clarification of what I mean exactly by saying that philosophical theses/theories are (non-psychological) norms is in order. To explain it, I will follow the arguments proposed by Marconi in "Wittgenstein and Williamson on Conceptual Analysis" (2011). There, Wittgenstein's view of philosophy as conceptual analysis and Williamson's view of philosophy as substantial inquiry are compared.

From an historical point of view, the conception of philosophy as an activity devoted to the discovering of norms can be brought back to Wittgenstein and can be presented in the following way. Let us take—for instance—an uncontroversial result of epistemology as that expressed by (**a**) "Knowledge entails truth", or by the equivalent: (**b**) "Propositions that are known are true". Wittgenstein would refuse to describe these expressions as propositions stating necessary connections between entities (or properties?), i.e., the entity/the property Knowledge and the entity/the property Truth, as Williamson does; but he would neither present them in terms of propositions about the concept under inquiry, as Williamson thinks any friend of conceptual analysis would do. So, if **a** and **b** are not propositions stating necessary connections between entities (or properties), and neither propositions on concepts (i.e., conceptual truths), what are they?

In the perspective Wittgenstein defends, statements as **a** and **b** express rules/explicit instructions for the use of the words 'knowledge' and 'truth' (or, as he would also say, rules for the application of the concepts *knowledge* and *truth*); or better, rules that set connections among concepts/ among "X" and other words (among *X* and other concepts). Among, for instance, the notion of knowledge and the notion of truth; or, as in the case of the complete formulation of the JTB theory of knowledge, the notion of knowledge, on the one side, and that of belief, truth and justification, on the other. So, being norms, **a** and **b** could be formulated as "Call a proposition 'a piece of knowledge' only if you are prepared to call it a truth", or "Only apply the concept of knowledge to contents to which you are prepared to apply the concept of truth".

It is important to remark that what epistemologists discover, when they discover a rule of this kind, is nothing but a fact: they discover that, in the practices in which "know" and "truth" are used, speakers use these words in a way that leads them to conclude that they do not generally call a propositional content "knowledge" if they do not take it to be true. Like any other fact, this kind of fact can be found out, Wittgenstein says, by *observing* (or by *describing*) how in fact things go, and specifically how, as a matter of fact, we use the words "knowledge" and "truth".

Let us not delve immediately into the problematic question of presenting the results of philosophical inquiry (i.e., norms) as the simple result of the description of semantic facts, and go back for one moment to Williamson's argument: is it true that, in order to defend the armchair methodology of philosophy, we *must* reconsider the nature of the evidence obtained by thinking about the cases? The answer is no: rather than revising the classic idea on intuitive judgements, we could decide to renegotiate the nature of the aims and results of the philosophical inquiry. As I argued in the last section, this does not amount necessarily to a psychologization of the subject matter of philosophy. Further-

more, this solution (the third solution) is more viable than Williamson's, since it allows one to find an "easier" solution to the gap problem and does not require a revision of the most obvious and widely shared view on intuitive judgements, i.e., the view that sees them as the expression of what we would say/would think in a given situation. Namely, if the object of the philosophical inquiry are norms governing the use of "X" in the community, and if intuitive judgements are the expression of the competence, i.e., the capacity to use "X" in a manner that tends to reflect how the term is used in the community, then the appeal to them is justified.

7. Few final remarks

Obviously, saying that the third solution is preferable under the particular respect I have just outlined does not amount to saying that there are no problems to face. Indeed, this view has several well-known problems, especially in the "crude" version I provided before, i.e., a version in which it seems that only by observing and describing uses, something like a norm can be achieved.

For instance, a classic objection is the following: everyday linguistic behaviour includes errors and idiosyncratic uses; it is then difficult to say how, by observing and describing uses, or just by considering judgements that are descriptive of the way people use/would use a certain term "X", one could define a theory on X, i.e., a norm for the use of "X". What one obtains will be, eventually, a list of all the idiosyncrasies within the speakers' community. Furthermore, uses are often incoherent or inaccurate. So one can ask: how could they then be subsumed by a theory that, in order to be a theory, has to be consistent and accurate?

Finally, the idea that philosophical theses and theories are just the description of uses, or of judgements on uses, or, in a broader sense, the registration of the *received opinion* on X, seems to be plainly false. First of all, as Williamson points out, describing our uses or our intuitions is not what philosophers say they are doing when they investigate X: "many epistemologists study knowledge, not just the ordinary concept of knowledge. Metaphysicians studying the nature of identity over time ask how things persist, not how we think or say they persist" (Williamson 2007: 211). Moreover, describing what we say or think, or what we suppose is correct to say or think, is not what philosophers have done up to now: many (perhaps all) philosophical theories are corrective of our uses and competence. In particular, they are seen as—and in fact are—means to discriminate between what is really correct to say and what we just say or we just think is correct to say.

So, in order to make the third solution plausible, the question one has to answer is: how can an inquiry starting from judgements that are the product of one's competence lead to a theory that **(i)** has the characteristics a theory has to have in order to be a theory (consistency, accuracy) and that **(ii)** satisfies the expectations philosophers have on a theory on X, i.e., those expectations that are plausibly subsumed by claims as "we, philosophers, want a theory on X, not just on what we think or say that X is"?

The problems I have just mentioned go beyond the specific question raised in this paper. Therefore I will not discuss them in detail, but simply hint at two aspects I believe can play a central role in outlining a strategy for supporting

my view. The first one concerns the normative nature of intuitions. The second one the structure of the method which takes from the intuitions to the theory.

Let us start from the first aspect. Against the first battery of objections, one could insist on the fact that the evidence philosophers appeal to in order to build, support or attack theories is not directly usage, but the judgements of *competent* speakers. Moreover, these judgements do not consist in mere descriptions of the usage, or in previsions on how people would use the term in a given case, but in the description of what a competent speaker would take to be the *correct* usage. In other terms, intuitions are hypotheses of correctness, developed in the light of competence and reflection.

Let us then pass to the second aspect. Against the second battery of objections, one could point out the corrective and enhancing nature of the method leading from intuitions to the norm. Theories are not the results of a plain generalization from an initial set of intuitions. They are rather the result of an articulated reflective process, in which different theoretical hypotheses—obtained by inference to the best explanation (IBE) from the initial set of intuitions—are confronted with other intuitions which have not been taken into account before and which emerge from the reflection on factual or counter-factual (i.e., TE) cases. Usually, when a theory contradicts these intuitions, then the theory is modified or eventually abandoned. The same thing, however, can happen to intuitions. Namely, we may happen to find a certain theory especially persuasive and the way we normally would have judged the case changes: we start seeing things in the way the theory predicted.

In sum: the process taking the analyst from his or her intuitions to the norm can be ideally described as a procedure which, through IBE from an intuitive base and through corrections and improvements, aims to set a reflective equilibrium between the norm and the intuitions. Eventually, also other beliefs (for instance scientific ones) we think are relevant to the problem at issue can be used in this process.

However, if all this is true, then it is clear that the product of a process of this kind cannot be seen as the simple description of what we think or say about X, and neither as the simple description of what we believe is correct to say about X. The norm that we obtain is corrective and enhancing of what we think and say. In particular, it can be used to discriminate between what is correct and what we just thought or said was correct. And this—it can be argued—is enough to satisfy the expectations philosophers have when they ask for a theory on X, as opposed to a theory merely describing what we believe X to be.

References

Brown, J. 2011, "Thought Experiments, Intuitions and Philosophical Evidence", *Dialectica*, 65 (4), 493-516.

Brown, J. 2012, "Words, concepts, and epistemology", in Brown, J. and Gerken, M. (eds.), *Knowledge Ascriptions*, Oxford: Oxford University Press, 32-54.

Glock, H. 2010, "Discussion: from armchair to reality?", *Ratio*, 23, 339-48.

Marconi, D. 2011, "Wittgenstein and Williamson on Conceptual Analysis", in Davies, R. (ed.), *Annuario della Società Italiana di Filosofia Analitica* (SIFA) 2011, Milano: Mimesis, 91-102.

Williamson, T. 2007, *The Philosophy of Philosophy*, Malden, MA: Blackwell.

Williamson, T. 2011, "Philosophical Expertise and the Burden of Proof", *Metaphilosophy*, 42, 215-29.

Wittgenstein, L. 1974, *On Certainty*, Anscombe, G.E.M. and von Wright, G.H. (eds.), Oxford: Blackwell.

Wittgenstein, L. 2009, *Philosophical Investigations*, Revised 4th edition, Oxford: Wiley-Blackwell.

