EUROPEAN JOURNAL OF
ANALYTIC PHILOSOPHY

UDK 101
ISSN (Print) 1845-8475
ISSN (Online) 1849-0514
https://doi.org/10.31820/ejap

Editors
Luca Malatesti
University of Rijeka, lmalatesti@ffri.hr
Majda Trobok
University of Rijeka, trobok@ffri.hr

Assistant editor
Marko Jurjako
University of Rijeka, mjurjako@ffri.hr

Managing editor
Borna Debelić
University of Rijeka, debelic@pfri.hr

Editorial administrator, web maintenance and layout
Ivan Saftić
University of Rijeka, isaftic@ffri.hr

Editorial boards
Elvio Baccarini (University of Rijeka), Carla Bagnoli (University of Wisconsin-Milwaukee), Boran Berčić (University of Rijeka), Clotilde Calabi (University of Milan), Mario De Caro (University of Rome), Katalin Farkas (CEU Budapest), Luca Ferrero (University of Wisconsin-Milwaukee), Pierre Jacob (Institut Jean Nicod, Paris), Carlo Penco (University of Genoa), Snježana Prijić-Samaržija (University of Rijeka), Michael Ridge (University of Edinburgh), Marco Santambrogio (University of Parma), Sally Sedgwick (University of Illinois, Chicago), Nenad Smokrović (University of Rijeka), Bruno Verbeek (University Leiden), Alberto Voltolini (University of Turin), Joan Weiner (Indiana University Bloomington), Berislav Žarnić (University of Split)

Advisory board
Miloš Arsenijević (University of Belgrade), Raphael Cohen-Almagor (University of Hull, UK), Jonathan Dancy (University of Reading/University of Texas, Austin), Mylan Engel (University of Northern Illinois), Paul Horwich (City University New York), Maria de la Conception Martinez Vidal (University of Santiago de Compostela), Kevin Mulligan (University of Geneva), Igor Primoratz (Charles Sturt University), Neven Sesardić (Zagreb), Mark Timmons (University of Arizona, Tucson), Gabriele Usberti (University of Siena), Nicla Vassallo (University of Genoa), Timothy Williamson (University of Oxford), Jonathan Wolff (University College London)

Publisher, Editorial office
University of Rijeka, Faculty of Humanities and Social Sciences, Department of Philosophy
Address: Sveučilišna avenija 4, 51000 Rijeka, Croatia
Phone: +385 51 265 794
Fax: +385 51 265 799
E-mail: eujap@ffri.hr

Web address
https://www.ffri.hr/phil/casopis/index.html

Printed by Impress, Opatija (Croatia)
200 copies available

TABLE OF CONTENTS

EuJAP | Vol. 13, No. 2, 2017
UDK
1 SWINBURNE, R.
164.031

DIFFERENT ARGUMENTS, SAME PROBLEMS. MODAL AMBIGUITY AND TRICKY SUBSTITUTIONS*

RAFAL URBANIAK
Ghent University
University of Gdansk

ABSTRACT

I illustrate with three classical examples the mistakes arising from using a modal operator admitting multiple interpretations in the same argument; the flaws arise especially easily if no attention is paid to the range of propositional variables. Premisses taken separately might seem convincing and a substitution for a propositional variable in a modal context might seem legitimate. But there is no single interpretation of the modal operators involved under which all the premisses are plausible and the substitution successful.

Keywords: *Church-Fitch paradox, futura contingentia, modal logic, modal operators, propositional quantification, Swinburne's modal argument*

1. Introduction

Certain arguments use modalities in close, but different meanings. This might lead to the situation in which premisses taken separately seem convincing (and substitution for a propositional variable might seem legitimate), but nevertheless, no single interpretation of the modal operators involved makes all the premisses plausible and the substitution legitimate.

While it's difficult to *a priori* point to a wider class of arguments in which the problem arises, the issue might be more common than it might seem: at least in philosophical arguments it occurs in quite different contexts. This suggests that philosophers should be on the lookout for this type of error whenever a philosophical argument involving both modalities and propositional quantification is involved.

*Received: 09.04.2016.
Accepted: 14.12.2017.

The goal of this paper is to diagnose this problem in a few fairly well-known philosophical arguments, which normally aren't discussed together, and whose similarity hasn't been previously noticed:

1. Swinburne's modal argument for the existence of the soul;
2. a logical argument for fatalism;
3. the Church-Fitch paradox.

The goal of discussing arguments concerning quite different topics is to emphasize that the flaw isn't too topic-dependent. The arguments were chosen because they are well-known, they concern different topics and yet they all commit the same fallacy.

From the assumption that it is logically possible that a human being survives the destruction of their body and a few additional modal premisses Richard Swinburne infers the actual existence of souls. Various variants of the argument against future contingents rely on modal and temporal premisses and seem to lead to the conclusion that there are no future contingent events. The Church-Fitch paradox leads to the conclusion that the existence of unknown truths excludes all truths being knowable. I start with presenting the first two arguments, then I argue that whatever appearance of soundness they have, they owe it to the ambiguity of the modal operators involved and lack of attention to propositional quantifier range. Then I describe the third argument and point out a similar issue with it.

While the arguments for the sake of clarity and brevity are to some extent formalized, the main point is not about the formal tools, but rather about their misuse in representation of the underpinning philosophical intuitions.

2. Swinburne's modal argument

Let's start with the original formulation of the argument. The argument in its fullest version can be found in (Swinburne 1986, ch. 8). It also occurs in (Swinburne and Shoemaker 1984, ch. 2). In (Swinburne 1996) the author develops a defense of the modal argument against certain objections raised in the literature of the subject.

First, some abbreviations. $\Diamond$ is the possibility operator, $\Box$ is the necessity operator, $\wedge$ is the conjunction symbol, $\rightarrow$ is material implication, $\Rightarrow$ is the logical/definitional implication, $\leftrightarrow$ is material equivalence, $\Leftrightarrow$ is the logical/definitional equivalence, and $\neg$ is the negation symbol.

The key difference between the single arrow symbols and double arrow symbols is that the former are connectives in the object language, while the latter are meta-linguistic. Moreover, material equivalence only says that it is not the case that one side is true and the other false, while the

definitional/logical equivalence requires that it is necessarily so, and allows for the substitution of equivalents in modal contexts.

I customized propositional constants for mnemonic purposes.

C ⇔ Swinburne is a **C**onscious person and exists in 1984.
D ⇔ Swinburne's body is completely **D**estroyed in the last instant of 1984.
S ⇔ Swinburne has a **S**oul in 1984.
E ⇔ Swinburne **E**xists in 1985.

Swinburne introduces a variable *p* that is supposed to range over propositions of a specific sort: "p ranges over all consistent propositions compatible with C∧D and describing 1984 states of affairs." (Swinburne 1996, 69) We'll work with **T** as the underlying modal logic (that is, apart from distributing □ over implication, we have reflexivity ($\Box p \rightarrow p$) which requires that whatever is necessary is true.

The first premiss of the argument is contingent. It says that Swinburne is a conscious person and exists in 1984:

(1) C

The second premiss states that for any sentence about 1984 compatible with C and D it is possible that Swinburne survives the destruction of his body, and yet that his compatible sentence is true:

(2) For all *p*, $\Diamond$(C∧D∧*p*∧E)

The third premiss says that it is not possible (at least for Swinburne) to survive the complete destruction of his body if he doesn't have a soul (an immaterial part):

(3) $\neg\Diamond$(C∧D∧¬S∧E)

Premiss (2) says that any sentence compatible with C∧D and describing 1984 states of affairs is compatible with C∧D∧E but premiss (3) says that ¬S is not compatible with C∧D∧E. Therefore, ¬S is not a sentence that is compatible with C∧D and describes 1984 states of affairs. Or, in other words, premisses (2) and (3) together entail that ¬S is not within the range of *p*. But if ¬S is not compatible with C∧D, then C∧D entails S. But D doesn't have any impact on the truth of S, and so, if C∧D entails S, then so does C alone.

The argument has been developed into a fully formalised form and reformulated into a version immune to what was considered the main objection put forward in (Zimmerman 1991; Alston and Smythe 1994; Stump and Kretzmann 1996). Full details of the construction and a longer discussion of known objections can be found in (Urbaniak and Rostalska, 2010). Here I just present the final effect.

To proceed with our analysis, we need three more abbreviations.

84(p) ⇔ it is a fact about 1984, that p

By 84(p) we only mean that p states something about an event or a state of affairs in 1984 and it does not state anything about an event or state of affairs "outside of" 1984. What is also important, a sentence does not have to be true in order to be about 1984. The notion of being about 1984 is a bit vague, but in fact we do have decent intuitions about whether a sentence is (purely) about 1984. So, for instance, we exclude sentences like:

It is the case in 1984 that in 1985 Swinburne will not exist.

Although, in a way, this sentence is about 1984, it is not purely about 1984, because it clearly implies a contingent sentence about 1985. However, both sentences:

Swinburne is purely material in 1984.
Swinburne is not purely material in 1984.

seem, on the face of it, to be purely about 1984. While *prima facie* it might seem that 84(-) is a predicate, it is intended as a connective: "*it is a claim about 1984 that …"* is supposed to be completed by a sentence, not a name thereof. Sometimes, in informal discussion I will simplify the discourse by speaking as if I was talking about a predicate, but in such cases nothing in the discussion prevents reformulation in which it is made explicit that 84(-) is a connective.

Now, we also add a piece of notation for expressing the property of *being true about 1984*:

(4) tr84(p) ⇔ 84(p)∧ p

The third abbreviation is:

(5) ◇(p∧C∧D) ⇔ p is compatible with C∧D.

It may seem slightly unclear what sort of compatibility Swinburne has in mind. He emphasises that it is the same notion as that of logical coherence, quite explicitly denying that there is a separate "metaphysical" kind of necessity: "…the contrast is misleading. For not merely is the necessity of both kinds equally hard, but has the same nature - the necessary is that which holds in all possible worlds, where 'possible' means 'coherently describable' " (Swinburne 1986, 314).

Now, we can carefully state the evolved version of the argument. The first premiss only states that Swinburne is alive and conscious in 1984:

(6) C

The second premiss is a modification of Swinburne's original second premiss. It also captures the assumption about the range of variables that

was mentioned but not included in the formula. At the first stab we might want to formalize the assumption as follows:

(7) For all *p,* [84(*p*)∧◇(*p*∧C∧D) →◇(C∧D∧*p*∧E)]

This, however, would leave the argument unnecessarily open to the following objection. Eliminate the universal quantifier and substitute M for *p*, where

M⇔Swinburne is purely material in 1984 for *p.*

84(M) seems intuitively true: Swinburne's being purely material in 1984 is a fact about 1984. Moreover, ◇(M∧C∧D) also seems true, unless we want to decide the issue at question beforehand: it is at least possible that Swinburne is purely material and conscious in 1984, and his body is destroyed in the last instant of 1984. This would make the antecedent of the resulting substitution at least strongly plausible. The consequent — ◇(C∧D∧M∧E) — however, would say that it is possible that Swinburne is conscious and purely material in 1984, his body is destroyed in the last instant in 1984, and yet he manages to survive into 1985. This doesn't seem plausible, and so (7) could be argued to entail a substitution which isn't very convincing.

Now, using tr84 instead of 84 we obtain:

(8) For all *p,* [tr84(*p*)∧◇(*p*∧C∧D) →◇(C∧D∧*p*∧E)]

(8) says about *any* proposition *p* that if it is true, purely about 1984, and compatible with C∧D, it is compatible not only with the claim that Swinburne is conscious and alive in 1984 and his body is destroyed in the last moment of 1984, but also compatible with the claim that Swinburne is conscious and alive in 1984, his body is destroyed in the last moment of 1984 and yet he survives and exists in 1985.

(8) is not susceptible to an objection analogous to the one that we just put forward against (7). For say we eliminate the universal quantifier and substitute M for *p.* To argue that this substitution is false, we need to argue that its antecedent is true. But the first conjunct now reads tr84(M) — 84(M)∧M — and while in the previous argument we only needed the assumption that being material and conscious is at least possible, now we would need the assumption that Swinburne indeed was purely material in 1984. But insisting that this is the case already decides the issue to be decided by the argument. To undermine a premiss, we'd be arguing that it is false, because the conclusion of the whole argument is — not an interesting criticism at all.

Premiss three is exactly the same as in the original argument:

(9) ¬◇(C∧D∧¬S∧E)

It is obviously equivalent to:

(10) $\Box(C \wedge D \wedge E \rightarrow S)$

(10) says that necessarily, if Swinburne is conscious and alive in 1984, his body is destroyed in the last moment of 1984 and yet he exists in 1985, he has a soul in 1984. It is meant to capture the intuition that to survive the destruction of one's body, one has to have a soul.

The next premiss says that 'Swinburne does not have a soul in 1984' is purely about 1984:

(11) $84(\neg S)$

and another one claims that if C and D necessarily entail S, then so does C:

(12) $\Box(C \wedge D \rightarrow S) \rightarrow \Box(C \rightarrow S)$

That is, if the facts that Swinburne exists and is conscious in 1984 and that his body is destroyed in the last moment of 1984 entail that Swinburne has a soul in 1984, the fact that Swinburne's body is destroyed in the last moment of 1984 has no relevance for this conclusion and the very fact that Swinburne is alive and conscious in 1984 already necessitates the fact that Swinburne has a soul in 1984.

These assumptions logically entail S. First, eliminate the universal quantifier from (8), substituting ¬S for p:

(13) $tr84(\neg S) \wedge \Diamond(\neg S \wedge C \wedge D) \rightarrow \Diamond(C \wedge D \wedge \neg S \wedge E)$

from (9) and (13) we obtain:

(14) $\neg[tr84(\neg S) \wedge \Diamond(\neg S \wedge C \wedge D)]$

we apply De Morgan's law to it:

(15) $\neg tr84(\neg S)$ or $\neg\Diamond(\neg S \wedge C \wedge D)$

At this point we split the disjunction into a proof by cases. Suppose $\neg tr84(\neg S)$. By (4) this means that either $\neg 84(\neg S)$ or $\neg\neg S$. But (11) says that $84(\neg S)$. So $\neg\neg S$ and hence S. Suppose on the other hand that $\neg\Diamond(\neg S \wedge C \wedge D)$. In this case we get:

(16) $\neg\Diamond(\neg S \wedge C \wedge D)$

Quite easily we now obtain:

(17) $\Box\neg(\neg S \wedge C \wedge D)$

Since:

(18) $\neg(\neg S \wedge C \wedge D) \Leftrightarrow [(C \wedge D) \rightarrow S]$

and because we can substitute logically equivalent expressions within the scope of modal operators, we can infer:

(19) $\Box((C \wedge D) \rightarrow S)$

Now, we apply *modus ponens* to (12) and (19) and obtain:

(20) $\Box(C \rightarrow S)$

With (6), by reflexivity for $\Box$ and *modus ponens*, this entails:

(21) S

Either way we obtain S, which completes the argument.

3. Assessing Swinburne's argument

The argument might be criticised for being epistemically circular in the following sense. Basically, (8) says that no *true* proposition compatible with Swinburne's being alive and conscious in 1984 and his body being destroyed excludes the possibility of him surviving the destruction of his body. This is a fairly strong claim, because it is equivalent to the claim that *any* proposition about 1984 compatible with C∧D, which excludes the possibility of Swinburne's survival (while C∧D), is already false.

To see the equivalence, unpack the expression in the scope of the quantifier in (8) as follows:

$84(p) \wedge p \wedge \Diamond(p \wedge C \wedge D) \rightarrow \Diamond(C \wedge D \wedge p \wedge E)$

It is now a matter of purely propositional manipulation (contraposition, really) to see that this is equivalent to:

$84(p) \wedge \Diamond(p \wedge C \wedge D) \wedge \neg \Diamond(C \wedge D \wedge p \wedge E) \rightarrow \neg p$

Thus, for instance, if one believes that M ($\Leftrightarrow$ Swinburne's is purely material in 1984) excludes such a possibility, and is compatible with C∧D, one is committed also to the falsity of M. But if this is the case, by accepting (8) we already seem to have a firm philosophical position on the issue.