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

Undermining Defeat and Propositional Justification

Giacomo Melis

University of Glasgow

Abstract

I extend the Higher-Order View of Undermining Defeat (*HOVUD*) defended in Melis (2014) to account for the defeat of propositional justification. In doing so, I also clarify the important notion of higher-order commitment, and I make some considerations concerning the defeat of externalist epistemic warrants.

Keywords: Defeaters, Undermining, Overriding, Propositional Justification, Higher-Order.

One of the things that epistemologists worry about is the relation of epistemic support, in virtue of which something gives a subject reason to believe that a proposition is true. While the notion of epistemic support might be accounted for in a variety of ways, for the sake of simplicity in what follows I will assume that what gives a subject reason to believe that a proposition *p* is true is the possession of a piece of evidence that supports *p*. We can thus say that evidence for *p* is a justifier for *p*, in that it provides the agent with justification to believe that *p* is true.

If justifiers—evidence—provide the epistemic agent with justification, defeaters take justification away from her. Just like justifiers, defeaters have epistemic force, but it is a force that speaks against believing a proposition, rather than in favour of it. For the purposes of this paper, let us understand defeaters as pieces of counterevidence—evidence that speaks against believing a proposition.

Since Pollock (1974: 42-43), defeaters are commonly distinguished in at least two different kinds. Say that *p* is a previously justified proposition for a subject *S*: *overriding* defeaters give *S* a reason to believe not-*p*; *undermining* defeaters, on the other hand, give *S* merely a reason to give up *p*, without thereby giving a reason to believe not-*p*. The present article develops a previous contribution to understanding the way in which undermining defeaters work.

In Melis (2014), I defended a view according to which undermining defeaters require the subject to engage in some higher-order epistemic reasoning, while overriding defeaters do not. One limit of that account was that it applied only to a doxastic notion of justification and defeat. In this paper I extend the proposal

Argumenta 1, 2 (2016): 271-280
ISSN 2465-2334

© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.MEL

to propositional justification and defeat, and I elaborate the crucial notion of higher-order commitments. The plan is the following: first (§ 1) I will present the view already defended and (§ 1.1) I will provide some details about the notion of higher-order commitment, then (§ 2) I will briefly recall the distinction between propositional and doxastic justification and show why the account of undermining defeat presented does not apply to propositional justification. Finally (§ 3), I will extend the account to propositional justification.

1. The Higher-Order View of Undermining Defeat

Let us begin by considering some examples to illustrate the difference between undermining and overriding defeaters. In the propositional triads below, *e* represents the evidence that supports *p*, *p* the proposition the subject is justified in believing, and *d* the defeater.[1]

(1) *e* = <Adam says that Paul McCartney was the drummer of The Beatles>
p = <Paul McCartney played in the Beatles>
d = <Lauren tells me that Adam's knowledge in matters of pop music is poor>

(2) *e* = <I remember having left the book on the desk>
p = <The book is on the desk>
d = <I now see that the book is not on the desk>

(3) *e* = <[*S*'s apparent proof of *p*]>
p = <[A seemingly logical theorem]>
d = <A logician tells *S* that there is a mistake in the (apparent) proof>

(4) *e* = <All swans observed at t_1 are white>
p = <All swans are white>
d = <At t_2 a black swan is observed in Australia>

The defeater *d* in cases (2) and (4) explicitly suggests that not-*p*, while in cases (1) and (3) it is compatible with the truth of *p*. In (2) and (4) the defeaters are overriders, in (1) and (3) they are underminers.

Taking the inspiration from different remarks made by Scott Sturgeon (2014) and Albert Casullo (2003: 45-46), I defended the following Higher-Order View of Undermining Defeat (*HOVUD*):

> *(HOVUD)* Underminers suggest that something was wrong with the source of justification or with the justificatory process, and they operate their defeat by appealing to the higher-order commitment that the belief in question was based on that source or that process. If the suggestion is that the process, rather than the source, was defective, the defectiveness is to be understood as the occurrence of a mistake or some other disturbing event.[2]

In order to assess the explanatory power of the view, we need to go back to examples (1)-(4). Before doing that, however, I will make a few clarificatory remarks about some of the notions I have appealed to in (*HOVUD*).

[1] I do not mean to suggest that all evidence (or counter-evidence) is propositional, but only that for every piece of evidence, there is a proposition that can be used to represent it.
[2] A refinement of (*HOVUD*) proposed in fn. 16 of Melis 2014 has it that the suggestion made with respect to the source or the process can also be that they *might have been* defective.

By 'source of justification' I mean something of a rather large scale, like the agent's five senses, her memory, her reasoning abilities, or someone's testimony. Justificatory sources provide the evidence for *p*, and deliver justificatory processes. By 'justificatory process' I mean the activity that begins with the gathering of the evidence and that ends with the formation of the belief. The stage of evidence-gathering often involves interaction with things external to the mental life of the subject, but overall, justificatory processes are mental affairs internal to the agent.[3]

Higher-order commitments are potential or actual cognitive attitudes (*e.g.* beliefs, acceptances, presuppositions) of the agent towards propositions about the ways in which beliefs are justifiably formed, retained and abandoned—propositions that describe and sustain the justification of the agent's doxastic state (the set of the agent's beliefs) at a time *t*.[4] In general, every epistemically respectable change in the set of the agent's beliefs involves some higher-order commitments.[5] The propositions towards which the *relevant agent* is committed are such that she would take them to be true (or just warranted) on reflection—at least for as long as she stands by the related piece of justification.[6] We might say that, in a loose sense, the higher-order commitments are part of the agent's justification, in that the agent could resort to them to defend the epistemic worthiness of her belief, if questioned. However, the agent need not be aware that the relevant commitments are in place in order to form, retain, and abandon beliefs in an epistemically worthy manner.

Let us consider an example. When I come to believe that it is ten o'clock by looking at my watch, I thereby form a justified belief which is sustained by commitments towards propositions like 'the belief that it is ten o'clock was formed by looking at my watch', 'my watch is reliable', or 'I would have not trusted my watch if I had a good reason to believe it had stopped working', and so on. However, I do not need to appreciate that such commitments are in place (*i.e.* that I am committed to take those higher-order propositions as true or warranted)[7] in order to be justified in believing that it is ten o'clock. Still, if some exigent interlocutor were to push me to lay down my reasons to believe that it is ten o'clock, I would, maybe after some reflection, probably bring up some of the higher-order commitments.

We can now go back to (*HOVUD*), according to which the phenomenon of undermining defeat is articulated in two distinct parts. Firstly, the underminer needs a certain higher-order commitment about the source of justification or about the justificatory process to be in place; secondly, the underminer challeng-

[3] That is not to say that the subject needs to be aware of the mental activity involved in the justificatory process.

[4] In so far as such propositions overtly describe the epistemic activity of the subject—as opposed to merely implicitly expressing the subject's engagement in it—they are higher-order propositions.

[5] Exactly which ones depends on the details about the source of justification and justificatory processes involved.

[6] More on the rather important notion of relevant agent below in § 1.1.

[7] Details here depend on the specific cognitive attitude. For example, while believing would require taking the relevant higher-order propositions as true, accepting and presupposing might require only taking them as just warranted, or at any rate something which generally is not truth-guaranteeing.

es the epistemic worthiness of that very commitment. Let us consider the examples.

The commitment that has to be in place for the undermining defeat to be effective in example (1) is a commitment about the source of justification, the relevant proposition being: 'the belief that Paul McCartney played in the Beatles was based on Adam's testimony'. That the commitment has to be in place so that *d* can do its defeating job can be seen by noting that if I had taken my belief to have been justified by a different source—maybe someone else's testimony—the acquisition of the information that Adam's knowledge in matters of pop music is poor would not have had any defeating effect.[8] That the epistemic worthiness of the commitment is challenged by *d* can easily be seen by considering the very content of *d*, whose general suggestion is that the source of justification is not reliable with respect to the subject matter.

By contrast, nothing of the sort goes on in the overriding defeat involved in examples (2) and (4). In both cases, the defeater *d* could have done its defeating work just as well if the relevant agent had taken the beliefs to have been justified by any other source rather than memory (case 2), or inductive reasoning (case 4). Thus, there is no need for a higher-order commitment concerning the source of justification or the justificatory process[9] to be in place so that the defeat can be effective. Such examples suggest that overriding defeat, in general, does not rely on the presence of commitments about the way in which the relevant belief was formed or justified.

Case (3) is an example of undermining defeat in which the challenge raised by the underminer concerns the justificatory process rather than the justificatory source. In the case at hand, the justificatory source is provided by the agent's proving abilities, as it were. Such source delivers the alleged proof that *p*, whose execution on the subject's part constitutes the relevant justificatory process. The commitment that has to be in place and that is challenged in case (3) concerns the following proposition: 'the belief that *p* was based on the execution of that specific proof'. If the subject had taken her proving abilities to have delivered a different computation, *d* would not have had any undermining effect. The epistemic worthiness of the commitment is challenged because *d* suggests that something went wrong in that justificatory process.[10]

The preceding paragraph offers the occasion for an interesting reflection about the difference in the way underminers and overriders impinge on higher-

[8] This example should also help to clarify what it means for an underminer *to appeal* to some higher-order commitments: in general, an underminer *d* appeals to some higher-order commitment *c* when *d* could not do its defeating job, unless *c* was in place (that is, unless the relevant agent took *c* to be true or warranted on reflection).

[9] Since justificatory processes are delivered by justificatory sources, if a defeater is effective regardless of the source of justification, it is to be expected that it is effective also regardless of the specific justificatory process delivered by the source.

[10] One might think that the underminer challenges the epistemic worthiness of the source *via* the challenge raised to the justificatory process. That might happen in some cases, but it is easy to conceive of a scenario in which that is not what happens in example (3). Just suppose that the subject is generally good at proving theorems, and that she executed the calculation in optimal circumstances (she was sober, in a quiet room, etc.). Since the occasional mistake is compatible with the trustworthiness of the source, the challenge raised by *d* in such a construction of case (3) does not extend from the process to the source. See Melis (2014: §§ 2-3) for more details.

order reasoning. Indeed, there is a sense in which overriders suggest that something went wrong with the justificatory process too. Overriders suggest that the previously justified belief is false, and since a justificatory process that produces a false belief cannot be a successful one (believing what is true is one of the main epistemic goals), if the overrider is right, something must have gone wrong with the justificatory process. Thus, we have to acknowledge that overriders do have some implications in the realm of higher-order epistemic reasoning too. However, such an acknowledgement does not question the point made in (*HOVUD*).

The point made in (*HOVUD*) is that while the acceptance of an underminer (and the belief revision that follows) compels the agent to engage in some higher-order epistemic reasoning, the acceptance of an overrider does not. And that is compatible with the acceptance of overriders having some implications in the higher-order sphere. Let us have another quick look at the examples. The agent can accept the overrider in cases (2) and (4), and accordingly update her set of beliefs by replacing *p* with not-*p*, without considering any thought about the relevant justificatory sources of processes. Of course, if the agent is a very responsible epistemic agent, she might ask herself some questions about the epistemic worthiness of the sources or the processes. But she does not have to. On the other hand, in cases (1) and (3), the agent gives up the belief that *p* precisely as a result of the emergence of doubts about the trustworthiness of the justificatory sources or processes that have been raised by the relevant underminer.[11]

Summing up, (*HOVUD*) does two things: on one hand it provides an account of how undermining defeat works; on the other, it explains the difference between underminers and overriders in terms of the impact they have on the sphere of higher-order reasoning. Contrary to what underminers do, overriders do not need to appeal to higher-order commitments about how the belief was formed, and do not need to challenge the epistemic worthiness of those commitments. If so, underminers force the subject to reflect about the ways in which her beliefs were formed, and thus cut at a deeper epistemic level than overriders do: unreflective agents can suffer overriding defeat, but they cannot suffer undermining defeat.[12]

[11] In Melis (2014: § 3), I gave more emphasis to another difference between underminers and overriders at this junction: while the suggestion made by overriders with respect to the faultiness of the justificatory process is compatible with the correct execution of the process—it is in the nature of justificatory processes delivered by fallible sources of justification that sometimes they fail to lead to truth—the suggestion made by the underminers is not. In other words, while overriders merely suggest that (as sometimes happens) the process failed to lead to truth, underminers suggest that a specific disturbing event has caused the process to fail. This point, reflected in the last clause mentioned in (*HOVUD*), is not crucial for the aim of extending (*HOVUD*) to propositional justification, and limits of space advice against expanding further on it.

[12] One might think that there might be cases in which defeaters work as both underminers and overriders. I do consider the possibility of such cases in Melis (2014: § 5).

1.1 More on Higher-Order Commitments[13]

Before moving on, I wish to clarify when exactly a higher-order commitment is in place. To that end, let us imagine that the agent has some misconceptions about the way in which she formed her belief. Say that the agent justifiably believes a proposition *p*, and that she explicitly believes (again, justifiably so) that she based her belief on source *X* while in fact she based it on source *Y*. In such a case, we can think of two kinds of higher-order commitments: those concerning propositions about the actual source *Y*—such as "the belief that *p* was based on source *Y*", or "source *Y* is reliable", etc.—and those concerning propositions about the putative source *X*—such as "the belief that *p* was based on source *X*", or "source *X* is reliable", etc. Are both kinds of commitments in place? Yes. For every justification, actual or merely putative, there are some higher-order commitments in place, and (*HOVUD*) is a proposal which is meant to account for the defeat of whichever justification (either actual or putative) is under scrutiny. But let us reflect on the case just described with some care.

One might raise the following worry. Since the agent described is completely unaware that she formed her belief in *p* on the basis of source *Y*, she would not be able to take a higher-order proposition like "source *Y* is reliable" to be true (or warranted) on reflection. How can the commitments about the actual source *Y* be in place then? It is at this juncture that the emphasis on the *relevant agent* becomes important. While in many ordinary circumstances the agent that would take the commitments to be in place is the actual agent, in cases where the actual agent has some misconceptions on the way she formed her beliefs (or simply ignores it without being in the position to figure it out—see next paragraph and fn. 14), the agent that would take a given commitment to be in place varies depending on the commitments (and the related justification) in question. If we consider the commitments about the *putative way* in which the belief was justified, then the relevant agent is the actual agent. In the case mentioned, it is the actual agent that, on reflection, would take the higher-order proposition "the belief that *p* was based on source *X*" or "source *X* is reliable" to be true or warranted. However, if we consider the commitments about the *actual way* in which the belief was justified, then the relevant agent is the idealized agent who is like the actual agent in all respects, except that she has no misconceptions about the way in which the belief was formed. Of course, in the case described the actual agent might not be able to appreciate the defeat of what I have referred to as 'actual justification', but that does not mean that that justification is not defeated—in an externalist sense. And (*HOVUD*) has the tools to account for that defeat.

The relation between (*HOVUD*) and the defeat of externalist justification can be seen clearly by considering the following similar worry. Suppose that the subject *S* is (non-culpably) unaware of the fact that she has formed her belief in *p* on the basis of source *X*, and that the defeater, although it does nothing to show that *X* is the source of *S*'s belief, challenges the epistemic worthiness of *X*. Does not (*HOVUD*) in this case predict defeat when intuitively none occurs? It is true that (*HOVUD*) predicts that, in the case described, there is undermining defeat. I stand by that. I also acknowledge that there is an intuitive sense in which there is no defeat, but the intuitions at play here involve an internalist conception of

[13] Thanks to two anonymous referees for alerting me to the issues discussed in this section.

epistemic justification, according to which a subject can only be justified if she is aware of the relevant justifying factors, as it were. In the case at hand, on an internalist conception of justification, *S* is not justified in believing that *p on the basis of source X* because *S* is not aware that she based her belief on source *X*. And, the objection goes, if there is no justification, there is no defeat. I have no quibbles with that. However, what the objection fails to mention is that, in an externalist sense—which does not require the subject to be aware of the relevant justifying factors—*S is* justified in believing that *p on the basis of source X* (assuming the conditions on the source posed by the relevant externalist account of justification are met). And such externalist justification is precisely the one that (*HOVUD*) predicts gets defeated.[14]

The lesson that we should draw from the cases illustrated in this section is that (*HOVUD*) is neutral with respect to at least some of the dimensions along which epistemic justification can vary. (*HOVUD*) is a theory of undermining defeat that explains how undermining defeat works in either internalist or externalist cases, and regardless of whether the justification in question is actual or putative. Let us now turn to the extension of (*HOVUD*) to propositional justification.

2. The Limit to Doxastic Justification

Before we see why (*HOVUD*) is limited to doxastic justification, and why that should be a reason of concern, let us recall the difference between propositional and doxastic justification.