A way out seems to be to say that no sentence purely about 1984 is incompatible with Swinburne's survival in 1985 because sentences purely about 1984 don't entail anything about 1985. So, no such sentence, even if true, could exclude Swinburne's survival of the destruction of his body. So to avoid the difficulty from the previous paragraph, one needs to interpret compatibility using a Humean notion of necessity on which no truth about one time necessitates anything about some other time, presumably Swinburne's notion of logical necessity.

Alas, the problem is that if we assume that our notion of compatibility is purely logical, we now have a reason to reject premiss (9) which says that

Swinburne cannot survive the complete destruction of his body and continue to exist in 1985 if he doesn't have a soul in 1984. After all, as Humean logical modality is involved, we have no reason to think that a sentence purely about 1984 should be analytically incompatible with any sentence about 1985. In other words, to accept (9) we have to use a stronger, presumably metaphysical interpretation of modality on which not having a soul in 1984 is incompatible with surviving the destruction of the body at the end of 1984.

Swinburne argued (in personal communication) that the truth of (8) is available even to children when he tries to explain the argument to them. They tend to agree when faced with statements like ‘Look, it is at least logically possible, whatever else is true and doesn't exclude us being conscious and our bodies being completely destroyed, that we survive the complete destruction of our bodies’. There are two points to make to explain these intuitions away. First, there is a notion of possibility on which (8) comes out true, but which falsifies (9). It is quite likely that some people, when faced with the sentence quoted in this passage, use this notion to assess its truth. Then, when they're faced with the informal reading of (9), they use quite a different metaphysical notion of possibility not noticing the difference. Second, there is an important scope distinction to be kept in mind. On one reading, the quoted sentence says exactly what (8) says, and yields a rather strong statement. On the other reading, it rather says that no matter what is true about 1984 and doesn't exclude C and D, it is still possible to survive the complete destruction of one's body. On the second reading, anyone who admits $\Diamond(C \wedge D \wedge E)$, a rather weak claim of mere logical possibility is committed to this claim. On this reading, however, the claim is too weak to constitute a premiss of a valid modal argument.

The issue can be rephrased in terms of substitutions: on one reading, substituting $\neg S$ for p when eliminating the quantifier yields a false antecedent, but we have no reason to accept the premiss, and on another, we have a reason to accept both the premiss and its substitution, but we have no reason to think that the consequent of the resulting sentence is false.

4. Arguing against future contingents

Aristotle in the IXth chapter of *On Interpretation* considers an argument for the non-existence of propositions about future contingent events. The argument is supposed to defend *fatalism*, the view that each future event will take place necessarily. The view that one can prove fatalism on merely logical grounds is called *logical fatalism*. I put well-known interpretative issues related to Aristotle's formulation aside (see however Øhrstrøm and Hasle 1995, 6-109) and look at a streamlined version of one of the best formulations of an argument for logical fatalism as developed by Prior and Rescher (see Prior 1967, 119-121).

- Formula F*yp* is read *in y units of time it will be the case that p*.
- Analogously we read the formula P*yp* as *y units of time ago it was the case that p*.
- Formula T*ap* reads *it is true at time a that p.*

Before we move to the argument itself, some preliminary inference rules are needed:

(RT) If $\vdash p$, *then* $\vdash \mathrm{T}ap$, for all a

(TC) $\mathrm{T}a(p \rightarrow q) \vdash \mathrm{T}ap \rightarrow \mathrm{T}ab$

(RD) If $\vdash p$, then $\vdash \Box ap$, for all a

(DC) $\Box a(p \rightarrow q) \vdash \Box ap \rightarrow \Box aq$

Roughly, (RT) says that whatever is provable is true at any time *a*, and (TC) says that truth at any time *a* distributes over implication. (RD) says that whatever is provable is necessary at any time *a*, and (DC) states that being necessary at *a* distributes over implication. All of these principles seem quite convincing.

We start the argument with a premiss saying that *p* will be true in *n* units of time just in case it was *m* units of time ago true that *p* would be true in *m*+*n* units of time, and another premiss saying that if it is true at *a* that *p* was true *m* units of time ago, it is necessary that *p* was true *m* time units ago:

(P1) $\mathrm{F}np \leftrightarrow \mathrm{P}m\mathrm{F}(m+n)p$

(P2) $\mathrm{T}a\mathrm{P}mp \rightarrow \Box a\mathrm{P}mp$

(P2) is supposed to capture the intuition that the past is necessary and cannot be changed. Now, (RT) and (TC) allow us to first add T*a* in front of the left-to-right implication from (P1) and then distribute it over the implication:

(AP2) $\mathrm{T}a\mathrm{F}np \rightarrow \mathrm{T}a\mathrm{P}m\mathrm{F}(m+n)p$

Let's now substitute $\mathrm{F}(m+n)p$ for *p* in (P2):

(AP3) TaPmF$(m+n)p \rightarrow \Box aPmF(m+n)p$

(AP2) with (AP3) give us:

(AP4) TaF$np \rightarrow \Box a$PmF$(m+n)p$

Now we take the right-to-left implication from (P1) and use (RD) and (DC) to introduce and distribute the necessity operator:

(AP5) $\Box a$PmF$(m+n)p \rightarrow \Box aFnp$

Finally, we put (AP4) and (AP5) together, obtaining the conclusion of the argument:

(AP6) TaF$np \rightarrow \Box a$Fnp

This conclusion says that whenever it is at *a* true that *p* will take place in *n* time units, it is necessarily the case.

Having described the arguments, let's focus on identifying the type of error that makes it seem plausible.

5. Assessing argument about future contingents

(RT), (RD) and (P1) seem rather innocent. One could be worried that (TC) and (DC) correspond to axiom K of the standard modal logic, which results in issues related to logical omniscience and some deontic paradoxes. These issues, however, don't seem to carry over to the temporal reading. As for the former, while perhaps it is a problem that actual agents' knowledge is not closed under logical consequence, there is no good argument to the effect that being true at a time shouldn't be so closed. As for the latter, most notable paradoxes are the Good Samaritan paradox and Ross's paradox. The Good Samaritan arises when one formalizes *It ought to be the case that Jones helps Smith who has been robbed* as $O(h \wedge r)$ and then uses K to infer *Or*, that Smith should be robbed. Part of the issue is that the formalization does violence to the original premiss, which, come to think about it, doesn't say that it ought to be the case that Jones helps Smith and that Smith has been robbed. But even if we're worried about conjunction elimination in deontic contexts, it doesn't seem to be problematic when applied to being true at a time.

Ross's paradox, on the other hand, arises when K is used to deduce $O(m \vee b)$, that *it ought to be the case that the letter is mailed or burned,* from *Om*, the claim that *it ought to be the case that the letter is mailed.* Again, while there might be reasons to think that obligation is not closed under disjunction, there is no analogous intuition about being true at a time.[1]

[1] I am grateful to an anonymous referee for pressing me on the issues with these prima facie suspicions about (TC) and (DC).

So it seems the most troubling move is the one captured by (P2). How exactly should we understand the modality in (P2)?

Alas, on this reading, no one who doesn't already accept causal determinism will accept the premiss itself. For then (P2) says that whatever happened in the past was causally determined, which is not something that someone who believes certain past human choices were made freely would accept.

So perhaps we should focus on a somewhat different reading, according to which it is not *being causally determined*, but rather *there being at time a no way of influencing the truth-value of p*. On this reading, the premiss only seems to say that whatever is in the past can no longer be changed. And indeed, we do seem to have this intuition if *p* is a simple sentence clearly about the past. For instance, if it so happened that student X failed his logic exam, nothing in the future can be done to change this fact from the past.[2]

But now, the key move is from (P2) to (AP3), where for *p* we substituted F(*m*+*n*)*p*. As pointed out already by Ockham (see Boehner and Ockham 1945), claims such as *it was the case m units of time ago that in m+n units of time p will be the case* are not strictly about the past. While an indeterminist might still have the intuition that we can do nothing to change past events (even though some of the past choices were free at the time when they were made), she will definitely deny that such spuriously past-tensed descriptions really describe a past event.

The problem is clear: as far as in this interpretation (P2) is convincing for sentences not about the future, if we allow substitutions of sentences involving the future it turns out to build in the intended conclusion: the necessity of all future events. (Having said that, developing a formal semantics for temporal logic which captures the distinction between legitimate and illegitimate substitutions is quite a task - see for example the discussion in Prior 1967.)

This can be avoided by restricting the range of legitimate substitutions (we did the same thing when making the range of quantifiers explicit in Swinburne's argument). Say by D<*ap* we mean *the truth value of p does not depend on any facts occurring after a*, (P2) should be replaced with

(P2′) D<*ap*→(T*a*P*mp*→□*a*P*mp*)

Switching to (P2′) avoids the objection we just raised: for (P2′) now says that if *the truth-value of p doesn't depend on any facts occurring after a,* then if it is true at *a* that *p* was true *m* units of time ago, it is also necessary at *a* that *p* was true *m* units of time ago.

---

[2] Note: the argument assumes the past cannot be changed. Already this is a non-trivial philosophical assumption excluding, for instance, time-travel. But let's play along. Also, there might be logical reasons to to take the possibility of time-travel too seriously (see Urbaniak 2007).

The problem now is that the conclusion of the whole argument with (P2′) in place of (P2) no longer follows because it clearly is not the case that D<*a*F(*m*+*n*)*p*.

Why did we buy into (P2) in its full strength to start with, then? For one thing, plain carelessness and not thinking about whether accepting (P2) with all its tricky substitutions captures our intuition about the *past* being irreversible. But I think there's an associated reason, that has to do with further confusion with another modality. We are so used to thinking about alethic modalities such as logical or metaphysical necessity and to the ultimate validity of logical principles involving such modalities, that we are without much hesitation willing to accept all substitutions of what seems to us to be a valid general principle about a modality. This however should not be the case with temporal modalities, which turn out to be a bit more tricky.

So we observed a situation where in a modal context our intuitions about which substitutions should be admitted lead us astray. The situation, it seems, isn't very unique. Let's take a look at yet another case.

6. Church-Fitch paradox

Here's another well-known argument that turns around a tricky substitution mixed with a modal context. The goal of the argument is to show that one cannot consistently claim that all truths are knowable and that there is at least one unknown truth. Let's start with formulating the premisses, next we'll see how the conclusion follows, and then we'll move to the assessment. Aside from classical logic in the background, the premisses and principles needed for the argument are as follows (K reads *it is known that/one knows that*):

(IK) For some p ($p \wedge \neg Kp$)

(Know) For all p, $\vdash (p \rightarrow \Diamond Kp)$

(Distr) $K(p \wedge q) \vdash Kp \wedge Kq$

(Fact) $Kp \vdash p$

(Contra) If $p \vdash \bot$, then $\Diamond p \vdash \bot$

(IK) says that our knowledge is incomplete, that there is at least one unknown true proposition. (Know) states the knowability of any true proposition. (Distr) allows to distribute knowledge over a conjunction, and (Fact) states the factivity of knowledge: that whatever is known is true. (Contra) says that the possibility claim of a contradictory proposition is already contradictory.

Now the reasoning. (IK) says that there is a true sentence, say n, that is not known. (in a formal proof this move would correspond to the elimina-

-tion of the existential quantifier, substituting a fresh propositional constant *n* for *p*),

(IK2) $n \wedge \neg Kn$

As (Know) contains a universal quantifier, we can eliminate it, substituting any formula whatsoever for *p*. In particular, let's substitute the content of (IK2):

(Know2) $n \wedge \neg Kn \rightarrow \Diamond K(n \wedge \neg Kn)$

Now we can simply apply detachment to (IK2) and (Know2) getting:

(IK3) $\Diamond K(n \wedge \neg Kn)$

Now, we will show that what's in the scope of $\Diamond$ in (IK3) entails a contradiction, which thanks to (Contra) will allow us to deduce contradiction from (IK3) itself. We apply (Distr) to $K(n \wedge \neg Kn)$ and eliminate the conjunction:

(IK4) Kn

(IK5) $K\neg Kn$

Now, we apply (Fact) to (IK5) to obtain

(IK6) $\neg Kn$

which together with (IK4) yields a contradiction. This means that by (Contra), a contradiction follows already from (IK3). We obtained the conclusion that a contradiction is possible, whose negation can be proven in a very rudimentary modal logic K. So it seems that (IK) and (Know) exclude each other!

Perhaps, we're to blame (Distr) or (Fact) for the pickle? Well, (Fact) seems like a principle capturing the factivity of knowledge, and isn't independently known to lead to undesired consequences. One might, however, have an issue with (Distr) and axiom K, pointing out that they build logical omniscience into the system (well, Distr does this only partially).

But what would an attempt to solve the paradox by insisting that (Distr) doesn't apply here look like? One would have to deny the inference from (IK3) to (IK4) — that is, one would have to deny that if a subject knows that (*n* and it is not known that *n*), then it follows that the subject knows that *n* and they know that it is not known that *n*. This doesn't seem too convincing and would require an independent motivation. Moreover, such a solution would, so to speak, come too late: already (IK3) seems rather absurd, and so something problematic must've preceded the application of (Distr).[3]

[3] Thanks to an anonymous referee for pressing me on the issue of logical omniscience.

It seems that the issues with logical omniscience and those brought up by the Church-Fitch paradox aren't too related — after all, the reasoning involved is very simple and doesn't involve any massive complication that could prevent the subject from grasping the fact, say, that conjuncts follow from a conjunction and so their knowledge is closed at least under eliminating conjunction once.

7. Assessing Church-Fitch paradox

Now, the whole trick is made possible by substituting $n \wedge \neg Kn$ for p in the knowability principle (Know), leading to (Know2). And come to think about it, if n is a true but unknown sentence, there is nothing amazing about $n \wedge \neg Kn$ not being knowable. After all, if you know this conjunction, you know n, so you at the same time falsify the second conjunct! The whole argument is just a way of explicating this fact.

But of course, you might think that this is a rather cheap shot, because this is not the sort of substitutions you had in mind. You can still accept the spirit of the knowability principle and say: *when I said all truths are knowable I meant actual truths about the non-epistemic world, not some tricky sentences involving claims about what is known, I don't know which tricky sentences involving epistemic operators are and which aren't knowable!* And once this restriction on substitution is made, the argument doesn't fly to far.

Having said this, the challenge of developing a formal framework explaining which substitutions exactly lead to problems and which don't is quite daunting. This is especially so if we want to be less conservative than we are by excluding all formulas involving epistemic operators. The problem generated quite a lot of literature (see for instance Kvanvig 2006 and Salerno 2009). People who think this is an important problem have spent considerable amount of time thinking about this without reaching agreement.

One notable proposal, for instance, is due to Tennant (2002). Take the example of "No thinkers exist". The proposition is consistent, but can't be known to be true (assuming only thinkers can know things). Such sentences (whose knowledge claims are not possible) Tennant calls *anti-Cartesian*. Sentences which are not anti-Cartesian are said to be *Cartesian.* A sentence of the type Kp can be impossible for various reasons: it might be that p itself is inconsistent, it might be the case that judging that p requires the falsity of some consequence of p, or it might be that Kp is impossible due do the logical structure of p itself, despite p being consistent. Tennant's antidote to the Church-Fitch paradox is: restrict the knowability claim to Cartesian propositions only.

The approach definitely blocks the paradox, and is definitely commonsensical in spirit: don't use tricky unintended substitutions which due to logical complexities involve you in impossibilities. It also nicely illustrates how rather unexciting the philosophical lessons from taking such paradoxes too seriously can get. At its core, the proposal simply is: *when I say that all truths are knowable, of course I don't mean truths which a priori can't be known for various rather trivial reasons that have to do pretty much with self-reference.* In all fairnes, it's hard to draw any deeper lesson here.