It is common to account for the distinction between propositional and doxastic justification along the following lines: when a subject has a reason or evidence to believe that *p*, she has propositional justification *to* believe that *p*; when a subject has a reason or evidence to believe that *p*, *and* bases her belief that *p* on those reasons or that evidence, she is doxastically justified *in* believing that *p*.[15] If a subject is doxastically justified in believing a proposition, she also has propositional justification to believe that proposition, but not vice versa: in other words, doxastic justification entails propositional justification, but not vice versa.

Here is a quick example: an agent that justifiedly believes that Cagliari is in Sardinia, and that Sardinia is in Italy, has propositional justification *to* believe that Cagliari is in Italy, but it is when the agent (competently) draws the relevant inference that she is (doxastically) justified *in* believing that Cagliari is in Sardinia.

To see why (*HOVUD*) is limited to doxastic justification, we need to understand that while doxastic justification requires that the proposition justified is believed by the agent, propositional justification does not. Recall that according to (*HOVUD*), underminers work by appealing to some higher-order commitments about the way in which the relevant belief was originally formed. But propositional justification does not require belief formation, and thus does not require

[14] To put the point in terms of the relevant agent, we might say that the higher-order commitments involved concern propositions that an *idealized agent*—alike the actual agent in all respects, except that she does not ignore how the belief was formed—would take to be true or warranted.

[15] See Feldman (2002: 46) and Pollock and Cruz (1999: 35-36) for just two examples of this way of drawing the distinction.

any commitments about belief-formation. So, (*HOVUD*) does not apply to propositional justification.

To see why the advocate of (*HOVUD*) should be concerned by this limitation, consider the following variation of example (1). Suppose that Adam tells me many things and that, before I have a chance to consider them and update my set of beliefs, Lauren tells me that Adam is a compulsive liar. In such a case, I obtain an underminer for the justification I get from Adam's testimony, before I have a chance to form the beliefs whose higher-order commitments (*HOVUD*) predicts that the underminer calls into question. Thus, (*HOVUD*) cannot account for such a case. And the reason is that the underminer in question affects the propositional justification to believe propositions before the agent actually forms any belief on the basis of that justification.

The lesson seems to be that the phenomenon of defeat arises with respect to propositional justification. If so, any account that, like (*HOVUD*), is tied to doxastic justification does not generalize far enough.[16]

3. Extending (*HOVUD*) to Propositional Justification

In this section I will argue that (*HOVUD*) can be extended to propositional justification in virtue of the close relationship that exists between propositional and doxastic justification. In a nutshell, propositional justification involves higher-order commitments analogous to those involved in doxastic justification, and those commitments are what underminers call into question in the defeat of propositional justification.

But let us begin with the relationship between propositional and doxastic justification. We said that a subject that bases her belief in p on some evidence e is doxastically justified in believing p. However, the justificatory process that brings the subject to believe p on the basis of e is grounded in a relation between e and p that is in place regardless of whether the subject uses it to her avail. That relation of support between e and p is propositional justification. In this sense, doxastic justification arises from propositional justification.

Let me briefly expand on this. Consider the set of the propositions that constitute the subject's evidence. Those propositions epistemically support a number of other propositions, to which they are related in propositional justification. The use of that epistemic link to form a justified belief (doxastic justification) is available to the subject since the moment in which she acquires the evidence, but it might take a while before she follows that link to form an actual belief, if ever. Justificatory processes proceed along that pre-existing epistemic link and lead the subject to doxastic justification. What matters for our purposes is that the relations of (propositional) justification between the propositions in the set of the evidence and those in the set of what is supported by the evidence involve higher-order commitments. Of course, they are not commitments about actual belief-formation, but rather commitments about the ways in which e supports p.[17] Those are the commitments that are challenged by undermining defeaters.

Let us consider the problematic case for (*HOVUD*) mentioned in the previous section. The (propositional) justification to believe the various propositions

[16] Thanks to Declan Smithies for bringing this issue to my attention in correspondence.

[17] They concern the source of justification, the acquisition of the evidence e provided by the source, and the relation of support between e and p.

that Adam is testifying involves the commitment towards the proposition that the source of justification is Adam's testimony. Such commitment does not concern belief-formation, but only the support relation that exists between the propositions that work as justifiers (propositions of the form 'Adam says that *p*') and the propositions justified (the various corresponding *p*'s). For the acquisition of the information that Adam is a compulsive liar to undermine the (propositional) justification to believe what Adam says, that commitment has got to be in place: should the relevant agent take the propositional justifications in question to be due to Peter's testimony rather than Adam's, the information that Adam is not to be trusted would not have any undermining effect. Moreover, it is the epistemic worthiness of that very commitment that is called into question by the undermining defeater, which suggests that Adam is not a reliable source. This is all very much in the spirit of (*HOVUD*).

Here is a formulation of a version of (*HOVUD*) that applies to propositional justification.

> *(HOVUD-prop)* Underminers suggest that something was wrong with the source of justification or with the grounds that support *p*, and they operate their defeat by appealing to the higher-order commitment that *p* is supported by that source or those grounds. If the suggestion is that something was wrong with the grounds, rather than with the source, the defectiveness is to be understood as the occurrence of a mistake or some other disturbing event that spoiled the epistemic worthiness of the grounds.

Of course, the main difference between (*HOVUD*) and (*HOVUD-prop*) is that the former talks about commitments concerning the formation of a justified belief, and the latter talks about commitments concerning the formation of a justification which need not result in belief formation. Just like in the case of doxastic justification, the commitments involved in propositional justification concern propositions that the relevant agent would take to be true on reflection (as long as she sticks to the corresponding justification). However, the agent that would take the higher-order propositions to be true or warranted on reflection is, once again, an idealized one: it is the idealized agent that, on the basis of the same body of evidence available to the actual agent, goes on to form all the beliefs that are supported by that body of evidence (including those that, for whatever reason, the actual agent fails to form).

A second difference worth noticing is that the notion of justificatory process included in (*HOVUD*) has been substituted with that of grounds in (*HOVUD-prop*). The reason is that justificatory processes are largely mental and have to do with belief-formation, and thus cannot play a role in (*HOVUD-prop*). However, propositional justification offers something that underlies the justificatory process involved in doxastic justification, and I have called that 'the grounds for *p*'. I understand grounds as a non-mental analogue of justificatory processes: the pre-existing epistemic path that goes from the subject's acquisition of *e* to the conferral of the positive epistemic status to *p*. Such a notion of grounds can thus refer both to the acquisition of the evidence (from which the epistemic path begins) and to the relation between the evidence and the proposition supported *p* (which constitute the remaining of the epistemic path). Just like justificatory processes, justificatory grounds have their origin in a source of justification.

Let us now briefly consider the previous examples of undermining defeat and suppose that the justification in question is propositional rather than doxastic. (*HOVUD-prop*) works in a way just parallel to (*HOVUD*). In (1) the underminer suggests that the source was not reliable with respect to the subject matter, and the higher-order commitment under attack is that the (propositional) justification to believe that Paul McCartney played in the Beatles has its source in Adam's testimony. In example (3) the underminer suggests that something was wrong with the grounds, and the commitment under attack is that the (propositional) justification to believe *p* is given by that alleged proof (as opposed to *the agent's execution* of that alleged proof, which constitutes the justificatory process).

4. Conclusion

To conclude, I acknowledge that, in an important sense, the phenomenon of defeat arises at the propositional level. Since the phenomenon of defeat concerns justification, and, at least in the way explained above, propositional justification comes before doxastic justification, it is no surprise that the phenomenon of defeat concerns primarily propositional justification. Yet, as I hope to have shown, (*HOVUD*) can easily be extended to account for the defeat of propositional justification. More generally, (*HOVUD*) promises to have the tools to account for undermining defeat regardless of several of the dimensions along which justification can vary: propositional vs. doxastic, actual vs. putative, externalist vs. internalist.[18]

References

Casullo, A. 2003, *A Priori Justification*, Oxford: Oxford University Press.

Feldman, R. 2002, *Epistemology*, Upper Saddle River, NJ: Prentice Hall.

Melis, G. 2014, "Understanding Undermining Defeat", *Philosophical Studies*, 170, 433-42.

Pollock, J. 1974, *Knowledge and Justification*, Princeton: Princeton University Press.

Pollock, J. and Cruz, J. 1999, *Contemporary Theories of Knowledge*, 2nd ed., New York: Rowan & Littlefield.

Sturgeon, S. 2014, "Pollock on Defeasible Reasons", *Philosophical Studies*, 169, 105-18.

[18] I would like to thank Declan Smithies and Carrie Jenkins for very helpful discussions, and everyone who attended the Philosophy Work In Progress seminar at the University of Aberdeen in February 2014, where an early version of this paper was presented. I am especially grateful to two anonymous referees for raising several precious questions, which led to a substantial improvement of the article.

Analytic Philosophy and European Culture[1]

Umberto Eco

University of Bologna

From among the problems posed in the letter of invitation that Marconi sent to us,[2] I have picked out one that is twofold, that seems to be the same thing twice, but is not quite. One aspect is the question of why there is a continental philosophy that drums up the interest of a broad public, while, at least in the United States, professional analytic philosophy passes almost unnoticed; the other concerns the relation between "professional" philosophy and what has come to be known as "edifying" philosophy. It seems to me that these are two very different issues that are intertwined with each other in a range of phenomena that we have been witnessing and of which the Modena phenomenon is a particularly salient case.[3]

Perhaps we are forgetting that we are living in a huge international supermarket in which much more philosophy is sold than once upon a time; but also much more relativity theory is sold and much more Bible study is sold: more of everything is sold. And the key recurring question is this: Is this all to do with the popularisation of philosophy? Is it to do with De Crescenzo?[4] I think not, because he does not popularise philosophy: he popularises Diogenes Laertius' anecdotes about the lives of philosophers, and he says hardly a word about their thought. Is it to do with what the newspapers manage to publish? Just the day

[1] By kind permission of the estate. Translation and notes regarding the Italian scene by Richard Davies.
This is a talk for the Round Table that rounded off the 6th National Conference of the Italian Society for Analytic Philosophy (SIFA), Genoa 2004. Given the subject of Eco's talk, it is worth noting that the Round Table was held in the Carlo Felice opera house and, perhaps because of the presence of Eco, there was standing room only. The conference to which the event was annexed was held in the classrooms of the Philosophy Department of the University of Genoa in via Balbi.

[2] Diego Marconi is Professor of Philosophy of Language at the University of Turin, a founder member of SIFA and co-panellist with Eco.

[3] Since 2001, the cities of Modena, Carpi and Sassuolo in Emilia-Romagna have organised thematic festivals of philosophy lasting several days, with Italian and international speakers, regularly attended by more than 100,000 people.

[4] Luciano De Crescenzo, a Neapolitan engineer turned writer and television personality, has written numerous books on ancient philosophy and mythology, which have sold 18 million copies worldwide, nearly half of which in Italy.

Argumenta 1, 2 (2016): 281-285
ISSN 2465-2334
© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.ECO

before yesterday, the *Corriere della Sera*[5] carried an opinion piece by Emanuele Severino, who is not a man to lend himself to popularisation,[6] but a newspaper may publish him, too. Is it then to do with the famous cafés of the Bastille, where I've never been but where I'm told that edifying philosophy is sold?

What we have to do with here is a massive reaction to philosophical problems by a public of the sort that goes to Modena, and it does no harm to recall that this was under way almost twenty years ago, when the public library in Cattolica[7] started organising meetings on "What philosophers are doing today": between 1,500 and 2,000 people coughed up and turned up. So Modena is just the latest in a series of phenomena that have become everyday business. Indeed, next Monday, I am down to speak about Bobbio,[8] and they sold out 1,200 tickets in an hour.

Now, what is going on here? We're all supposed to know the answer. It's all down to the collapse of the grand narratives and of the old ideologies. So there are droves of people who want to go and hear that sort of stuff where it is still talked about, given that no one is inclined to talk that way in institutions such as political parties. And among the youth, there is a rebellion against the trivialisation in the mass media: they cannot put up with evenings in front of the television seven days a week, so they say, "I'm off to listen to people talking about philosophy, for crying out loud". Likewise, in the period since I was at University, when 100 presences at a conference was an event, the world population has burgeoned from 2 billion to 6 billion. We have to keep track also of the numbers here. I find nothing to be surprised by or to be worried about, but on the contrary something to be comforted by, if 2,000 people go to listen to philosophers talking among themselves.

Rather, the key question seems to me to shift the terms of the debate. Is philosophy an exclusively technical discipline or is there also a philosophy that I hesitate to call "popular"? Is there a "folk philosophy" as there is a "folk psychology"? For sure there is a folk psychology, which is studied by psychologists because it is the way that normal folk interact with the world. Likewise, there is a philosophy that might be called "folk" or a "popular philosophy" of the sort that my taxi driver or my grocer might do.

So we move onto the second theme: what is philosophy about? Philosophy is about questions that have no answer or, to put it less drastically, that do not have a single verifiable answer. Once a question is answered, it is no longer philosophical. For Epicurus, it was a philosophical question how large the Sun is, whether it is larger than we see it or not. Epicurus said that it was about a foot across. But as soon as it was established that it is more that a foot across, the question was no longer philosophical. At the worst, and we might find someone in internet who has posed it in these terms, the philosophical question might be "why is the Sun not a foot across?". This is an interesting philosophical

[5] The *Corriere della Sera*, founded in 1876, is Italy's establishment broadsheet.

[6] Emanuele Severino, Emeritus Professor of Philosophy at the University of Venice, is often associated with an attempt to revive the thought of Parmenides of Elea; his prose is challenging also for Italian native speakers.

[7] Cattolica is a seaside resort on the Adriatic coast, with a population of about 17,000.

[8] Norberto Bobbio (1909-2004), Professor of Philosophy of Law at the University of Turin and Life Senator of the Italian Republic, was a prolific and respected commentator on social and political questions.

question that does not, after all, have an answer and that, if you think about it, is just a homespun instance of the question of why there is something rather than nothing.

Now both professional and popular philosophy have to do with these questions, except that there are questions that cannot be posed in popular terms and that can be formulated only using technical terms. For instance, the yokel of Anaximander's day who was watching over his field and asking himself "where the devil does all this stuff come from?" was asking the same question that Anaximander raised. In other words, this is a question that an ordinary person might ask: "Is there a God who created the world or not?" or "To what extent do I have a right to tell lies?". This is popular philosophy, and, if I carry on, I'll end up re-writing Augustine's treatises against lying, such as the *De Mendacio*. But it is a question that my next-door neighbour understands: "What does it mean for something to be true?". This is popular philosophy.

But there is no answer that can be formulated as popular philosophy to the question of what it is for something to be. It is even difficult to explain what is meant by the question "What is being?". Indeed, there are no popular terms to explain what is meant, to pick an example dear to analytic philosophers, by "rigid designation" because the popular terms would be: "Oh! that fellow there is Diego Marconi"; but that is not the problem of rigid designation.

What we find is that, if you look into a history of philosophy,[9] three quarters of the philosophy that has been done in the last 2,500 years is not technical philosophy. You may recall that Ermete Zacconi[10] and Ruggero Ruggeri[11] acted out a good number of Plato's dialogues. You can go and describe the *Phædo*, the *Phædrus* or the *Crito* to your grocer and he will be deeply moved.[12] But try relating the *Parmenides* to him. We haven't understood it ourselves. Santambrogio[13] will recall that not so many years ago, we devoted a seminar running over two years to reading and re-reading the *Parmenides*, and all we came away with was a headache. You can't recount the *Sophist* to your grocer and perhaps not even the *Meno*, given that the geometrical demonstration calls for some effort. In these ways, therefore, you can see that, even within the works of a single philosopher, there are parts that are popular or popularisable, and there are parts about which there is nothing to be done.

The diagnosis of non-being in the *Sophist* is a technical problem. And St Thomas Aquinas' *Reflexio ad Phantasmata* is a technical problem: don't try and explain it to just anyone. Even the Five Ways of demonstrating the existence of God might count as popular philosophy because, irrespective of whether they are good arguments or not, everyone understands them. As for Aristotle, the whole of the *Politics* and the whole of the *Ethics* are popular philosophy; likewise

[9] Eco probably has in mind the three-volume review-anthologies that are produced in Italy for the philosophy courses of the last three years of high school instruction, one volume per year in chronological order, starting with the precursors of Thales of Miletus.

[10] Ermete Zacconi (1857-1948) took the leading role in a film based on Plato's versions of the trial and death of Socrates in 1939.

[11] Ruggero Ruggeri (1871-1953) took the leading role in the mise en scène of the *Phædo* and the *Crito* in Brescia and Milan in 1950.

[12] Interestingly, Eco does not mention the television film version by Marco Ferreri (1988) of the *Symposium* with Irene Papas in the role of Diotima.