Again, the phenomenon seems to be that initially a formula seems like an adequate formalisation of our pre-formal claim, and only after further formal development which uses tricky unintended and unexpected substitutions it is revealed that the formula in its whole unintended generality wasn't so convincing to start with.

8. So what do they have in common?

I surveyed three different arguments about quite different issues, which turned out to share the following features:

(A) **They all employ somewhat unexpected substitution for a propositional variable or for a schematic letter.**

- In the case of Swinburne's argument, premise (8), which roughly captures the intuition that nothing consistent with Swinburne's being conscious prior to the destruction of his body excludes his survival, is used by substituting a claim on the very matter at hand "Swinburne doesn't have a soul" for the propositional variable, and the interplay of various modalities involved in the argument with our shaky intuitions are used to make the argument seem plausible.
- In the case of the logical argument against future contingents, a claim involving tricky reference through time, $F(m+n)p$, is substituted for p in (P2).
- In the Church-Fitch paradox a clearly a priori unknowable claim is substituted for a variable in the knowability principle, which doesn't seem to be intended to be applicable to such claims.

(B) **The substitution is applied to a *prima facie* plausible premiss.**

- In Swinburne's argument the intuition supporting (2) is that at least we should treat the possibility of Swinburne's survival as an open possibility.
- In the fatalistic argument, (P2) at least prima facie is supported by the intuition that past cannot be changed.

- In the Church-Fitch paradox the assumption is an expression of cognitive optimism to the effect that at least in principle any aspect of the world can be the subject of knowledge.

(C) **The substitution is made in a modal context.**

- In Swinburne's argument, as discussed, how the modalities are understood bears on how plausible particular premisses are.
- In the fatalistic argument the modalities involved are temporal, and various considerations arise from different interpretations of what it means for the past to be unchangeable.
- In the Church-Fitch paradox, the modal context is epistemic.

Given the discussion so far, the following claims also clearly hold:

(D) **The arguments lead to very strong conclusions from prima facie innocent premisses.**

(E) **The unclarity about which modality is involved and which substitutions are intended as admissible, given the choice of modality, makes the premisses prima facie plausible and the argument at least initially plausible.**

Problems with these arguments are not at the formal level: they arise at the level of formalisation. The goal of the paper wasn't to show that there are *logical* flaws in the arguments. Quite the opposite: with all assumptions in place, the (semi-)formal arguments are logically correct in the sense that the conclusion follows from the premisses by means of the rules stated. The goal, however, was to show that logically correct arguments can be used to philosophically suboptimal effects, as long as not enough attention is paid to formalization, to the underlying intuitions, and philosophical justifications of the premisses.

Given that the problems are with philosophical moves and not with formal systems, it is difficult to formulate simple general principles that would help to avoid such infelicities. I hope my critical survey, however, will raise the reader’s sensitivity to the meaning of modalities and to the range of propositional variables used in various arguments. These issues, while they might sound quite distant from the point of view of a serious philosopher who wants to argue for a strong philosophical position, cannot be ignored. The devil is in the detail.

Acknowledgments

I would like to express my gratitude to two anonymous referees of the journal, whose in-depth comments were extremely useful.

REFERENCES

Alston, W. P. and Smythe, T. W. 1994. Swinburne's argument for dualism. *Faith and Philosophy* 11(1): 127–133.

Boehner, P. and Ockham, G. 1945. *The Tractatus de Praedestinatione Et de Praescientia Dei Et de Futuris Contingentibus of William Ockham: With a Study on the Mediaeval Problem of a Three-valued Logic*. New York: Franciscan Inst. St. Bonaventure College.

Kvanvig, J. L. 2006. *The Knowability Paradox*, Oxford: Oxford University Press.

Øhrstrøm, P. and Hasle, P. F. 1995. *Temporal logic: from ancient ideas to artificial intelligence*. Dordrecht: Springer.

Prior, A. N. 1967. *Past, Present and Future*. Oxford: Clarendon Press.

Salerno, J. eds. 2009. *New Essays on the Knowability Paradox*. Oxford: Oxford University Press.

Stump, E. and Kretzmann, N. 1996. An objection to Swinburne's argument for dualism. *Faith and Philosophy*,13: 405–412.

Swinburne, R. 1986. *The Evolution of the Soul*. Oxford: Clarendon Press. Page references to the second edition published in 1997 by Clarendon Press.

_____ 1996. Dualism intact. *Faith and Philosophy* 13(1): 68–77.

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

Tennant, N. 2002. *The Taming of the True*, Oxford University Press.

Urbaniak, R. 2007. Time Travel and Conditional Logics. In *Logica 2007 Yearbook*, ed. M. Peli, 247–255.

Urbaniak, R. and A. Rostalska. Swinburne's modal argument for the existence of the soul. *Philo*, 12:73–98, 2010.

Zimmerman, D. W. 1991. Two cartesian arguments for the simplicity of the soul. *American Philosophical Quarterly* 3: 217–226.

EuJAP | Vol. 13, No. 2, 2017
UDK
111:32
165.2

EXPLANATION AND INDIVIDUAL ESSENCE*

MÁRTA UJVÁRI
Corvinus University of Budapest

ABSTRACT

*In this paper I show that a novel ontic reading of explanation, intending to capture the de re essential features of individuals, can support the qualitative view of individual essences. It is argued further that the putative harmful consequences of the Leibniz Principle (**PII**) and its converse for the qualitative view can be avoided, provided that individual essences are not construed in the style of the naïve bundle theory with set-theoretical identity-conditions. Adopting either the more sophisticated two-tier **BT** or, alternatively, the neo-Aristotelian position of taking essences as natures in the Aristotelian sense, can help to evade these main charges against the qualitative view. The functional parallels with the alternative haecceitistic view of individuation and individual essence will also be considered.*

Keywords: *qualitative individual essence, bundle theory, Identity of Indiscernibles, Aristotelian essentialism, ontic explanation, explanatory-gap argument*

The qualitative view of individual essence assumes that an individual has, apart from its specific or kind essence, an individual essence as well, one that is unique to it, which can be spelled out in terms of qualitative features. Authors working on the topic further assume that the qualitative features can be captured by sets of nontrivial essential properties (Forbes 1985, 99; Lowe 1995, 69-70). The set-theoretical construal of the qualitative features, however, has the risk of excessive rigidity when it comes to the issue of identity through worlds and times: any change, even the most trifling one, in the set of qualitative features would result in the numerical change of the individual. This is precisely what the converse of

*Received: 31.3.2017.
Accepted: 28.2.2018.

the Leibniz principle of the Identity of Indiscernibles (**PII**) says: property-discernibility entails numerical difference.[1] The highly disputed (**PII**) itself guarantees that property-indiscernibility is sufficient for numerical identity. Though (**PII**) is typically not formulated for modal cases, the qualitative view seems to lose much of its appeal under the potential risk of its being committed to (**PII**) and its converse.

The present account markedly deviates from the set-theoretical reading of the qualitative view. My claim is that the qualitative features constituting the individual essence or nature of an individual make a structured, function- and goal-oriented constellation. Individual essences cannot be captured by a mere conjunction of properties. Therefore, this account does not invoke the mere listing of the features obeying set-theoretical identity conditions . To note, (**PII**) and its converse concern the mere listing of properties. I suggest that by adopting a version of the qualitative view that is more refined than the naïve bundle theory, one can avoid the main charge that the qualitative view is vulnerable to objections that stem from (**PII**) and its converse.

Apart from evading this charge, a positive support for positing qualitative individual essences can be found in the form of the ontic or metaphysical version of explanation. The explanatory role of the qualitative features has already been explored (Ujvári 2013a), but now I think that the explanatory approach can be strengthened with the metaphysically committing version. In this version the epistemic reading of explanation is replaced by its ontic reading; this represents also a decisive step in the essentialist-explanatory account of modalities.[2] The point is that once *de re* essentialism has been liberated from the modal-logical view of necessity, explanation in the ontic sense could be given a crucial role in determining the essential features of things.[3] And the ontic reading of essential features is naturally coupled with conceiving 'individuation in the metaphysical sense (as opposed to the cognitive or epistemic)' as it is recently stressed by Lowe.[4]

Admittedly, the nature of support for qualitative individual essences on the part of explanation in the ontic sense is not a strict, compelling argument: it inclines rather than necessitates its conclusion in virtue of

---

[1] The converse of (**PII**) often referred to as Leibniz's Law is obviously impeccable in the context of classical logic as a characterization of identity. However, the metaphysical problems of counterfactual identity and identity through times situate the converse of (**PII**) into another perspective.

[2] Kment (2014, ch. 4-5), Gorman (2005, 276-289). As a forerunner of the view, see Salmon (1984, 84-134).

[3] See Fine (1994) and Gorman (2005).

[4] Lowe (2014, 216).

the fact that the notion of qualitative individual essence coheres well with the essentialist-explanatory conception. 'Cohering well' means that once this conception is adopted, the significance of the very notion of qualitative individual essence increases.

The familiar alternative account of the identity of individuals through worlds and times is the haecceitistic account. Concerning the contemporary versions of haecceitism, I take them as elaborations on Kaplan's original claim that individuals can be identified through counterfactual situations without making recourse to their attributes or qualitative features (Kaplan 1975). While not denying the relevance and particular suitability of haecceitism in certain contexts, I think that much can be claimed in favor of the qualitative view of individual essence as well. The paper does not intend to contribute to any current debate between proponents of the qualitative view of individual essence and proponents of its haecceitistic view. Rather, the goal is only to bolster the position of the qualitative view by underlining its role in explanation in the ontic sense and by evading the main standard charge of the view being committed to (**PII**) and its converse.

In the first section of the paper, I consider the prospects for the qualitative view of individual essence backed by the ontic reading of explanation. I also show here that (**PII**) is not a real threat to the tenability of this view. In the second section, I point out that individuation in the metaphysical sense requires that individuation should be told apart from mere particularization of the type. In the metaphysical sense individuation is separated from mere identification, say, by ostension. Consulting the literature, one finds a striking overlap between the functions and roles attributed to haecceity by advocates of haecceitism, and the functions and roles attributed to qualitative individual essence by supporters of the qualitative view. So, in the third section, I explore the parallels, and in the assessment, I find that the explanatory function is a further crucial factor in the comparison. Turning to the tenable versions of the qualitative account, I shall present only briefly the two-tier view of **BT** (bundle theory) since it has already been discussed in the literature (Simons 1994; Cleve 1985; Ujvári 2013b). The neo-Aristotelian version of essentialism has also been spelled out many times; therefore, I shall add only, in the fourth section, a further point about how this position is applicable to a sound theory of individuation and individual essence.[5] To anticipate, I shall argue that Aristotelian natures can fruitfully be explored as bearers of the qualitative aspect of individual essence. In the last section of the paper, I visit the question of relationality, and clarify what I take to be 'genuine' individuals.

[5] See Gorman (2005), Fine (1994), Oderberg (2011), Lowe (1999).

1. The ontic reading of explanation and the threat of (PII)

When it comes to the task of explaining the actions, behavior, dispositions, etc. of individuals, it seems that this task requires considering individual essences in addition to their specific essences. Otherwise our purported explanations about individuals would remain hopelessly incomplete and gappy. According to the *explanatory-gap argument* suggested in the literature, certain possibilities of individuals would remain unexplained, and even unexplainable, if sortal essences plus accidental features were the only candidates for the explanans.[6] Say, the sortal essence of 'being human' admits a wide range of possibilities that humans are capable of realizing. But considering the actual possibilities of determinate human individuals, the sortal possibilities should be tailored to the idiosyncratic character of the given individual aimed at explaining why, for example, certain future developments are open to this individual, and some others are excluded, while both are tolerated by the specific human essence.

The significance of the explanatory function can get further support from the familiar symmetry-thesis concerning explanation and prediction. It is a truism that a good explanation could function as a prediction as well. Predictions, if they are not mere projections of past regularities to the future in the Humean fashion, should rely on explanatory connections. It is hard to imagine how our predictive practice could manage successfully without analyzing and assessing the qualitative features of the items, including those of the individuals, that show up in the predicted scenario.

The explanatory approach to positing individual essences can be strengthened with the metaphysical version of explanation. This version presupposes a departure from 'logical' modalism: as Kit Fine has argued in his criticism of the modal view of *de re* essentialist claims, the logical-modal notion of necessity cannot fully capture the metaphysics of *de re* essentialist locutions. Since these locutions should do with the very nature of things whereby they receive their identity, therefore, *de re* essentialism enjoys a distinctive metaphysical status not to be reduced to logical necessity. Departure from logical modalism continued with Gorman's suggestion that 'explanation' is the clue to decide whether a given feature of a thing is essential. He says that '**F** is essential to x just in case **F** is (i) a characteristic of x and (ii) not explained by any other characteristic of x' (Gorman 2005, 284). Not being explained by other characteristics of x does not, however, render the notion of 'essential' mysterious: features which are fundamental, and thus unexplained relative to the very nature of a given thing, may require explanation in some other contexts.

[6] Ujvári (2013a).

The notion of ’explanation’ applied here is ’ontic’. As Gorman says, ‘to speak thus of explanation is to use the word in its ontic and not in its epistemic sense.’ The key to this ontic use is that ‘explanation is a real relation between things’ (Gorman 2005, 283). The mind-independence of such real explanatory relations was also stressed long ago by Wesley Salmon. Recently, Kment has argued in a similar way: when the issue is that ’x explains y’ or ’x is the reason why y obtains’ then ‘explanation in this sense is a metaphysical relation, not an epistemic one’ (Kment 2014, 5).

It seems to me that the best way to render explanation intelligible as a metaphysical relation is to associate it with *grounding*. Grounding and explanation (in the metaphysical sense) are relations which are constitutive of things, and they may be contrasted with projected regularity patterns in the Humean epistemic sense. As Maurin has argued quite recently, grounding is ‘involved’ in ‘metaphysical explanations’ in virtue of grounding being ‘tracked’ by these explanations (Maurin 2017). Though the precise nature of the grounding - metaphysical explanation link is currently still disputed, it is clear from the discussions that it is metaphysical explanation about the real nature or essence of things that constitutes one of the relata and not explanation in the epistemic sense (Thompson 2016). Advocates of the metaphysical theory of grounding are keen on avoiding the epistemic sense of explanation: although ‘grounding is often called an *explanatory* notion ... to us it seems advisable to separate the objective notion of grounding which belongs to the field of metaphysics from an epistemically loaded notion of explanation’, emphasize Correia and Schneider in their introduction to the topic (Correia and Schneider 2014, 24). Congenial to these efforts, we can witness Kment's recording a ‘shift of focus from the modal to the explanatory domain’. The shift covers concerns about ‘grounding, essence, fundamentality’. These new concerns change the direction of entailment: according to Kment, ‘entailment seems to hold only in one direction – from the metaphysical claim to the modal one’ (Kment 2014, 14).

Without taking a position about these essentialist-explanatory concerns, I argue conditionally: *given* these concerns, particularly the emphasis on the priority of explanation in the ontic sense *and* the metaphysical reading of *de re* essential features, one might feel a temptation to consider seriously the qualitative account of individual essence and individuation. Clearly, the qualitative view cannot be dismissed, since, how could a non-qualitative haecceity, a bare ’thisness’ constitute the metaphysical nature of a thing, let alone the explanatory role of that nature in the ontic sense. The special virtue of the ontic reading of explanation is that it can help ruling out a possible objection to the explanatory-gap argument: one might object, for example, that a mere epistemic point is irrelevant from the metaphysical perspective.

A further argument, in favor of the qualitative view, rests on the presumed *analogy* between *modal explanations for continuants* by appealing to their individual essence *and* explanation of *singular event causation* by appealing to individual event essences (Ujvári 2013a). Though the analogy has some appeal, it has not gained universal acceptance. For example, Diekemper has claimed that events, unlike continuants, can be shown to have a primitive nonqualitative thisness or haecceity essential to them[7] (Diekemper 2009).