[13] Marco Santambrogio was Professor of Philosophy of Language at the University of Parma, a founder member of SIFA, and a co-panellist with Eco.

most of the *Poetics* and the first part of the *Rhetoric*, which gets harder as it goes on. But if you try to explain book Zeta of the *Metaphysics* to your grocer, you'll get nowhere. And even the *De Interpretatione* … well, until just the other day, we thought we understood it and then Lo Piparo comes along and tells us that we haven't understood a blind thing and we must start over.[14]

A good deal of Descartes is good popular philosophy. The whole of Locke's *Essay* is the upshot of table talk and it lends itself to being read. Fontenelle explains the plurality of worlds to ladies; but all the same he is a bit of a chauvinist and at the end says "but, for all the world, I don't want to trouble your little head any further, and you should be getting on with something else". Yet, as the book progresses, he does *philosophie pour dames*, as well as *philosophie pour messieurs*. Spinoza: the first half of the *Ethics* is technical, but the whole of the theory of the passions is good popular philosophy: you can get your grocer to read Spinoza on jealousy and you can bet that he will be carried away with it.

But there is nothing to be done with Zeno of Elea: he is technical. And, indeed, all that De Crescenzo says about him is that he was gay. Perhaps there are charlatans who could do good popular philosophy but who pretend to do technical philosophy. In exchange for an outrageous sum of money, I think I could relay in popular terms almost everything in Heidegger's *Being and Time*: inauthenticity, happiness, small talk, being-toward-death, and so on. This can be got across to your grocer. But he made heavy weather of it because he wanted to seduce Hannah Arendt, who fell for it like a sucker.

Are there popular sides to analytic philosophy? With Quine there's no way, nor even with Putnam. But Grice and Austin can easily be done. Is it true that Americans do not follow these things? For sure, there is nothing like Modena, but when it comes to critical discussions, the Y Center in New York pays out handsomely and sells out six months in advance for 1,500 to go and hear cultural debates.

What, rather, is analytic philosophy's problem? One has already emerged from the discussion, with a touch of hubris that was straightaway subject to severe self-criticism. This s is the taste, the need or the noble aspiration to feel oneself a specialist who works on an extremely limited range of problems and aims to resolve marginal puzzles. Then I happened once to talk with Putnam about the Mutakallimun (otherwise known as Motocallemin), an Arabic sect known to only a few Medieval scholars, and he knew everything about them. But he would not allow his knowing to be known, because it would not be serious, in line with the American campus ethic. He would rather keep it hush hush, and yet he had read about the Motocallemin, who held a curious view not unlike Malebranche's occasionalism, but with an Islamic flavour.

And then there is the other problem, as Marco Santambrogio has noted, that cuts the analytical philosopher off from the broader public, which is the divide between theoretical philosophy and history of philosophy. Gilbert Harman once put a notice on his office door in Princeton's Philosophy Department, saying, "History of Philosophy: Just Say No!". When we Europeans talk about rigid designation, we do what we can to find points of contact with Anselm's account of paronymy. Not in America, though. This is an obstacle to communica-

[14] Francesco Lo Piparo is Professor of Philosophy of Language at the University of Palermo. Eco is referring to his *Aristotele e il linguaggio. Cosa fa di una lingua una lingua*, Laterza, Rome-Bari, 2003.

tion with a much broader cultural network that could strike and stimulate interest in the public at large. This is the real vice, not so much of European analytical philosophers, who know their history of philosophy well enough, as of their American cousins, who sometimes risk reinventing the wheel because they are in the dark about how, for instance, the Stoics had long since cleared up a certain question or shown it to be unsolvable.

And here we come to the heart of the problem that accompanies the well-known "town-and-gown" conflict from Oxford and Cambridge, England to Cambridge Massachusetts, and that is very marked and cuts thinkers off from the society that surrounds them. This is something that continental philosophy gets right: it is practiced in cities and in light of the problems that they face. But, above all, this is a practical division from the wider culture and so is less a philosophical problem than a sociological problem. In my view, what upsets certain analytical philosophers who pose the question "why don't others listen to what we have to say?" is a merely sociological question.

One last thing. I had a teacher by the name of Giovanni Cairola, who died very young.[15] He was an assistant of Abbagnano[16] and worked on Duns Scotus. When I had finished the second year, he died of a stroke while out sailing. I was very fond of him and, though he published relatively little, he once gave a talk contesting Boethius, of which I remember only the title: "Philosophy does not console". It seems to me that, relative to the problems of technical against non-technical philosophy, of philosophy as edification, of education to responsibility, of the growth of applied philosophy, of the status of embryos and so on, the point to hang on to is this: if we let people believe that philosophy is a source of consolation, then we are doing psychoanalysis without even the analysis, and we are selling spurious knowledge. If we manage to get across the idea that philosophy does not console precisely because it tells us that, for some questions, there is no one answer, then philosophy will still have a certain educative role to play.

[15] Giovanni Cairola died in 1954. His article "La filosofia non consola" appeared in the review *aut aut* in 1952 (N° 10, pp. 332-37).

[16] Nicola Abbagnano (1901-1990) was for many years Professor of History of Philosophy at the University of Turin, where Eco was among his students. His eight-volume *History of Philosophy* (1946) and his *Dictionary of Philosophy* (1960), both many times re-edited, expanded and reprinted, are standard reference works in Italian.

Filosofia analitica e cultura europea[1]

Umberto Eco

Università di Bologna

Nella lista di problemi che ho trovato nella lettera che Marconi ha mandato a tutti noi, ne identifico uno duplice, che sembra che si parli della stessa cosa, ma non è la stessa cosa. Uno è la domanda del perché c'è una filosofia continentale che riscuote gli interessi del grande pubblico mentre una filosofia analitico-professionale, almeno negli Stati Uniti, non riscuoterebbe l'interesse del pubblico, e l'altro è il rapporto tra filosofia *professionale* e filosofia *edificante*. Mi pare che siano due problemi molto diversi che vanno a innervarsi in una serie di fenomeni ai quali assistiamo e di cui, è già stato detto, il fenomeno di Modena è un fenomeno particolarmente importante.

Ci stiamo dimenticando che viviamo in un grande supermercato globale in cui si vende molto più filosofia di un tempo, ma si vende molto più teoria della relatività, si vende molto di più di studi biblici, si vende di più di tutto. E il nodo di problemi che ci si ripresenta è: stiamo parlando di filosofia divulgativa? De Crescenzo? No, perché De Crescenzo non divulga la filosofia; divulga Diogene Laerzio, cioè gli aneddoti sulla vita dei filosofi; sul loro pensiero non dice una parola. Stiamo parlando del fatto che i quotidiani possono pubblicare? Diego, l'altro ieri sul *Corriere* c'era un elzeviro di Severino, che non è uomo che indulga alla facile divulgazione, però un giornale può pubblicare pure lui. Stiamo parlando dei famosi caffè della Bastille, dove non sono mai andato ma dove credo vendano filosofia edificante?

Stiamo parlando della risposta massiccia ai problemi filosofici da parte del pubblico tipo Modena, e io vorrei ricordare che la cosa è incominciata quasi vent'anni fa quando alla biblioteca di Cattolica sono iniziati gli incontri "Cosa fanno oggi i filosofi": c'erano 1500-2000 persone che pagavano e venivano. Quindi il fenomeno Modena è solo l'ultima coda di un qualcosa che avviene ogni giorno: lunedì io debbo andare a parlare su Bobbio; in un'ora hanno fatto fuori 1200 biglietti. Ci sono 1200 persone [che vogliono assistere]. Ora, questo a che cosa è dovuto? Lo sappiamo tutti: il grande crollo delle grandi narrazioni e

[1] Per gentile concessione degli eredi.
Questa è la trascrizione di un intervento di Umberto Eco a una tavola rotonda organizzata per il VI Convegno della Società Italiana di Filosofia Analitica, tenuto nel 2004, anno in cui Genova era capitale europea della cultura. Il convegno (con circa 200 relazioni) ebbe luogo presso la Facoltà di Lettere e Filosofia in via Balbi, mentre la tavola rotonda si svolse all'Auditorium Montale del teatro Carlo Felice.

Argumenta 1, 2 (2016): 287-290
ISSN 2465-2334
© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.ECO

delle ideologie, e allora c'è un vasto pubblico che vuole andare a sentire qualcosa là dove lo dicono, visto che nelle sedi istituzionali, partiti o altro, non glielo dicono più.

Uno, direi da parte dei giovani, è la reazione a una banalizzazione dei mass media: non è sopportabile per sette giorni una serata televisiva, quindi "vado a sentirmi gente che parla di filosofia, santiddio".

L'altro è che noi non consideriamo che dai tempi della mia Università, quando se andavano 100 persone a un convegno era una cosa straordinaria, oggi siamo passati da 2 miliardi a 6 miliardi. Mettiamo in conto anche gli elementi quantitativi. Io non trovo niente di stupefacente né di preoccupante, se mai incoraggiante, perbacco, che 2000 persone vadano a sentire una discussione di filosofi.

Piuttosto il punto nodale mi pare un altro: la filosofia è un discorso esclusivamente tecnico o esiste una filosofia che non oso chiamare "popolare" – c'è una *folk philosophy* come c'è una *folk psychology*. C'è una *folk psychology* che gli psicologi studiano perché è il modo in cui la gente normale reagisce al mondo. C'è una – per non dire sempre "folk", che a pronunciarlo bene poi ci si confonde con un altro termine, parliamo pure di – "filosofia popolare", la filosofia che può fare il mio taxista o il mio droghiere.

Anzitutto, secondo tema: di cosa si occupa la filosofia. La filosofia si occupa solo di porre le questioni che non hanno risposta, o, per rendere meno radicale l'affermazione, che non hanno una risposta unica verificabile. Non appena una questione trova risposta, esce dalla filosofia. Era questione filosofica in Epicuro l'apparente *magnitudo solis*: se il sole fosse più grande di quello che vediamo o no. Epicuro diceva che era largo 30 centimetri. Nel momento in cui si è appurato che il sole ha un diametro più grande di 30 centimetri, la questione è uscita dalla filosofia. Casomai la questione filosofica – andando a guardare su Internet magari si trova già qualcuno che [ne] ha scritto – sarebbe "perché il sole non è largo 30 centimetri?". Questa è un'interessante questione filosofica che in effetti non ha risposta e che, se ci pensate bene, è una forma volgarizzata della questione perché c'è dell'essere piuttosto che il nulla.

Ora, le due filosofie, quella professionale e quella popolare, hanno a che fare tutte con queste questioni, salvo che ci sono delle questioni che a livello popolare non possono essere formulate e che possono essere formulate solo a livello tecnico.

Voglio dire: il contadino dei tempi di Anassimandro che stava seduto a guardare il campo e diceva "ma, da dove diavolo viene fuori tutta 'sta roba?" si poneva una questione che era la stessa di Anassimandro. In [altri] termini è questa la questione che si pone la persona normale: "Ma insomma, esiste Dio che ha fatto il mondo oppure no?"; "Sino a che punto ho diritto di mentire?". Questa è filosofia popolare (se continuo ci scrivo l'intero *De Mendacio*). Però è una questione che capisce anche il mio vicino di casa. "Ma cosa vuol dire che una cosa è vera?". Questa è filosofia popolare.

Dire che cosa è l'essere è una questione che a livello popolare non ha risposta, non può essere formulata. È difficile spiegare a uno qual è la questione "cos'è l'essere". Beh, per farvi contenti, perfino la designazione rigida non può essere esposta in termini popolari, perché in termini popolari sarebbe "Ah, Diego Marconi è quello lì" (ma non è questo il problema della designazione rigida).

Allora avviene che, se prendete in mano una storia della filosofia, i tre quarti della filosofia filosofata negli ultimi 2500 anni sono filosofia popolare,

non filosofia tecnica. Voi sapete che buona parte dei dialoghi platonici li recitavano a teatro Ermete Zacconi e Ruggero Ruggeri. Potete andare a raccontare il *Fedone*, il *Fedro*, il *Critone* al vostro droghiere e ne viene profondamente colpito. Provate a andargli a raccontare il *Parmenide*. Non l'abbiamo capito neanche noi. Santambrogio si ricorda che ancora tre o quattro anni fa abbiamo dedicato due anni di seminario alla lettura e rilettura del *Parmenide* di Platone e ne siamo usciti con la testa in fiamme. Non potete andare a raccontare al droghiere il *Sofista*; forse neanche il *Menone*, perché la dimostrazione matematica impone qualche sforzo; quindi vedete come all'interno di uno stesso filosofo c'è un'ampia zona di filosofia popolare – popolarizzabile – e un'altra zona che lì non c'è niente da fare.

La diagnosi del *Sofista* è un problema tecnico. La *Reflexio ad Phantasmata* di san Tommaso è un problema tecnico: non lo andate a spiegare. Anche le cinque vie per la dimostrazione di Dio sono filosofia popolare, che poi – siano buone o cattive – le capiscono tutti. Aristotele: tutta la *Politica* e tutta l'*Etica* è filosofia popolare; quasi tutta la *Poetica* e la prima parte della *Retorica* (dopo diventa un po' densa). Provate invece ad andargli a fare il libro Z della *Metafisica*. Niente, non si può. E già il *De Interpretatione* ... sino all'altro ieri lo capivamo e poi viene fuori Lo Piparo che dice non avevamo capito un tubo e quindi bisogna ricominciare.

Buona parte di Cartesio è buona filosofia popolare. Locke: tutto il *Saggio sull'intelletto umano* nasce da discorsi che facevano intorno alla tavola. Si può leggere tutto. Fontenelle spiega la pluralità dei mondi a una signora, vabbe' poi è maschilista e alla fine dice "ma non voglio più turbare la sua testolina, per carità; dovete occuparvi d'altro". Ma per tutto il libro fa della *philosophie pour dames*, e anche *philosophie pour messieurs*. Spinoza, niente, fino a metà dell'*Etica* è tecnica. Tutta la teoria delle passioni è buona filosofia popolare: potete far leggere la gelosia di Spinoza al droghiere e vedete se non ne viene preso. Zenone no: tecnico. Infatti De Crescenzo dice solo che era frocio.

Forse il filosofo mestatore che potrebbe fare della buona filosofia popolare fa finta di fare la filosofia tecnica. Per una somma più che ragionevole sono capace di mettervi in termini popolari quasi tutto *Sein und Zeit*: l'inautenticità, la felicità, la chiacchiera, l'essere per la morte... si può raccontare al droghiere. Lui la metteva difficile perché voleva sedurre Hanna Arendt, che c'è cascata come una pera.

C'è degli aspetti popolari nella filosofia analitica? Con Quine non ce la fate; neanche con Putnam. Grice e Austin sì, si può fare benissimo. È vero che il mondo americano non segue queste cose? Non ci sono fenomeni come Modena, ma adesso, non parlando solo di filosofia ma di discorsi critici, allo Y Center di New York pagano un sacco di soldi, prenotando sei mesi prima, per andare a sentire in 1500 persone anche dibatti culturali.

Qual è il problema, piuttosto, della filosofia analitica? Uno è venuto fuori un po' nella discussione, con qualche venatura di *hubris* immediatamente corretta da una severa autocritica: il gusto, il bisogno, la nobile necessità di sentirsi specialisti che porta a lavorare su un sillabo estremamente ristretto e a risolvere *puzzles* periferici. Poi succede che io una volta parlo con Putnam dei Mutakallimun (o Motocallemin), setta araba nota solo a pochi medievisti, e lui sapeva tutto. Non si permetterebbe mai di lasciare capire che lui lo sa. Perché non sarebbe serio secondo l'etica del Campus americano. Starebbe zitto zitto.... e invece

aveva letto cose sui Motocallemin, [che sostenevano] una particolare teoria, quasi un occasionalismo alla Malebranche di stampo islamico.

E poi l'altro problema che taglia in parte fuori – dice Marco – gli analitici dal grande pubblico è la scissione tra filosofia teoretica e storia della filosofia. Gilbert Harman aveva messo un cartello a Princeton al Dipartimento di filosofia: "Proibito l'ingresso agli storici (della filosofia)". Se noi parliamo della designazione rigida facciamo il possibile per metterci dentro anche i paronimi di Sant'Anselmo. Là no. Questo impedisce delle prese di contatto con un tessuto culturale molto più ampio che può colpire e interessare il grande pubblico. Questo è il vero vizio non degli analitici europei, che poi sanno anche la storia della filosofia, ma certamente degli analitici americani, che poi certe volte rischiano di riscoprire l'acqua calda perché non sanno che gli Stoici quel problema lì lo avevano già messo a posto o dimostrato che era irresolubile.

Ecco, questo è in particolare il problema, insieme alla nota dicotomia *town and gown* che va da Oxford e Cambridge in Inghilterra a Cambridge Massachusetts, che è molto forte e quindi taglia fuori il pensatore dal tessuto cittadino. Quello che fa la filosofia continentale è il fatto che viene fatta nelle città in presenza dei problemi. Ma più che altro è questa scissione pratica nei confronti delle grandi culture, e quindi è un problema non filosofico ma un problema sociologico, secondo me, quello che può mettere in crisi certi analitici che chiedono "perché il nostro discorso non arriva agli altri": è un problema soltanto sociologico.