Here I am not going into the event-continuant dispute. Haecceitistic commitments just mentioned, it is worth considering though what would count, from the perspective of the haecceitistic position, as a sound objection to the positing of qualitative individual essences. It is not only Diekemper who seems to hold that Leibnizian (**PII**) is lethal for the qualitative view; Penelope Mackie also considers (**PII**) and its converse from the point of view of Leibnizian individual essence (Mackie 2006, 22). Let us see these arguments in turn. Diekemper's strategy for defending non-qualitative thisness with events centers around seeking sound counter-examples to the Leibnizian Principle of the Identity of Indiscernibles. His point is that by demonstrating the numerical distinctness of Leibniz-indiscernible individuals with the help of such counter-examples, there opens the path to account for their distinctness in terms of a primitive, non-qualitative thisness (Diekemper 2009, 260). This treatment of Leibniz-indiscernible individuals seems to be one of the main merits of the haecceitistic approach. I think Diekemper's point can be conceded with the provisio that it applies only to particulars that are merely tokens of some common type; typically, mass products qualify as such particulars. However, individuals with their unique essences do not illustrate the failure of (**PII**). One can feel here the need for spelling out the distinction between mere particulars and individuals; this will come in the next section.

It is also implied in these discussions that the qualitative view of individuals, presumably fleshed out in terms of bundles of qualitative features, is vulnerable to the converse of (**PII**) since any change in the bundle or set of properties would, according to the converse of (**PII**)**,** yield a different individual. Penelope Mackie explores this sort of critical

[7] It is noteworthy that the way Diekemper describes nonqualitative features does not match the standard approach in the literature. Briefly, he identifies nonqualitative features with impure qualitative features. See Diekemper (2009, 256). One can consult the literature about the distinction between purely qualitative properties and impure ones: see Adams (1979, 7), Armstrong (1978, 146-147; 2004, 13), Loux (1997, 128, footnote 19). They all point out that a qualitative property mixed with referential devices essentially occurring in the reference to that property qualifies only as an 'impure' qualitative property. While 'being the son of a king' is a pure relational qualitative property, 'being the son of an Anjou king' is an impure relational qualitative property. Now Diekemper's example for nonqualitative property exactly matches the examples given by these authors for impure relational qualitative properties.

reasoning. Her starting point is the Leibnizian notion of individual essence which is characteristically super-essentialist. As she remarks, 'Leibniz holds that *every* property of an individual is essential to it'. This amounts to the position of 'super-essentialism' which is standardly attributed to Leibniz with respect to the notion of individuals. Mackie then goes on to show how this essentialist notion of individuals naturally gives rise to the Leibnizian identity condition in terms of complete property-indiscernibility governed by (**PII**). 'The Identity of Indiscernibles guarantees a non-trivial distinction between the properties of any two individuals. But the consequence of this view – that of rendering false all claims to the effect that anyone or anything could have had a history in any way different from the actual history – is one that, for obvious reasons, most philosophers find unacceptable' (Mackie 2006, 22). Though Mackie refers here to the 'Identity of Indiscernibles' (**PII**), obviously, what she takes to be 'the consequence of this view' is the consequence of the converse of (**PII**). The latter states that the identity of things rules out any change in their properties; a claim found unacceptable by philosophers who are concerned about identity through worlds/times.

While I do not deny problems with (**PII**) and its converse in metaphysical contexts, this is not the last verdict on the issue. If it were, the qualitative view would be doomed to failure, and even the explanatory reading with the ontic notion of explanation could not help to save it. Mackie is right in claiming that Leibnizian super-essentialism and the converse of (**PII**) are tailored for each other; but it does not follow that qualitative individual essence should be taken as coinciding with the complete set of properties of an individual. In fact, the individual essence does not cover the whole qualitative space for locating the individual. For example, some Leibniz scholars take efforts to point out that the complete notions of the individuals contain 'incremental' features as well apart from their essential features (Grimm 1970, Ishiguro 1979). The incremental features comprise the spatio-temporal location and other contextual aspects of the individual according to Grimm. Further, the individual essence *within* the qualitative space characterizing an individual need not be seen as a set of features whose members are supposed to be listed in agreement with (**PII**). Fortunately, the qualitative view has the appropriate resources to avoid the forced track of (**PII**). First, as I have said, the qualitative character of individual essence does not have to be captured by the Leibnizian complete notion of the individual. Recall that Leibnizian completeness is understood as the completeness of the set of the properties of the individual. Second, the qualitative character need not to be fleshed out in terms of a set-theoretical construal, since the structural aspects of the individual essence require a different approach. The qualitative account recommended here avoids falling prey to the vulnerable set-theoretical construal: i. e., that even a trifling change in the

set of properties amounts to shifting to another individual. This uncomfortable consequence follows only if a naïve bundle theory is adopted with set-theoretical identity conditions.

It is a false dilemma, however, that either one opts for qualitative individual essences, - but then one has to buy into a naïve bundle theory - or, in view of the failure of the naïve (**BT**)**,** one is compelled to accept the haecceitistic account. In fact, there are more options. For example, the friend of the qualitative account can embrace a more sophisticated, two-tier version of the bundle theory which is not vulnerable to the risks of the Leibniz Principle, or, alternatively, (s)he can appeal to the Aristotelian notion of nature or essence.[8]

As to the first option, a core of essential features can be posited in the relation of tight bundling going proxy for an individual nature. Simons has adopted the notion of bundling from Husserl and he made an amendment to that notion by replacing Husserlian bundling on the specific level with tropist bundling on the individual level. The result is a two-tier (**BT**) with a qualitative nucleus in the core and accidental features in the outer fringe (Simons 1994). The crucial point is that the qualitative features in the tight bundling are not to be analyzed with the Leibniz principle.[9] The same can be vindicated, perhaps even more directly, for Aristotelian natures leading to the second option. Here, again, the claim is that Aristotelian natures or essences, while qualitative, cannot be resolved into sets of qualitative features. In fact, a distinction can be made between a mere *list* of the properties and the qualitative essence (Oderberg, 2011). The details of the non-set theoretical construal of individual essence come a bit later. So, we can anticipate the conclusion that the tenability of the qualitative account of individual essence and individuation does not hinge on the tenability of (**PII**).

2. Individuation vs Particularization

By 'individual' I do not mean just an arbitrary token of a type. Though instantiation captures the tokening of the type, it cannot serve as a clue to the individuation of the particular. The reason is that instantiation of a type yields only a bunch of particular tokens such that each token indiscriminately illustrates the type. In other words, each token is just a particularizer of the type. But particularization itself is indifferent to the issue which of the tokens of the relevant type fills the slot since any arbitrary token would do the same job. Further, 'individuation' of the par-

[8] It is worth noting that the Aristotelian natures are natures of kinds, that is, they are about specific essences. The present suggestion is to explore the Aristotelian approach to individual natures as well.

[9] The claim that the bundle theory is not committed to **PII** is discussed by Matteo Morganti (2011, 37-40).

-ticular is ambiguous: it can mean the process or the result of securing unique reference to a particular or, alternatively, it can mean the grasping of the particular through its unique nature. The supposed unique nature is vital to counterfactual discourse and explanation about the individuals. Obviously, 'uniqueness' is a requirement on both readings: but, on one reading, only *referential uniqueness* is achieved while, on the other reading, the goal is to secure the *uniqueness of the individual nature,* that is, attributive uniqueness. In short, the referential/attributive ambiguity, very much familiar in the semantics of definite descriptions, is operating behind the metaphysical issues of individuation.

The ambiguity inherent in the very notion of individuation is captured by Lowe in a similar vein. He recommends taking 'individuation' as an ontic notion, rather than as an epistemic one. He notes that corresponding to the two notions of individuation there are two principles of identity: the one which yields criteria *for us* to distinguish entities, while the other principle hangs together with 'genuine individuation'. According to the latter principle, the identity of things 'is received from that what makes them the very things what they are', as Lowe quotes Locke. While agreeing with Lowe about this ontic notion, I am inclined to take the first as purely *referential* identification, used, in the first instance, for telling things apart. However, Lowe's 'genuine individuation' targets the specific essence of things and not their individual essence. When explicating individuation as a 'determination relation between entities' Lowe says as an illustration, that ' x determines or "fixes" .... *which* of its kind y is' (Lowe 2014, 216). So, individuation, for him, is the selecting of a particular token from among other tokens of a given specific type. It seems to me that this approach tells us about the *particularization* of the universal type rather than accounting for the individual as such. The latter task, however, requires separating individuation from mere particularization.

There is still a temptation, at least in the Platonic tradition, to think of particulars merely as instances of types or species. The question, in that tradition, is what it is to be the particular instance of a type or species; and the answer, according to Boethius, is that the particular instantiates the species by virtue of 'dividing' it. Say, particular rabbits divide the species 'Rabbit'. An equally important other question would be what makes some of the particulars genuine individuals, over and above the mere instantiation of the species. This question is typically ignored, or, even worse, individuality is practically identified with particularity as it can be illustrated by Jorge Gracia's account of individuation. For him 'particular' and 'individual' are 'coextensive', and he devotes a book to defending this position. The extensional overlap between being a particular and being an individual suffices for explanatory purposes as well according to him, since he explicitly says that 'there is no great advantage in making a distinction between particularity and individuality'

(Gracia 1988, 54).[10]

Gracia's effort is directed to draw the line between universals and particulars in a realist way which is obviously a legitimate enterprise; but it falls short of being a complete metaphysical theory of individuals.

Obviously, for a thing to be a particular, and for it to be an individual, are two different aspects that can be told apart. The question arises then, what account can be provided for individuality, once particularization is well-explained by the realist theory of instantiation. I think genuine individuality requires the notion of qualitative individual essence and individuation as supported here. The account that can fill the slot explores a 'thick' notion of individuals as opposed to the 'thin' notion of being an instance of a species, i. e., a mere particular. How 'thick' the notion should be is a further question. Presumably, not as thick as for Leibniz for whom only the complete set of the properties is individuating. It would be safe, though a truism, to say that the qualitative account of individual essence and individuation should avoid the Scylla of bare haecceity, and the Charybdis of the Leibnizian complete-concept notion of individuals.

3. Roles of Haecceity and Qualitative Individual Essence Compared

So far, I have been arguing for the qualitative view of individual essence. If one visits though the functions/roles attributed to haecceity and the functions/roles attributed to non-haecceitistic individual essence in the literature, one finds a substantial overlap between the two lists. A comparison will be illuminating.

Rosenkrantz records the various functions of *haecceity* (Rosenkrantz 1993). These are the following: 1. As a primitive thisness, it helps securing identity through worlds (see, also, Adams' account); 2. In its semantic role, it turns *de re* discourse into *de dicto* eliminating thereby the problematic *de re* locutions (see Plantinga 1974); 3. In its epistemic role, discussed by Chisholm, the special status of self-knowledge is explained by grasping one's own haecceity; 4. It functions as the intension of proper names (see Plantinga 1974; Chisholm 1976).

Losonsky summarizes the functions of *individual essence* (Losonsky 1987). His item 1' is the same as 1. at Rosenkrantz, i. e., to secure transworld identity. 2' says : individual essence fills the slot at those worlds where a certain individual fails to exist. Its role is to match *de re* claims with the contingency of empirical existence. Say, Socrates is essentially/necessarily human, but he does not exist in every possible world. However, his individual essence, as an abstract and necessary existent, exists at those other worlds where he fails to exist. 2' has no equivalent in the haecceity-list. 3' says: individual essences as qualitative

[10] On Gracia's position see, also, Ujvári (2017).

bundles illustrate the property-bundle view of individuals and, consequently, they fall under (**PII**). This use has no matching item - a pendant - in the haecceity list. 4' matches with 2 and 3 of the haecceity-list: i. e., 'to account for the knowledge we have of ourselves and other persons'. 5' says: individual essence is used as the principle of individuation: historically, the principle was applied to continuants, and recently the application is extended to events as well. 6' says: individual essences are used as senses of proper names. This semantic function has its pendant in the haecceity-list under 4.

The survey shows that haecceity and qualitative individual essence as metaphysical posits are both supposed to explain: a) trans-world identity, b) self-knowledge and other *de re* attributions, and c) the intension of proper names. Assessing the functions, one finds that securing transworld identity is the strongest support for both posits. The essential nature of the individual without which it would not be what it is, saves identity-claims from the fragility of accidental changes. Haecceity can achieve the same goal albeit with relying on direct referential devices instead of qualitative aspects. The idea is that a thing can never lose its haecceity, and we can keep track of the same individual with referential devices.[11] As to the role they are supposed to play in self-knowledge and also in turning other *de re* attributions into *de dicto* claims, I am rather skeptical.[12] It seems to me that the reductive move of eliminating *de re* locutions is a misapplication of haecceity and/or individual essence since both are the vehicles of *de re* metaphysical claims. As to their involvement in self-knowledge, I do not see any reason why they should be involved. We enjoy privileged access to our mental items and their ownership, but it would be false to think that there must be something with entitative status that is responsible for this immediate grasp. And the last point, motivating the posit of both haecceity and individual essence,

---

[11] It seems to me that the distinction between the 'strong' and the 'weak' versions of haecceitism suggested by Cover and Hawthorne can help the haecceitist in accounting for this function. According to them strong haecceitism does not supervene on general propositions in explaining transworld identity, while weak haecceitism accepts the role of some general propositions in the explanation of transworld identities. See Cover and O'Leary-Hawthorne (1999).

[12] The reductive move is conceived by Rosenkrantz in the following way: when a person S directly attributes F-ness to an object X, S grasps the conjunction of the haecceity of X and F-ness. Thus, one can eliminate de re beliefs in favor of de dicto beliefs with the help of haecceity (See Rosenkrantz 1993, 34). Haecceity is claimed to have a role in self-knowledge as well: it is Chisholm who makes a Kantian start by noting that first-person reference is indispensable to knowledge of external things and that reference is secured via grasping one's own haecceity. Thus, haecceity becomes, in his account, a precondition of knowledge of external things (Chisholm 1981). However, I think that the necessary self-ascription of knowledge-claims with Kant is not an epistemological K-K thesis. Rather, it is a transcendental precondition of knowledge in general; thus, it is a de dicto claim and not a de re one.

is that they serve as intensions of proper names. I am not going to get into the semantic issue.

It has become clear from the survey that the explanatory role has not been taken into account either by Losonsky or Rosenkrantz. However, with the explanatory role in mind, the significance of the notion of qualitative individual essence increases.

4. Aristotelian Natures, Individual Essence and Existence

In one version of the qualitative view individual essences are closely connected to the property-bundle construction of individuals (Plantinga 1974; Forbes 1985). As I have said earlier, the qualitative aspect of individual essences is not to be fleshed out in terms of the set-theoretical construction of the naïve bundle-theory taking every qualitative feature homogeneously, as equally contributing to the identity of the individual. This approach is vulnerable even to trifling changes in the members of the bundle or set: this is a familiar defect of the set-theoretical construction. As a result, excessive essentialism ensues with the naïve bundle theory. In view of this problem it seems more reasonable to hold that qualitative individual essences are either tight bundles of essential properties with the special gluing relation of mutual foundation, as it was suggested in the first section, or, one might hold that they are natures in the Aristotelian sense. Let us consider now the virtues of this latter.

Nature or essence in the Aristotelian sense, though qualitative, is clearly not identifiable with a set of properties. Aristotle seems to be aware of the difference. He says, that by giving the essence of a thing the 'what' question is answered; but this important question is not answered by specifying some, or all, of the properties of the thing. He writes in the *Metaphysics* Book Z: 'when we say of what quality a thing is, we say that it is good or beautiful ... but when we say *what* it is, we do not say "white" or "hot" or "three cubits long", but "man" or "God"' (Aristotle 1984, 1028.15.).