Per finire, una sola cosa. Io ho avuto un maestro che è morto giovanissimo. Era assistente di Abbagnano, si chiamava Giovanni Cairola, lavorava su Duns Scoto, quando io ho finito il secondo anno è morto di un colpo, mentre andava in barca; io lo amavo molto; ha scritto poco, ma aveva fatto una conferenza anti-boeziana di cui ricordo solo il titolo: "La filosofia non consola". Questo, mi pare – rispetto al problema tecnico/non tecnico, edificante/non edificante, educare alla responsabilità, la crescita dei compiti essoterici… – questo è un punto da tenere fermo: se lasciamo credere che la filosofia ci consola, allora facciamo della psicanalisi senza neanche l'analisi e vendiamo del sapere fittizio. Se riusciamo a raccontare che la filosofia *non* consola per il semplice fatto che vi dice che alla domanda non c'è risposta, allora la filosofia ha ancora un certo compito educativo.

Reply to Peter M.S. Hacker

Diego Marconi

University of Torino

In his article for *Argumenta*, Professor Hacker essentially reiterates views he had presented, at much greater length, in his recent book *Wittgenstein: Comparisons & Context* (Oxford University Press, 2013). Here I will not discuss such views, as I already did so, albeit briefly, in a review I wrote for the *Notre Dame Philosophical Reviews* (25.10.2014). However, as Hacker polemically discusses my review in his article, I should like to reply to some of his criticisms. I thank the editor and board of *Argumenta* for providing this opportunity.

In my review, I commented on Hacker's view that understanding a word, sentence or utterance is no mental state or process but "is more akin to an ability" (Hacker 2013: 104, 148). What I said was:

> [Hacker's view] is borne out by some uses of 'understand', as in "Though she understands written English quite well, she has trouble with spoken English", but not by others ("She knows how to understand this sentence"—?). Moreover, an ability *to do what*? Abilities are individuated by their complements; e.g., sight is an ability, it is the ability to see. Once we have established so much concerning our use of the word 'sight' [...] what we are really interested is in what seeing *is*—what the exercise of the ability consists in. In the case of understanding, we seem to use the same word for both the ability ("She understands English") and its exercise: when we say "Did you understand what I said?" we are not asking whether our interlocutor possessed a certain ability but whether he exercised it.

The first sentence between parentheses, i.e. "She knows how to understand this sentence", was (obviously, I would believe) *not* intended as itself a counterexample to Hacker's thesis, as he took it to be; it was intended to show that the sentence "She understands this sentence" cannot be paraphrased into a sentence explicitly stating someone's possession of an ability (as it should, if 'understand' always indicated an ability). Indeed, as far as I can judge I fully share Hacker's opinion that "She knows how to understand this sentence" "is not very good English": that was exactly the point, and the reason for the '?'. The intended counterexample was "She understands this sentence", which, in my opinion, is not easily interpreted as being about possession of an *ability*. Hacker disagrees: "If [the sentence] means 'She understands this sentence', then it confirms the claim that understanding is ability-like, for she can say what it means." Now, perhaps in most cases if S understands *p* then S can "say what *p* means", i.e. provide an adequate paraphrase of *p*; however, few would take the two sentences ("S understands *p*", "S can say what *p* means") as synonymous or even neces-

Argumenta 1, 2 (2016): 291-293
ISSN 2465-2334
© 2016 University of Sassari
DOI 10.14275/2465-2334/20162.MAR

sarily equivalent (just suppose S is aphasic). If one sees a mountain one can usually describe it: this does not entail that seeing a mountain *is* being able to describe it, or that seeing a mountain is an ability (rather than the exercise of an ability).

Later on (49: fn. 4), Hacker says: "Professor Marconi surprisingly ascribes to me the view that grammatical propositions are not merely formulations of rules for the use of words, but 'more precisely exclusionary rules'. This is not a view I have ever advanced." He then goes on to give examples of grammatical propositions that are not exclusionary rules. Fine. However, in his book (Hacker 2013: 166), he wrote the following: "Clearly, it is not a theory, let alone a hypothesis, that red is a colour, that red is darker than pink, or that nothing can be red and green all over—any more than it is a theory, let alone a hypothesis, that bachelors are unmarried. Nor is it a theory or hypothesis that there can be no such thing as a private language or a private ostensive definition—even though it is not immediately obvious [...]. These are exclusionary *rules*—and what they exclude is a meaningless form of words." To be sure, Hacker does not say in so many words that *every* grammatical proposition is an exclusionary rule; but, given his list, perhaps it is not so surprising that I saddled him with such a view. I am happy to learn he does not hold it.

At some point in my review I mentioned Wittgenstein's suggestion that behaviour is not inductive but criterial (logical) evidence for the mental. Once more, I was quoting Hacker: "In third-person cases, psychological attributes are predicated of agents on the basis of what they do and say (including their avowals of thought and experience) but this is not inductive evidence for the inner, it is *logically* good evidence or 'criteria'" (Hacker 2013: 91). Notice that Hacker is speaking of what agents do and say *in general*, not of certain forms of behaviour as distinct from others: it is what agents do and say that, we are told, "is *logically* good evidence or 'criteria'" of the inner. Based on this, I pointed out that "this seems to entail that we do not *conjecture* the mental from behaviour, which, in turn, could be taken to entail that we cannot go wrong." Hacker now replies (52: fn. 11) that "it seems no such thing, and it could not be so taken". For it is one thing to "conjecture from the fact that one's wife is taking an aspirin, that she has a headache (she suffers from headaches and takes aspirin to alleviate them)", and quite another to say she has a headache based on the fact that "she holds her head moaning 'I have a terrible headache'". True, or anyway, plausible; however, no such distinction had been originally introduced. Surely taking aspirin is part of what an agent does and says, which—we had been told—"is *logically* good evidence or 'criteria'" of the inner.

Finally, on p. 56 Hacker quotes me as saying: "If our 'conceptual scheme' is to be investigated by surveying our ordinary use of words, can such use really be conceived as completely segregated from scientific uses? E.g., is our ordinary use of 'mind' and related words entirely isolated from scientific theories of the mind and the brain?". His reply appears to presuppose that we can and should sharply distinguish between "those parts of our conceptual scheme that belong to the province of technical terms in science", which should "be examined and described by scrutiny of the technical uses of words", and those parts of our conceptual scheme that do not belong to the technical province, which "are to be examined and described by reference to the ordinary, non-technical uses of words." This is, essentially, the distinction I intended to question: my doubt was that, nowadays, our talk of the mental may be so penetrated with loans from

psychology and the neurosciences that it has become hard to isolate "purely ordinary" uses of mind-related words. That such loans do not often come with full understanding of their original scientific "grammar" may originate all sorts of confusion, as Hacker points out, but it does not make it easier to isolate *really* ordinary uses, as Hacker appears to require.

Advisory Board

SIFA former Presidents

Eugenio Lecaldano (Roma Uno University), Paolo Parrini (University of Firenze), Diego Marconi (University of Torino), Rosaria Egidi (Roma Tre University), Eva Picardi (University of Bologna), Carlo Penco (University of Genova), Michele Di Francesco (IUSS), Andrea Bottani (University of Bergamo), Pierdaniele Giaretta (University of Padova), Mario De Caro (Roma Tre University), Simone Gozzano (University of L'Aquila)

SIFA charter members

Luigi Ferrajoli (Roma Tre University), Paolo Leonardi (University of Bologna), Marco Santambrogio (University of Parma), Vittorio Villa (University of Palermo), Gaetano Carcaterra (Roma Uno University)

Robert Audi (University of Notre Dame), Michael Beaney (University of York), Akeel Bilgrami (Columbia University), Manuel Garcia Carpintero (University of Barcelona), José Diez (University of Barcelona), Pascal Engel (EHESS Paris and University of Geneva), Susan Feagin (Temple University), Pieranna Garavaso (University of Minnesota, Morris), Christopher Hill (Brown University), Carl Hoefer (University of Barcelona), Paul Horwich (New York University), Christopher Hughes (King's College London), Pierre Jacob (Institut Jean Nicod), Kevin Mulligan (University of Genève), Gabriella Pigozzi (Université Paris-Dauphine), Stefano Predelli (University of Nottingham), François Recanati (Institut Jean Nicod), Connie Rosati (University of Arizona), Sarah Sawyer (University of Sussex), Frederick Schauer (University of Virginia), Mark Textor (King's College London), Achille Varzi (Columbia University)
Wojciech Żełaniec (University of Gdańsk)