The Aristotelian notion of essence is the specific essence of substance-kinds captured by definitions: 'there is an essence only of those things whose formula is a definition' (Aristotle 1984, 1030.5.). So, a definition captures the specific essence of a thing since a thing is *what* it is, by its specific essence. What about those other properties, including propria and accidents, that are not covered by the definitions? According to Aristotle these features have less substantial roles. While things 'participate' in these other qualitative features, they do not have their essences in this way: things do not 'participate' in their essences since they are what they are just by their specific essence. As Aristotle puts it: 'nothing, then, which is not a species of a genus will have an *essence* – only species will have it, for in these the subject is not thought to participate in the attribu-

-te' (Aristotle 1984, 1030.10.). It is clear, then, that for Aristotle, to have a (specific) essence is not tantamount to participating in a set of features. I would say that the same applies for individual essences. These, while qualitative, cannot be *analyzed compositionally* for the same reason that specific essences cannot be so analyzed: they are natures, and not simply sets of properties.

Non-compositionality does not mean, however, that the usual significance-conditions are disobeyed, and that we cannot meaningfully talk about the properties being entailed by the specific and/or the individual natures. As to the latter, one can say that an individual essence or nature entails each and every essential property of the individual including the specific essential properties. I can only agree with Plantinga who endorses such entailment (Plantinga 1974, 73; 2003, 56). But it does not follow that the individual nature itself could be analyzed reductively in terms of the *conjunction* of the component essential features. This point is important if we consider that the functional/teleological unity of the whole individual is under-determined by the mere conjunction of its component features. A conjunction is neutral with respect to the structure responsible for such unity. An anti-reductive claim, similar to mine, is put forth by Loux with respect to kinds or specific essences in his 'substance theory of substance' (Loux 1997). Recently, Oderberg defended the same point, by making a distinction between a mere *list* of properties and (qualitative) essence. He says that 'an essence is more than a list: it is a structural, organizational unity'. He takes this distinction between properties and essence 'a key feature of Aristotelian essentialism' (Oderberg 2011, 99).

To note, the compositional analysis, discarded here, is accepted by Rosenkrantz. He says, that 'a conjunction of a haecceity and a universal essential property is an individual essence' (Rosenkrantz 1993, 43). However, I have doubts as to whether individual essences can be resolved into such conjunctions.

A further feature to be considered is the dependence of individual essences on the existential condition since it would be highly problematic to postulate individual essences for merely possible individuals. Let us visit now the existence-claim from the perspective of individual essences. While specific essences may remain unexemplified, at least in the Platonic tradition, most authors agree about the requirement that individual essences should be exemplified. In short, essence goes with existence in the case of individual essences. Even Plantinga, construing individual essences in a Platonic way, claims that 'an essence is necessarily instantiated in some world or other' (Plantinga 1974, 76). So, individual essences must be exemplified somewhere. But the contingency of existence is still preserved, on his account, since each world is such that it is not necessary for any individual essence that it should be exemplified there.

Still, there is something faulty with this solution that purports to provide for the contingent existence of concrete things. First, if instantiation is a trait common to specific essences and individual essences, as Plantinga holds, then the distinction between general existence-claims and singular existence-claims is obliterated. In this case, it applies to both kinds of existence-claims that to exist is to be instantiated. But the obliteration is hardly acceptable. The metaphysical difference that species are instantiable while individuals are not, has to be reflected in the logical distinction between the two kinds of existence. While common natures or forms have a threefold existence, at least according to medieval Aristotelians, as existing in themselves, and by 'informing' the things and the minds respectively, concrete individuals have only one 'mode of existence', and it is their bare existence. It would be mistaken to describe such existence as the *instantiation* of an individual essence.

Second, Plantinga's scheme implies a Leibnizian arrangement of existent beings at a world. The existent beings as the instantiations of their individual essences must be compossible. Thus, the individual essence of each individual must contain a clause about the essences of every other individual at a given world, precisely in a Leibnizian way. Plantinga incorporates such clause into his definition of individual essence. He says, that the individual essence *E* of Socrates must be such that it is essential to Socrates; that it incorporates all the other entailed essential properties of him; 'and finally, the complement of *E* is essential to every object distinct from Socrates' (Platinga 1974, 76). It seems to me that this clause invites an unnecessary abundance of individual essences. Moreover, their abundance is promoted by Cambridge changes. Say, if a new contingent being comes into existence, then the rest of the population at that world acquires a new essence: the complement of essence *E* of the newborn individual.

The upshot is that individual existence should not be conceived as the exemplification of an individual essence, rather, individual essences can be seen as having an existential precondition. I recommend here the existence-conditioned characterizations of individual essence given by Kit Fine, and developed by Roca-Royes (Fine 1994; Roca-Royes 2011). Though their position is formulated for the essential properties of individuals, it surely applies to individual essences as well. These properties are supposed to help 'answer the question "what is *a ?*"' where *a* is the individual whose nature is to be discovered. According to the characterization 'if *P* is an essential property of *a*, *a* could not exist without being *P'* (Roca-Royes 2011, 66). It is evident that being *P* is not tantamount to *P'*s being instantiated, for individual essences make sense only on the precondition of the existence of the individuals. As to the unrealized possibilities, we can say that the virtual realm of singular possibilities is something that actually does not obtain, rather than remaining 'uninstantiated'.

5. Qualitative Individual Essence and Relationality

I owe the reader the support of my claim that some particulars are 'genuine individuals'. After all, what makes some particulars to be so? Also, it is time to consider what features are entailed by the qualitative individual essences. Are their features all monadic such that a 'relational individual essence' sounds like a *contradictio in adjecto*?

Individual essence or nature is thought to belong so intimately to the thing that it is typically construed as an *intrinsic* feature. Obviously, it cannot be external to the thing; it cannot depend on its relation to other things. Thus, relationality seems to be undesirable in this context. Chisholm draws our attention to an ambiguity in referring to haecceity that seems to threaten with an undesirable relationality. '"That thing" could be taken in relational sense i. e. "the thing I am now looking at". When it is taken in this way, then, of course, it does not intend the individual essence or haecceity referred to' (Chisholm 1976, 35). Reference to haecceity obviously cannot be indexical reference though it takes the same form. Fortunately, the ambiguity cannot even arise in connection with individual essences of the qualitative sort since they can be referred to only non-indexically in virtue of their qualitative aspect.

So, the problem of indexical reference is not a real threat to qualitative individual essences, but the issue is still open whether relational properties in general can be entailed by them. I do not think that relationality should be dismissed across the board. The reason becomes clear if we consider what the 'genuine' individuals are. Here, I have talked about some particulars being 'genuine individuals' but the criterion for being 'genuine' was left open. It seems that this criterion cannot be afforded by inspecting carefully the physical world. A promising criterion, however, needs the acknowledgment of some relational aspects in the notion of 'genuine' individual. Persons are undeniably genuine individuals - except for those who deny the existence of composites, and admit in their ontology only particles arranged in a certain way. But once persons are acknowledged, then those works of art and engineering, or even natural phenomena, that are endowed with features projected by persons, should also be considered as genuine individuals. To be an individual is thus derivative from being a person. Applying the notion of grounding, personhood *grounds* our claims about individual essences/ natures. The relational character of attributed or *projected individuality* is not conceived here in a Humean manner. Rather, I think that to be a genuine individual is a feature determined ultimately by our prevailing cultural scheme. The only safe point to start with is that being a person entails being a genuine individual. As to the further question what other things count as genuine individuals in the derivative mode, the answer depends on the prevailing cultural scheme.

To sum it up, I have argued that the recent emphasis on the ontic notion of explanation increases the significance of qualitative individual essences. The explanatory-gap argument strengthened with the ontic variant supports postulating qualitative individual essences. It was claimed further, that their qualitative character is not to be conceived as the mere conjunction of properties; that is, it is not to be conceived in the way characteristic of (**PII**) and its converse. The alternative view recommended here takes individual natures or essences as structured, organizational units in the Aristotelian sense thereby exempting individual essences from falling under (**PII**) and its converse. Similarly, applying the two-tier bundle theories of individuals, instead of the naive (**BT**), would achieve the same result. It was also argued that individuation should be told apart from mere particularization of the specific nature. When comparing the roles of qualitative individual essence and haecceity, one finds that securing trans-world identity is the strongest argument in favor of both. However, if the explanatory role is considered, the notion of qualitative individual essence becomes increasingly important.

Acknowledgments

I am grateful for valuable comments and criticism of an earlier draft of this paper, read at a recent workshop in Neuchatel, to Fabrice Correia, Massimiliano Carrara, Christian Wüthrich, Shamik Dasgupta and Gonzalo Rodriguez-Pereyra. Thanks also to two reviewers for their valuable comments. I also thank Zs. Zvolenszky and A. Ryder for their helpful comments.

REFERENCES

Adams, R. 1979. Primitive Thisness and Primitive Identity. *The Journal of Philosophy* 76: 5-26.

Armstrong, D. M. 1978. A Theory of Universals. *Universals and Scientific Realism.* Volume II. Cambridge University Press.

_____ 2004. *Truth and Truthmakers.* Cambridge University Press.

Aristotle 1984. *Metaphysics.* Book Z. In *The Complete Works of Aristotle,* ed. J. Barnes, volume II. Princeton: Princeton University Press.

Chisholm, R. 1976. *Person and Object.* London: Allen and Unwin.

_____ 1981. *First Person.* Minneapolis: University of Minnesota Press.

Cleve, J. 1985. Three Versions of the Bundle Theory. *Philosophical Studies* 47: 95-107. reprinted in *Metaphysics. Contemporary Readings.* ed. S. D. Hales, Belmont, Ca: Wadsworth, 1999.

Correia, F. and Schneider, B. eds. 2014. *Metaphysical Grounding.* Cambridge University Press.

Cover, J. A. and O'Leary-Hawthorne, J. 1999. *Substance and Individuation in Leibniz.* Cambridge University Press.

Diekemper, J. 2009. Thisness and Events. *The Journal of Philosophy* 106: 255 – 76.

Fine, Kit 1994. Essence and Modality. *Philosophical Perspectives* 8: 1-16.

Forbes, G. 1985. *The Metaphysics of Modality.* Oxford: Clarendon Press.

Gorman, M. 2005.The Essential and the Accidental. *Ratio* 18: 276 – 289.

Gracia, J. E. 1988. Individuality. An Essay in the *Foundations of Metaphysics.* University of New York Press.

Grimm, R. 1970. Individual Concepts and Contingent Truths. *Studia Leibniziana.* Reprinted in *W. G. Leibniz: Critical Assessments*. 1994. Oxford: Routledge, ed. Woolhouse, R. vol. I. 308 – 329.

Ishiguro H. 1979. Substances and Individual Notions. In *The Philosophy of Nicholas Rescher.* Philosophical Studies Series in Philosophy, Sosa E. ed., vol. 15. Dordrecht: Springer.

Kaplan, D. 1975. How to Russell a Frege-Church. *The Journal of Philosophy* 72: 716-29.

Kment, B. 2014. *Modality and Explanatory Reasoning.* Oxford University Press.

Losonsky, M. 1987. Individual Essences. *American Philosophical Quarterly* 24: 253 – 60.

Loux, M. 1997. *Metaphysics: A Contemporary Introduction.* London: Routledge.

Lowe, E. J. 1999. The Possibility of Metaphysics. *Substance, Identity and Time.* Oxford: Clarendon Press.

_____ 2014. Asymmetrical Dependence in Individuation. In *Metaphysical Grounding*, F. Correia and B. Schneider eds., 214-233. Cambridge University Press.

Mackie, P. 2006. *How Things Might Have Been.* Oxford University Press.

Maurin, A. S. 2017. Grounding and Explanation: A Cautionary Tale. Conference talk at ECAP9, 21 – 26 August, Munich, Germany.

Morganti, M. 2011. Bundles, Individuation and Indiscernibility. *European Journal of Analytic Philosophy* 7: 36-48.

Oderberg, D. S. 2011. Essence and Properties. *Erkenntnis* 75: 85-111.

Plantinga, A. 1974. *The Nature of Necessity*. Oxford: Clarendon Press.

_____ 2003. *Essays in the Metaphysics of Modalities*. Oxford: Oxford University Press.

Roca-Royes, S. 2011. Essential properties and Individual Essences. *Philosophy Compass* 6: 65 – 77.

Rosenkrantz, Gary S. 1993. *Haeccceity. An Ontological Essay*. Dordrecht: Kluwer Academic Publishers.

Simons, P. 1994. Particulars in Particular Clothing: Three Trope Theories of Substance. *Philosophy and Phenomenological Research* 54: 553 – 575.

Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World.* Princeton: Princeton University Press.

Thompson, N. 2016. Grounding and Metaphysical Explanation. *Proceedings of the Aristotelian Society*, 116, 395 – 402.

Ujvári, M. 2013a. Individual Essence: gibt es solche? *Metaphysica* 14: 17 – 30.

_____ 2013b. The Trope Bundle Theory of Substance. *Change, Individuation and Individual Essence.* Frankfurt: Ontos Verlag.

_____ 2017. Haecceity Today and with Duns Scotus: Property or Entity? In *Perspectives on the Self*, B. Berčić ed., Rijeka: Faculty of Humanities and Social Sciences at the University of Rijeka, 331-340.

EuJAP | Vol. 13, No. 2, 2017
UDK
111 INWAGEN, P. VAN
165.023.1

SAVING THE SHIP*

JOHN BIRO
University of Florida

ABSTRACT

In defending the startling claim that that there are no artifacts, indeed, no inanimate material objects of the familiar sort, Peter van Inwagen has argued that truths about such putative objects can be paraphrased as truths that do not make essential reference to them and that we should endorse only the ontological commitments of the paraphrase. In this note I argue that the paraphrases van Inwagen recommends cannot meet his condition. Read one way, they lose us some truths. Read another, they entail the existence of the very objects they are supposed to rid us of. However, we need not share van Inwagen's distaste for the latter: to say that they exist is not to say that anything exists in addition to the simples composing them.

Keywords: *van Invagen, paraphrase, composites, simples*

As part of his argument that there are no composite objects, van Inwagen claims that statements appearing to assert or imply their existence, while allowable as true "in the ordinary business of life," are not *strictly* true, by which he means that they are not to be taken at face value when we are doing serious metaphysics. What are strictly true are paraphrases of such statements that do not make reference to anything other than the simples of which these putative are supposed to be composed.[1] If this is so, one welcome - at least to van Inwagen - consequence is that hoary old

[1] van Inwagen stresses that by 'paraphrase' he does not mean 'translation,' if the latter is understood as preserving the meaning of, or expressing the same proposition as is expressed by, the original sentence (1990, 112-3). All that is required is that they describe all the facts without essential reference to composites (1990, 113 *et pass*). The problems with his proposal to be canvassed in this note arise even if we understand paraphrase as he does.

*Received: 09.04.2016.
Accepted: 14.12.2017.

puzzles about their identity cannot as much as arise. For example, we can re-tell the Theseus story in terms of planks ("honorary simples," that is, simples relative to the putative ship they are thought to compose) "arranged ship-wise," without mentioning ships. Even though in the re-told story there are no ships, nothing philosophically important is left out (1990, 128-9). Furthermore, we do not have to deny that the ordinary, non-philosophical, description of the facts is true. Even though there are no ships, sentences ordinarily taken to be true about them can be allowed to be literally, if not strictly, true, albeit "perhaps…in some sense… misleading" (1990, 101).

In this paper I challenge van Inwagen's claim that his paraphrases "preserve *everything*" that is true of the thing, event or state of affairs described in the original, everyday, statement (1990, 129, van Inwagen's emphasis). I argue that the locution on which the paraphrases rely – 'simples arranged *x*-wise' – cannot be understood as referring either to the simples taken severally on pain of losing some truths or to their arrangement on pain of readmitting *x*s into our ontology.

In van Inwagen's metaphysically serious version of the story, what we have before the first plank is replaced is what he calls the First Planks. After one of these is replaced, we have the Second Planks, after one of *those* – one that had been one of the First Planks as well as one of the Second Planks – is replaced, we have the Third Planks, and so on. When each plank that had been one of the First Planks has been replaced, the First Planks are re-arranged in exactly the same way they were at the beginning of the story. The First Planks and the Last Planks are clearly not identical, since there is a plank that is one of the Second Planks and not one of the First Planks – and, again, so on. Since there is no mention of any ship, "there is no such question as 'Which of the two ships existing at the end of the story is the ship with which the story began?'" (1990, 129). In general, "If there are no artifacts, then there are no philosophical problems about artifacts" such as those that have exercised philosophers for millennia.

The viability of eliminating apparent reference to artifacts by paraphrase has not gone unquestioned. Rosenberg, for example, suggests that "…the notion of simples 'being arranged chairwise' is one that we can and do understand *only* to the extent that we understand references to ships *per se*" (702). Elder voices a similar misgiving: "…allowing that dogwise arrangement obtains at all is allowing that there are dogs" (132).

One may also wonder how 'This is a ship' can be literally true in any, even in a "loose", everyday, sense if it is not true in the strict sense. van Inwagen asks us to imagine Copernicus saying "According to my theory, the sun does not move. Nevertheless, sentences like 'It was cooler in the garden after the sun had moved behind the elms' can, when uttered in the course of the ordinary business of life, express truths" (1993, 684-685).

If we take Copernicus to be saying that, strictly speaking, the sun does not move, it is not clear what we are to take him to be saying when he says that the *quotidian* sentence expresses a truth. Hirsch makes the related complaint that he does not understand what van Inwagen means when he says that there are, strictly speaking, no apples (690-691). In saying this, is *he* speaking strictly or not? Is Copernicus?

Here I offer a different objection: contrary to what van Inwagen says, his proposed paraphrases do not preserve all the everyday truths we – and he – accept about artifacts such as ships.

What can we take 'the First Planks' etc. to refer to? Suppose we take the plural noun in the expression at face value and treat the expression as having plural reference to the planks that are arranged ship-wise. These are van Inwagen's "honorary simples" that stand to the putative ship as do the true simples, whatever they are, to the planks themselves (and as would the relevant simples to any other putative composite material object). If we understand the expression in this way, the claim that everything true in the standard version of the story remains true in the re-telling becomes hard to maintain. The ship is bigger than the planks – even when these are arranged ship-wise – but the planks, even when so arranged, are not bigger than the planks not so arranged. (More on this below.)

Suppose, instead, that we take the capitals seriously and treat 'the First Planks' as a proper name, ignoring the plural. What can it be thought to name? The only thing in the offing is the ship-wise arrangement of planks we have. Trouble again: the planks of the ship, as we ordinarily say, are nailed to each other but the ship-wise arrangement of planks is not nailed to anything. In any case, *arrangements* of planks are just what van Inwagen is in the business of eschewing. It is essential to his strategy that whatever truths there are be truths about planks, not about arrangements of them. Once we let arrangements in, we may as well call them ships.[2]

van Inwagen's own understanding of the expression 'the First Planks' is as "a rigid plural designator, like 'the British Empiricists'" (1990, 128). The idea is that the latter does not refer to the individual philosophers so grouped, nor to an additional entity, the group they form; in the same way, 'the First Planks' should not be taken to refer either to the individual planks that are arranged ship-wise or to an additional entity, the ship they

[2] As Uzquiano observes, "…if propositions expressed by ordinary statements apparently concerned with ordinary material objects like bricks and tables and chairs turn out to be propositions concerned with sets of simples under certain arrangements, then one should consider *identifying* objects like bricks and tables and chairs with sets of simples under certain arrangements" (446, my emphasis). Goldwater advocates precisely this. However, doing this is not open to van Inwagen. He cannot be seen as saying what ordinary objects are, given that he thinks there are none. It is one thing to be a reductionist, quite another to be an eliminativist. (See also fn. 6 below.)

compose. If this is right, the properties of individual planks, including the property of – as we say – being nailed to other planks. are irrelevant, and thereby so is the fact that the First Planks lack(s?) that property.

It is not clear that the expression 'the British Empiricists' behaves in the way van Inwagen claims. What exactly does 'the British Empiricists' refer to? Do I, when I use it, really leave behind the properties of the individuals? Not if I say that the British Empiricists were English, Irish, and Scottish, respectively: the property of being English or Irish or Scottish is a property of the individuals Locke, Berkeley, and Hume. In a similar way, if I say 'the First Planks came from trees in Norway,' I must be taken to be referring to the individual planks. This is especially clear if we add that they came at different times, some in one shipment, some in another. Note that I could have said 'were cut from trees in Norway,' instead of 'came from Norway.' The plural 'were' would have made it clear (as 'came' does not) that the property in question – having been cut from a tree in Norway – is being attributed to the planks individually. Of course, since that is what is in question here, we should not put things in a way that begs the question. The point can be made, however, even if we use the number-neutral 'came.' The property of having come from Norway is as much a property of the individual planks as is the property of having been cut from a tree in Norway. So, too, is the property of being nailed to other planks.

It may seem that the property of having come from Norway is a property of the plank*s*, as much as of each plank. So, then, is the property of having been cut from tree*s* in Norway, one that the plank*s,* as well as individual planks possess. But it is not clear what sort of property that could be. How does one cut a number of planks from a number of trees, except by cutting individual planks (one or more) from individual trees?[3]

The capital 'E' in the third word in 'the British Empiricists' signals that the expression refers to a particular group of philosophers, namely, the group comprising Locke, Berkeley and Hume.[4] Without it, the expression would be naturally taken to refer to whatever philosophers are British and share a certain outlook – Mill certainly included. But a group is a single thing, a composite, and as such is no more acceptable to van Inwagen than is an arrangement.

van Inwagen's 'rigid plural designator' hovers uneasily between the plural and the singular. It is worth noting that he himself insists that *tertium non*

[3] If Smith, Brown and Jones are students in the class, each one has the property of being a student in the class. Who has the property of being students in the class? Not Smith, not Brown, not Jones, only the trio – which is, again, one thing.

[4] As in the title of Johnathan Bennett's book.

datur. He criticizes those who think that the thesis that the whole is the sum of its parts can be expressed by 'A whole *is* its parts' precisely for that reason. He says: "This sentence seems to me to be syntactically radically defective. There is the predicate 'is identical with', which yields a sentence when flanked by singular terms or singular variables. There is the predicate 'are identical with', which yields a sentence when flanked by plural referring expressions or plural variables. I do not see how there could be any sort of "identity" predicate that yielded a sentence by being put between a singular term (or variable) and a plural referring expression (or variable)" (1990, 287)[5]

What, then, are we to make of van Inwagen's claim that we can paraphrase 'Some chairs are heavier than some tables' as "There are *x*s that are arranged chairwise and there are *y*s that are arranged tablewise and the *x*s are heavier than the *y*s" (1990, 109)? If the *x*s and the *y*s are all simples, albeit arranged differently, it cannot be that some of them are heavier than others. Presumably, all simples weigh the same. For this not to be so, some of them would have to have some property such as being of a different size, or of different density, than others, properties only composites can have. I can see no other way of understanding the last clause of the proffered paraphrase than as short for '(some of) the *x*-wise arrangements are heavier than (some of) the *y*-wise arrangements.'

But, as already noted, a chairwise arrangement of simples, being a composite as much as a chair is, is not something van Inwagen can allow. Such an arrangement is, in fact, nothing other than a chair.[6]

What gives van Inwagen's proposed paraphrase (and with it, his rejection of inanimate composites) plausibility is that 'the planks arranged shipwise' un-hyphenated is easily taken to refer to something that, while it has all the properties of a ship, is nevertheless not a single thing, hence not a ship. But we cannot have it both ways. The referent is either one thing – a planks-arranged-ship-wise – or many things – the plank*s*, even if, as it happens, arranged ship-wise. If the former, we still have ships. If the latter, we lose some truths.

---

[5] Uzquiano argues that the paraphrase strategy cannot handle "apparent plural quantification over composites': "…singular quantification over composites can be paraphrased as plural quantification over simples, but plural quantification over composite cannot be…". For another interesting discussion of plural reference and plural quantification see Yi (2014).

[6] van Inwagen offers, as an alternative: "There is an *x* such that *x* is a set and the members of *x* are arranged chairwise and there is a *y* such that *y* is a set and the members of *y* are arranged tablewise and the members of *x* are heavier than the members of *y*" (110). But a set of simples arranged chairwise is, surely, a composite, composed of its members, and as such should be off limits for van Inwagen. And, again, it is presumably the sets that differ in weight, not their respective members. Thus the property of being heavier than is a property of some of the sets of *x*s arranged chairwise – of some of the chairs, as we can surely say.

Consider a simple case where we have only three simples, A,B, and C. They may be arranged in six different ways: ABC-wise or ACB-wise, and so on. But the property of being arranged in one or the other of these ways does not distribute: it does not belong to A, or to B, or to C, any more than does the property of being large to the individuals who make up a large crowd. Contrast the property of being silent if the crowd is silent: each person in the crowd had better be, if 'the crowd is silent' is to be true. So with 'the planks are here,' 'the planks are too short' etc. distribute: each plank had better be here or be too short. However 'the planks are arranged ship-wise' does not distribute. While a hundred planks may be arranged ship-wise, no single plank can be so arranged. Nor, for van Inwagen, can being arranged ship-wise be a property of an aggregate or collection of planks: once we let these in, we can hardly avoid thinking of these as being composed of planks in the way van Inwagen does not allow putative composites such as ships to be. Furthermore, such an aggregate is as much a single thing as is a single plank (or a ship), so *it* cannot be arranged any more than a plank can.

Suppose that as we proceed with our repairs, each wooden plank is replaced by a metal one. Metal planks do not float. To say that they do when they are arranged ship-wise is to commit a double mistake. The first is to suggest that being arranged distributes, so that each metal plank has the property of being arranged ship-wise. The second – even if a single plank could have the property of being arranged ship-wise – is to think that that property would be sufficient to make it such that *it* can float. Conversely, the planks are a hundred in number, but the ship (a-hundred-planks-arranged-shipwise) is not. It was the Titanic that was thought to be unsinkable; no-one thought that any of the sheets of metal which when riveted together made it up were.

I said earlier that the property of being arranged does not distribute and thus cannot be a property of the planks taken severally. Nor, as already noted, can it be a property of the set or aggregate of the planks, since it makes no sense to say of a single thing, even if that thing is a set or aggregate, that it is arranged in a certain way. Only a plurality of things can have the property of being arranged.

It may be suggested that being arranged ship-wise can be a property of the planks in something like the way as being arranged in a circle may be said to be a property of the chairs in the room. It may seem that here we are attributing the property neither to the individual chairs nor to a single thing, their collection or set: we are saying neither that any single chair is arranged in a circle nor that there is a circular thing the chairs compose that is. Indeed, there is no single concrete thing the chairs compose. However, not every way simples (or honorary simples) may be arranged yields an object of a familiar sort. If the chairs are, in addition to being arranged in a circle, attached to each other — as is not uncommon with

rows of chairs in an auditorium – the fact that we have no ready label for what we then have does not mean that what we have is not a single thing.[7] And it is not difficult to make up stories about how we could find it useful to recognize it as such and coin a name for it.

Not only that: when we say that the planks are arranged ship-wise, there seems to be no way to express the property we supposedly have without mentioning ships.[8] With our chairs, we have serviceable grammatical paraphrases: we can say that they make a circle – an abstract object – even if they compose no recognized circular material object. But no such paraphrase is available with our planks: if, being arranged shipwise, they make a ship-shape, we have not an abstract object but a ship.

There are any number of locutions we understand perfectly well that resist the kind of paraphrase van Inwagen recommends. We speak of the ship's planks, of the ship and its planks, of the ship and the planks that compose it, all perfectly fine. By contrast, 'the planks arranged ship-wise's planks,' 'the planks arranged ship-wise and their (?) planks,' and 'the planks arranged ship-wise and the planks that compose them (?)' resist parsing. Put in the hyphens, and all is well. But now we have our ships back.

Things are equally clear in van Inwagen's leading example of the fort (as we say) built of sand by legionnaires. If it is true that the fort is twenty feet high and can withstand an attack, it must be true that the grains of sand arranged fort-wise are (?) and can. None of the grains of sand (our honorary simples here) are or can. So, 'the grains of sand arranged fort-wise' must be understood as referring to a fort-wise arrangement of grains of sand. That thing *is* twenty feet tall and *can* withstand an attack. But that thing is a fort, even if calling it grains of sand arranged fort-wise instead of a grains-of-sand-arranged-fort-wise (or a fort-wise arrangement of grains of sand) obscures this. The first four hyphens in the latter, and the indefinite (or definite) article they licence, make all the difference.[9]

---

[7] Even if they are not attached to each other, we may have a single thing if we think of them as, so arranged, constituting an object of interest to us, one, perhaps, with a religious significance. Some think that we would then have a scattered object, as in the familiar example of a watch disassembled for repair or a pipe for cleaning (e.g., Cartwright (1975) and Smart (1973)). For resistance on this, see Biro (2017a).

[8] Recall Rosenberg's and Elder's complaints.

[9] The point is a general one: arrangements have properties that the things arranged lack, and vice versa. This is so even if, unlike in the case of the ship and its planks, it is the same property that is in question. At the county fair, your roses may fail to win first prize, yet your arrangement of them may do so, or the other way around. The property of being the most beautiful at the show may attach to the arrangement or to the things arranged (or, of course, to both.)

Van Inwagen insists that in building (so to speak) a fort or a ship we do not bring anything into existence. Just so, if by that is meant that we did not add to the world's stock of simples. But when one says that we have brought a fort or a ship into existence one is, obviously, not saying *that.* One is saying that we have made certain propositions true, for example, that the world now contains a fort or a ship, something it did not do before. It does so because the appropriate (honorary) simples have been arranged in the appropriate way, in the shape of a fort or a ship. So arranged, they *are* ship or fort, and there are now truths about forts or ships when there were none before. When van Inwagen says that we have merely "rearranged the furniture of the earth without adding to" (van Inwagen 1990, 124) he is using 'furniture' to refer to the simples. Once again, we cannot take him to be talking about individual simples, since these cannot be *re*arranged, since they were not – and could not be – arranged in any particular way before. A certain number of them must have been arranged X-wise if we are to talk of something's being *re*-arranged so that that something is now arranged Y-wise. And if all we have to start with are the simples, to arrange a number of them Y-wise is to make something, namely, a Y.

It sometimes happens that the host is one bed short for the guest who sensibly prefers not to drive home after the party. Never mind – we can push two capacious armchairs together. We have rearranged the furniture – this, time literally. Have we made a bed? What matters is not what we call what we have made but that there is now something there was not before, something for our guest to spend a tolerably comfortable night in. Neither of the armchairs (our honorary simples here) was like that before the two were pushed together. To say that there is now a bed is just to say that there is now something that has a certain property, can serve a certain function. For that to be the case, we need not have *added* to the furniture. *Of course*, we did not do *that.* There are just as many (honorary) simples as there were before, and there is nothing in addition to them. It is they that are now a bed.

According to van Inwagen, what I have claimed are truths about ships, forts, beds and mountains are merely apparent truths, truths really about simples.[10] But if we accept the paraphrasability-without-loss criterion, we have to conclude that this cannot be so. If there are even loose-talk truths about something that are not reducible to loose-talk truths about anything else, that thing exists.[11] But to say that it does is not to say that the thing is something in addition to the (honorary) simples that compose it; thus to

[10] There are, he allows, genuine truths about living things and abstract objects, neither under discussion here.

[11] This is not to say that the thing is real as opposed to being fictional. The principle holds within fiction, too.

say that there are composites is not to say that these exist in addition to whatever ultimate simples there are. Since there are truths about ships distinct from any truths about the planks that compose them, truths about composites in general in addition to whatever truths there may be about simples, composites, natural or artifactual, exist, and they do so strictly speaking. Van Inwagen seems to assume that to hold this is to hold that there are such objects in addition to simples. But in arguing that this is not so, he is tilting at windmills.

It may be objected that I have overlooked the fact that, as van Inwagen insists, all he needs his paraphrases to capture is what is strictly true, and since some of the things we ordinarily say about ships – those that imply their existence – are not strictly true, they do not need to be saved. If we can say something that *is* strictly true that captures all the facts without saying what we ordinarily say, the "truths" of the latter need not be captured by our translation. This may seem plausible with respect to the Copernicus example. We know that it is not true, strictly speaking, that the sun moved behind the elms. Of course, in everyday discourse we continue speaking as before. But we can say what *is* strictly true by replacing the everyday description with one that does not carry the same commitments: we can say that the rotation of the earth has brought the elms to be between us and the sun. However, no such paraphrase is available for ordinary utterances about ships and the like, and no scientific discovery can make one available in the way that Copernicus' did for what we said about the garden. It is not that we lack some knowledge about what ships really are such that if we had it, we would no longer need to speak of them. We know that they are composed of planks or the like; knowing this, we still have to speak as we ordinarily do to say what we want to say of *The Queen Mary* when we say that it weighs 81237 tons. Thus there is an important disanalogy between what we are imagining Copernicus as saying and what van Inwagen says about ships. With the former, we are told that we were mistaken about the facts. The latter merely proposes a new way to describe them. We are, post-Copernicus, willing to say that the sun did not move, but we are not, post-van Inwagen, willing to say that the ship has not sailed.[12]

While we have reason to say of what we said in the garden that it was not strictly true, we have no reason to say this of what we say about *The Queen Mary*'s tonnage. So, to say that our paraphrase need not capture it, since it is true only in a loose sense is to beg the question. There has to be an independent reason for thinking that it is not strictly true, beyond the fact that it eludes our favoured paraphrase.

One reason van Inwagen offers is that if we accepted that the legionnaires built a fort by pushing the grains of sand around, we would have to accept that whenever we alter the shape of some collection of simples, we

[12] Korman (2009) makes a similar point.

create a new object. His example is kneading a lump of soft clay absent-mindedly into "some complicated and arbitrary shape. Call anything essentially of that shape a gollyswoggle… [i]f you can make a gollyswoggle by accident by kneading clay, then you must, as you idly work the clay in your fingers, be causing the generation and corruption of the members of a compact series of objects of infinitesimal duration. That is what seems to me to be incredible" (van Inwagen 1990, 126). But is it? At time *t* we have an object with shape *g*, at t_1, an object with shape *h* (call it a hollyswoggle), and so on. We have, of course, no interest in such objects – though we might have, if, say, we suddenly noticed one and found it particularly beautiful or in some other way remarkable. But even if this does not happen, there will be a true proposition 'there is a gollyswoggle,' at time *t*, a true proposition 'there is a hollyswoggle' at t_1, another – 'there is a jollyswoggle' – at t_2, and so on. I see nothing puzzling in this.[13]

van Inwagen also thinks that the only way to avoid the – to him as unwelcome as to me – consequence that two things, a statue and a piece of clay, can be in the same place, is to say that statues exist no more than do gollyswoggles. After all, the shape of the lump of clay that the sculptor formed into the statue changes every instant, if only ever so slightly and even if not by the sculptor's action but the weather's. So, as time passes, we have a succession of statues (let us suppose, to piggy-back on Gibbard's (1975) well-known example, Goliath at time *t*, Holiath at t_1, and so on – perhaps, eventually, no statue). If that means that at every instant we have two things, one of the statues and the lump of clay, Lumpl (since the latter does not go out of existence with every change of shape, as each successive Goliath_n is thought to do), we do have a puzzle. But that is a different puzzle from the one that puzzles van Inwagen.[14]

---

[13] While one can stipulate, as van Inwagen does here that every change in the shape of a piece of clay yields a different object, that is not the way we usually think of pieces of clay. Arguments for co-incident entities typically turn on the opposite assumption, namely, that the piece of clay can survive being deformed, while the statue cannot. Even statues are not essentially of the shape they are, as they can survive changes of shape such as losing an arm.

[14] True, the effects of the weather on the statue – erosion or encrustation – change its, and thus Lumpl's, size, as well as shape, so that if we apply Gibbard's mereological essentialist definition of a piece (no loss or addition of part) strictly, the piece of clay, Lumpl, is as ephemeral as is Goliath. But it is obvious that pieces of clay can tolerate small changes, as in fact Gibbard himself allows. (For a different way of shunning coincident entities, see Biro 2016 and Biro 2017b.) And, of course, we do not think about statues in that way, so that Holiath etc. need not be distinct from Goliath, not even, perhaps, when what we have no longer resembles the subject it represents. (Think of the many examples of abstract, non-figurative painting or sculpture intended and seen to represent without resembling.)

Denying that composite material objects exist is typically motivated by a desire for a sparse ontology and a reluctance to multiply entities beyond necessity.[15] But such Ockhamite goals do not require the eschewing of composites altogether. All one needs to deny is that they exist in addition to whatever simples (ultimate or honorary) compose them. When some planks are put together in a certain way, they make a ship. There is, indeed, nothing where the ship is other than those planks so put together. But to insist on this one need not deny that when they *are* put together in that way, there is a ship, even while agreeing that it is nothing over and above the planks.[16] When they are put together in some other way, there may be something else – a house, perhaps, or a pile of planks, which, too, are single things. If they are scattered, there is no single material thing (though each plank is one). We can hold that ships are not something in addition to their planks without agreeing with van Inwagen that there are no ships. More generally, we can hold that there are composites, even if these are not things in addition to the things that compose them.

Acknowledgments

My thanks to referees of this journal for their helpful suggestions.

[15] For an example, see Merricks (2001).

[16] This is to endorse the strong version of the so-called composition-as-identity thesis, on which composition *is* identity (Lewis 1991 and Cotnoir 2013).

REFERENCES

Baxter, D. and Cotnoir, A., eds. 2014. *Composition as Identity*. Oxford University Press.

Biro, J. 2016. Co-location and separability. *Philosophical Inquiries* VI/2: 9-16.

_____ 2017a. Are there scattered objects? *Metaphysica* 18/2: 155-166.

_____ 2017b. Constitution and identity. *Erkenntnis.*

Cartwright, R. 1975. Scattered Objects. In *Analysis and Metaphysics*, K. Lehrer, ed. (Dordrecht: Reidel): 153-170.

Cotnoir, A. J. 2013. Composition as General Identity. *Oxford Studies in Metaphysics*, 8, 294–322.

Elder, C. 2007. On the Phenomenon of "Dog-wise arrangement". *Philosophy and Phenomenological Research* 74 (1): 132–155.

Gibbard, A. 1975. Contingent Identity. *Journal of Philosophical Studies* 4: 187-221.

Hirsch, E. 1993. Peter van Inwagen's *Material Beings. Philosophy and Phenomenological Research* 53 (3): 687-691.

Korman, D. 2009. Eliminativism and the Challenge from Folk Belief. *Nous* 43:2: 242-264.

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

Merricks, T. 2001. *Objects and Persons,* Oxford University Press.

Rosenberg, J. 1993. Comments on Peter van Inwagen's *Material Beings. Philosophy and Phenomenological Research* 53 (3): 701-708.

Smart, B. 1973. The Ship of Theseus, the Parthenon, and Disassembled Objects. *Analysis* 34/1: 24-27.

Uzquiano, G. 2004. Plurals and Simples. *The Monist* 87/3: 429-451.

van Inwagen, P. 1990. *Material Beings*. Ithaca, New York: Cornell University Press.

_____ 1993. Precis of *Material Beings. Philosophy and Phenomenological Research* 53 (3): 683-686.

Yi, B. 2014. Is there a plural object? In *Composition as Identity*, eds. D. Baxter and A. Cotnoir, 169-191. Oxford: Oxford University Press.

EuJAP | Vol. 13, No. 2, 2017
UDK
165:141:113
165.6

EVALUATION OF RESEARCH(ERS) AND ITS THREAT TO EPISTEMIC PLURALISMS*

MARCO VIOLA
Moscow State Pedagogical University, Russian Institute for Advances Studies

ABSTRACT

While some form of evaluation has always been employed in science (e.g. peer review, hiring), formal systems of evaluation of research and researchers have recently come to play a more prominent role in many countries because of the adoption of new models of governance. According to such models, the quality of the output of both researchers and their institutions is measured, and issues such as eligibility for tenure or the allocation of public funding to research institutions crucially depends on the outcomes of such measures. However, concerns have been raised over the risk that such evaluation may be threatening epistemic pluralism by penalizing the existent heterodox schools of thought and discouraging the pursuit of new ones. It has been proposed that this may happen because of epistemic bias favouring mainstream research programmes. In this paper, I claim that (1) epistemic pluralism is desirable and should be preserved; (2) formal evaluation exercises may threaten epistemic pluralism because they may be affected by some form of epistemic bias; therefore, (3) to preserve epistemic pluralism, we need some strategy to actively dampen epistemic bias.

Keywords: *Economic Epistemology, Epistemic Pluralism, Research Policy, Research Evaluation*

1. A new governance for research

At the end of the last Century, many national research and higher education systems underwent major reforms toward a new style of governance, named *steering at a distance*. According to his inventor, it involves

*Received: 04.10.2017.
Accepted: 01.03.2018.

> a departure from classical government steering by means of legislation, prohibitions and regulations and a move towards the autonomy and the delegation of responsibilities to institutions of higher education [...] Under the new system, intervention is restricted to *ex-post facto* adjustments based on quality assessments of results (Kickert 1995, 135).

This style is inspired by the ideals of *New Public Management*, that prescribes to reproduce quasi-market condition to prompt market-like competition in public institutions. As a consequence of this shift, and in the name of accountability and efficiency of public expenditure, many countries implemented large scale and centralized *ex-post* research evaluation systems (Whitley and Gläser 2007). Rather than merely portraying the state of research in a given country, these evaluation systems significantly affect it, especially (but not exclusively) in the many countries where public funding is based upon their results (Hicks 2012). Furthermore, some countries implemented centralized and formalized procedures to regulate hiring. In the remainder of this article, by speaking about *formal evaluation* I will indicate either these kinds of large scale exercises.

2. The unwelcoming reception of formal evaluation in Italy

In Italy, formal evaluation has been introduced later than in most Western countries. Notably, a National Agency for the Evaluation of Research and University System (ANVUR) was founded in 2010, which designed both a centralized system to regulate academic recruitment, i.e. the *Abilitazione Scientifica Nazionale* (ASN),[1] and a nation-wide research assessment, named *Valutazione della Qualità della Ricerca* (VQR).[2] The ASN was then taken in charge by the Ministry of Education and University, while ANVUR has steered two editions of the VQR, assessing research produced through 2004-2010 and through 2011-2014, respectively. Motivated by the explicit goal of compensating for late implementation, governments are making abundant use of the results of such procedures. In particular, the results of VQR have been employed to regulate a large share of public funding for universities, as well as for other issues that were beyond its original scopes (e.g. to set the standards for allowing professors to teach at some PhD-level courses).

The introduction of those formal evaluation systems by ANVUR raised several harsh criticisms. Most of them focused either on the ideological roots of evaluation (most notably Pinto 2012), or on technical mistakes

[1] See http://abilitazione.miur.it/public/index.php?lang=eng, accessed 12/12/2017.

[2] See http://www.anvur.org/index.php?option=com_content&view=article&id=1206:vqr&catid=2:non-categorizzato&lang=it&Itemid=789, accessed 12/12/2017.

made by ANVUR (e.g. Banfi and De Nicolao 2013; Baccini 2016). However, in this paper we want to highlight another problematic feature of formal evaluation systems such as the Italian one. Namely, the risk that they could impoverish the epistemic pluralism of scientific communities. Some reflections were devoted to this issue by one of the very people who designed these formal evaluation, a former member of the ANVUR Steering Committee, in a book conceived as a rebuttal of several criticisms of formal evaluation (Bonaccorsi 2015; see also Bonaccorsi 2018).

Recognizing that some scientific communities, particularly in social sciences and humanities, may not share uniform methodological standards, Bonaccorsi acknowledges the possibility that evaluation may generate epistemic conflicts between different schools of thought that co-exist within a discipline. Advised by the philosopher Carla Bagnoli, he acknowledged the possibility that social mechanisms for value attribution may produce what Miranda Fricker (2007) called *epistemic injustice* (Bonaccorsi 2015, 76, fn. 11).

Notably, in order to introduce the concept of testimonial injustice,[3] Miranda Fricker presented the following example, inspired by a discussion with a scientist friend:

> Imagine […] a panel of referees on a science journal who have a dogmatic prejudice against a certain research method. It might reasonably be complained by a would-be contributor that authors who present hypotheses on the basis of the disfavoured method receive a prejudicially reduced level of credibility from the panel. Thus, the prejudice is such as to generate a genuine testimonial injustice. (Fricker 2007, 27)

However, while Bonaccorsi (2018) explicitly acknowledges and even endorses epistemic pluralism (at least in social sciences and humanities), he is optimistic that such injustices can be avoided by finding a common ground for assessing the research quality across different schools of thought. Contrary to his optimism, in this article I claim that epistemic pluralism is likely to be compromised by a bias that is rooted in any kind of evaluation – but gets amplified when evaluation procedures are highly formalized.

The discussion will proceed as follows: first, I will motivate the claim that epistemic pluralism is a desirable feature for the social organization of science. Then, I will describe a possible kind of epistemic bias that may negatively affect epistemic pluralism and show that it is very likely

[3] Fricker describes two varieties of epistemic injustices: testimonial injustice, occurring when a speaker is given less epistemic authority than she deserves; and hermeneutical injustice, occurring when a social group lacks the conceptual resources to make sense and to express its social experience.

to be at work in all sorts of evaluative practices. Therefore, since even informal evaluation might endanger epistemic pluralism, *a fortiori* particular care should be made with formal evaluation. I conclude by briefly hinting at some strategy that might be adopted to counter this risk.

3. Epistemic Pluralism

3.1. What is Epistemic Pluralism

In science, the word 'pluralism' may refer to many different, if partially overlapping, concepts. For instance, we can have the following three kinds of pluralist stance:

a) *Ontological pluralism.* Contrary to the neo-positivist ideal of unifying science by reducing special sciences to physics (Oppenheim and Putnam 1958), post-positivist philosophy of science argued in favour of a plurality of unreducible ontologies (e.g. Fodor 1974; Suppes 1978; Dupré 1993).

b) *Sociological pluralism.* Feminist philosophy of science has denounced the underrepresentation of some social groups in science (e.g. women and ethnic minorities), and argued in favour of a more balanced composition, for both epistemological and political reasons (Anderson 2015).

c) *Disciplinary pluralism.* More recently, some researchers compared the scientific productivity of various countries, revealing that it was higher in those that diversify their research efforts across more domains as opposed to specialising in some specific one; therefore, they suggest that national science policy-making should try to promote a pluralism of domain of inquiries (Cimini, Gabrielli, and Sylos Labini 2014).

Notwithstanding their relevance, the abovementioned kinds of pluralistic stances are not addressed in the present discussion. Rather, I focus on a fourth variety of pluralism, which I dub *epistemic pluralism*. My working definition is the following:

> epistemic pluralism = the compresence of two or more rival schools of thought within a same domain of inquiry.

Given the lack of undisputable criteria for setting the boundaries between 'rival schools of thoughts', I shall settle for the following stipulation:

> rival schools of thought = distinct research groups who endorse conflicting conceptual and/or methodological commitments, but whose explanatory scopes are at least partially overlapping – i.e., they are competing to explain some shared set of phenomena.

Usually – though not necessarily – such rivalries are revealed by different institutional features such as distinct scientific societies or scientific journals for each competing schools of thought, or by pragmatic features such as different technical languages (or if you prefer, ontologies) and heuristics. Their peculiar disagreement is not much about what they hold to be true about the world – members of a same school of thought may also disagree on that – but rather about *how to verify* these truths, i.e. by means of which methods, heuristics, models, idealizations. In a nutshell, what is at stake here is not the disagreement between specific theories *per se*, but rather between second-order conceptual frameworks – be them construed as thought collectives (Fleck 1935), paradigms (Kuhn 1962/1970), research programs (Lakatos 1970), research traditions (Laudan 1977) or something else. To name but a few intuitive examples of actual rivalry, think about psychodynamics and cognitive psychology; continental and analytic philosophy; neoclassical and heterodox economics.

Is epistemic pluralism desirable for science? If so, in which form? Divergent opinions existed in classical 20th century epistemology. Notably, describing the convergence on a single paradigm as a prerequisite for normal science, Thomas Kuhn (1962/1970) interprets the co-existence of rival schools of thought within a same discipline as a cue of immature science. However, while his justification for such convenience may be regarded as a transcendental argument for endorsing epistemic monism on a synchronic plane, his praise of 'progress through revolution' qualifies him as a supporter of diachronic pluralism (Viola 2015). Other philosophers held that epistemic pluralism should be pursued also on a synchronic plane. Notoriously, Paul Feyerabend (1975) argued for a very radical form of pluralism, expressed in the slogan 'anything goes' (see also Kellert, Longino and Waters 2006). However, one needs not commit to such radical stances to defend epistemic pluralism. More modestly, siding with Lakatos, one can recognize that

> [t]he history of science has been and should be a history of competing research programmes (or, if you wish, 'paradigms'), but it has not been and must not become a succession of periods of normal science: the sooner the competition starts, the better for progress. 'Theoretical pluralism' is better than 'theoretical monism': on this point Popper and Feyerabend are right and Kuhn is wrong (Lakatos 1970, 60).

Epistemic pluralism does not entail antirealism, nor ontological irreducible pluralism such as Duprè's (1993): while being a realist, one could still maintain some form of convergent realism (Kellert, Longino and Waters 2006), holding that while in the long run the one true ontology will be discovered, no option should be foreclosed in advance.[4]

[4] I thank Giuliano Torrengo for pointing this out to me.

In the following sub-sections, I summarise some discussions concerning the desirability of pluralism in social epistemology. Analytical and agent-based models both seem to speak in favour of epistemic pluralism. Nonetheless, since the interpretation of these models is by no means straightforward, I will turn to another source for a modest defence of pluralism – namely, theoretical arguments based on historical considerations.

3.2. Why Epistemic Pluralism matters #1: economic epistemology

Starting from the nineties, social epistemologists began to reflect both on the desirability of epistemic pluralism and on the conditions that may promote or reduce it. An exemplar case of Goldman's (2011) system-oriented social epistemology, this literature addresses both normative and descriptive concerns. Moreover, it borrows some methods from social science – in particular, mathematical and agent-based modelling from economics: that is why it gained the label of 'economic epistemology' or 'economics of scientific knowledge' (Zamorra-Bonilla 2012).

Concerning the present discussion, the first milestone was posed by Kitcher in his 1990 article *The Division of Cognitive Labor*. Adopting a framework originally elaborated by Peirce (1879), Kitcher envisioned a scenario in which a community of scientists, who ought to make a given discovery, must choose if and how to split their cognitive efforts between two methods, i.e. they must answer the question: how many scientists should pursue each method?[5]

It is assumed that the probability that each method produces a discovery in a given timeframe depends on a return function that is increasing in the number of scientists. However, these return functions are concave: consequently, while a method M_1 may be intrinsically superior to another method M_2[6], overcrowding of the former can make it less efficient than hedging the bets. Therefore, hedging the bets by also having a minority of scientists who pursue M_2 is wiser than having everybody pursuing M_1.

Kitcher then goes on discussing whether the social reward structure of science may be particularly fit for achieving this optimal allocation. Strevens (2003) further expands that point by comparing alternative reward structures, and claiming that the one that is more likely to sustain

[5] Kitcher's discussion is based built on the example of the discovery of the molecular structure of some molecule, which can be investigated either by empirical observation or by building toy-tinkers. While this example does not count as a genuine case of epistemic pluralism according to my definition, he specifies that his discussion is meant to refer to various kind of 'cognitive objects' such as "set of rival theories, research programs, methods for approaching a problem, etc." (Kitcher 1990, 10).

[6] M1 is superior to M2 if, given any number of scientists pursuing only one method, the probability of the discovery is always superior when this method is M1.

the optimal allocation of scientists is indeed the actual system, based upon the *priority rule* (first described by Merton 1957), which prescribes that only the first one(s) who make a discovery get a prize for it ('winner takes all'). According to Strevens, this is the most rational allocation, because nothing would be gained by making the same discovery twice (but see §2.4). Nonetheless, Strevens in a later paper (2013) acknowledges that this optimistic assessment of the division of cognitive labour 'naturally' emerging from the adoption of the priority rule may be seriously endangered by the presence of herding behaviour. He notices that the 'golden share' for undertaking the correct scientific project often takes a (indeterminately) long time to unfold. Therefore, risk aversion might drive scientists to settle for more modest sources of credit, such as the recognition of their peers – typically expressed in the form of citation. But then, being into a more crowded school of thought make it easier to be recognized by a wider number of peers – thus making mainstream schools of thought more appealing than it is would be rationally desirable.

Muldoon and Weisberg (2011) refer to these mathematical models as a Marginal Contribution/Reward approach, and criticize them for relying on controversial assumption (mostly inherited from classical economics: see Hands 1997; Mäki 2005; Viola 2015; Fèrnandez-Pinto 2016).

3.3. Why Epistemic Pluralism matters #2: agent-based models

Instead of these models, Weisberg and Muldoon propose to investigate the division of cognitive labour through an agent-based model where the agents (scientists) must explore a three-dimensional 'epistemic landscape', representing the many possible approaches within a scientific field (Weisberg and Muldoon 2009). The 'landscape' is composed by many patches, each one representing a different approach within a given domain of inquiry. Some of these patches are higher than others (representing more fertile approaches), delineating some 'hills' of scientific fruitfulness. Agents ought to explore as many patches as they can among those whose epistemic significance above 0; to put it bluntly, they must climb the hills and its epistemically significant surroundings as soon as possible.

Each agent has only limited information: it only knows the epistemic significance of the patch it occupies, as well as that of those adjacent patterns that have been already explored by some other agent (i.e., scientists left traces of the in form of publications about the patterns they explored). However, Weisberg and Muldoon designed two kinds of agent[7], distinguished by different behavioural patterns: *followers,* who

[7] Plus, a third kind used as control, that I ignore here.

follow the trails of other agents; and *mavericks*, who privilege the exploration of yet unknown patterns over the known ones.

Populations made entirely by followers perform worse than those made entirely by mavericks, because followers tend to cluster and get stuck in low significance regions instead of making brave explorations as mavericks do.[8] However, mixed populations perform even better.

Being aware of the high level of abstraction of their model, Weisberg and Muldoon refrain from drawing strong lessons out of it. Nonetheless, further considering that being a maverick is costlier than being a follower, they propose the tentative conclusion that "optimum research communities are going to be composed of a healthy number of followers with a small number of mavericks" (251). For the sake of the current debate, their conclusion can be interpreted as an indirect endorsement of epistemic pluralism (represented by the exploration of several patches), paired with the suggestion that pluralism is easier to achieve when scientists are biased toward the exploration of unknown approaches.

Other intriguing agent-based simulations with different architectures have been proposed, that are either moderately in favour or against synchronic epistemic pluralism. In Balietti, Mäs and Helbing's (2015) model, scientists ought to find a scientific truth, represented as a point within a bi-dimensional space. Scientists are 'dragged' along the space by the combined effect of three vectoral forces: first, they are attracted by the intrinsic force of the truth-point; second, they are influenced by their neighbour colleagues directions, which they mimic, provided that these colleagues stand within a given agent's 'sensory range'; third, for each agent there is some noise, i.e. some random force. According to this model, a marked epistemic pluralism – represented by scientists being sparsely distributed all over the landscape – hampers progress toward the truth because of the lack of forces that are strong enough to prevent self-reinforcing herding behaviour. In fact, in this scenario noise might deviate small clusters of researchers toward the wrong direction, arguably representing the self-reinforcement of prejudices held by a sub-community due to mimicking behaviours.

Instead, Zollman (2010) argues in favour of transient diversity. He models a scientific community as a network of interconnected Bayesians who play two-armed bandits. Each arm represents a different scientific approach and is characterized by a different payoff distribution. The payoff structure, however, is unknown to scientists. Rather, they have prior beliefs about which one is the better arm and update them by considering both the result of their own choices and those of the scientists

[8] In a follow-up, Weisberg (2013) refers to such a phenomenon as of *herding behavior*. He shows that while some strategies may reduce the herding behavior of followers, none of them would make them as efficient as mavericks.

they are linked with in the network. After testing several kinds of networks, Zollman concludes that while networks with *less* connections are slower, they are more reliable in making everyone converge on the (objectively) better arm, whereas highly interconnected networks sometimes converge on a self-reinforcing consensus over the wrong answer. However, a community with stubborn scientists having extreme priors will manage to test alternative hypotheses without discarding them too soon, and eventually it will converge on the right outcome even in highly interconnected networks. Yet, combining extreme priors and low interconnections tend to fossilize the disagreement and paralyze various clusters of scientists into their prejudice, thus failing to achieve consensus.

All things considered, while both Zollman and Weisberg and Muldoon's models suggest that a certain amount of epistemic pluralism might be beneficial (at least in some conditions), *prima facie* Bailetti and colleagues' model seems to point toward the opposite conclusion. This disagreement mainly depends on the different scopes and assumptions made by these models. Given that these model assumptions are presently "still rather disconnected from the real-world social organization of scientific research" (Martini and Fernàndez-Pinto 2016), we would refrain from drawing strong conclusion from them.

3.4. Why Epistemic Pluralism matters #3: historical and sociological considerations

Given the uncertainty surrounding the models found in social epistemology, my endorsement of epistemic pluralism will mainly bear on two more modest epistemological arguments inspired by some simple historical and sociological considerations. I dub them the *prudence argument* and the *convergence argument* and discuss them in turn.

> *[Prudence]* we cannot reliably foresee which one, among many rival schools of thought, is more likely to produce correct or more significant findings.

Often, a school of thought may fail to explain some phenomena which are easily accounted by another one. In contrast to what is assumed in Kitcher's (1990) model, the history of science seems to suggest that we cannot reliably foresee which school of thought is more likely to produce a given answer. This is vividly expressed in the case discussed by Zollman (2010), i.e. the discovery that peptic ulcers are caused by the *helicobacter pylori*. In 19th century, two competing hypotheses were proposed to explain the disease: the presence of some unobservable bacteria and an excess of acid. When in 1954 the prominent gastroenterologist Palmer published a study that appeared to demonstrate

that no bacteria can colonize the human stomach, this was taken as a conclusive evidence against the bacterial hypothesis. Sadly, his conclusion was unwarranted, since the kind of stain he used to investigate the biopsies was 'blind' to the *H. pylori*. It took about thirty years to Marshall and Warren to discover that the disease was caused by a bacterium. Yet, at first their discovery (yielding them the Nobel for Medicine in 2005) was dismissed by the medical community, since Palmer's conclusions had been crystalized into received wisdom among the medical community. Fortunately, the frustrated Marshall behaved as Zollman's stubborn scientist, and he himself drank a solution containing *H. pylori*. He manifested the symptoms of the peptic ulcer, and then effectively cured himself with an antibiotic, thus convincing his peers. Despite this story has a happy ending, Zollman cannot help wondering about how many more patients could have been successfully cured if only the bacterial hypothesis was not dismissed too soon. And we may also ask: how many correct hypotheses could have been overlooked if they have had no stubborn advocates such as Marshall?

Nonetheless, science must not only find truths: indeed, it should find *significant truths* (Kitcher 1993, 94). However, the significance of some scientific discoveries (which I take to indicate their potential to contribute to social well-being) cannot be unequivocally estimated *ex ante*, also because each piece of the puzzle of science might gain value depending on the availability of other pieces. This dynamic character of epistemic significance is nicely explored by Avin (2015a, ch. 4).[9] For instance, Avin stresses how the laser gyroscope, which required advancements in engineering and in theoretical physics made in the Sixties in order to be built, was only conceivable because of an experiment made in 1913 by Georges Sagnac, and published in France, whose original scope was to test *ether wind*.

Furthermore, many important discoveries in science were not due to some specific theory-testing. Rather, many ground-breaking discoveries emerged as the unexpected result of some fortuitous event – a circumstance for which the word *serendipity* has been coined. An evocative example is the discovery of penicillin: Alexander Fleming noticed that the cover of a Petri dish containing bacterial culture had not been properly set, and that a mould had grown, killing the bacteria.

Arfini, Bertolotti and Magnani stress that in order to make a serendipitous discovery, it needs to be "not expected, but […] still recognizable, at least to certain cognitive systems. Otherwise, it would be pushed aside by consciousness". Since the school of thoughts scaffold scientists' cognitive system, a monopolistic school of thought with an overly restrictive ontology might work as a blindfold for some phenomena that are not predicted by its ontology. Indeed, according to Kuhn (1962), this is the

[9] This facet of Avin's work was brought to my attention by Carlo Debernardi.

routinely way to deal with anomalies. Kyle Stanford (2015) makes a similar point, expressing the worry that the actual structure of science, due to the concentration of incentives toward conservative research, might make it harder to grasp unconceived alternatives.

These considerations stress the risks of allowing for the monopoly of a single school of thought, thus vindicating a cautionary rationale for preserving at least a *minimal epistemic pluralism*. Lastly, it is worth keeping in mind that scientific activity is imbued with *tacit knowledge* (Polanyi 1966). This kind of knowledge is hardly translatable onto explicit knowledge; rather, it is usually transmitted through long apprenticeship – which is why reading textbooks is not enough to become a scientist, but a doctorate or some other equivalent form of apprenticeship is in order. Thus, allowing a school of thought to completely extinguish likely implies a loss of tacit knowledge – perhaps right before the availability of some piece of the puzzle would make it priceless: a despisable loss, compared to the relative small price of letting some minoritary school of thought continuing its legacy, if just for a few stubborn followers.

> *[Convergence]* if a multiplicity of independent rival schools of thought converges on a same result, this result is more reliable.

Strevens's (2003) abovementioned defence of the priority rule was based upon the assumption that we do not need the same discover to be made twice. However, since the reliability of science significantly bears on the reproducibility of its findings,[10] giving no incentives at all for replications is tantamount to deprive science of its antibodies, because scientific frauds and mistakes will lurk for longer, and perhaps forever – a topic which is daunting for current research (e.g. Ioannidis 2005).

In particular, the better guarantee for the reliability of (a piece of) scientific knowledge comes from the convergence of many independent sources (Kosso 1989). Jean Perrin's discovery of the Avogadro's number counts as an exemplar case of a reliable knowledge, since he "measured the same physical quantity in a variety of different ways, thereby invoking a variety of different auxiliary theories" (Kosso 1989, 247). Given that rival schools of thought employ, by definition (see above), different methods for testing scientific statements, it follows that whatever scientific theory is deemed true by distinct, even rival schools of thought, is *ceteris paribus* more robustly validated than one that is backed solely by the followers of a single school of thought.

---

[10] Eugenio Petrovich brought to my attention that now that we are in the era of Big Science, where some experiments are simply too expensive to be replicated (think about the CERN in Genève), replication might belong to the idealized image of science, rather than to its accurate description (Collins 1992). I concede that this might be true when it comes to big science. Yet, I contend that replication remains possible in many domains outside Big Science – and that failures to replicate present serious reasons of concern.

4. Epistemic Bias

4.1. Why should evaluations be threatening to Epistemic Pluralism?

Evaluation is usually aimed at measuring the scientific quality of some research products and/or of those who produced it. However, provided that scientific quality exists, what makes us think that evaluators would be able to grasp it objectively, without being affected by their prejudices? Even excluding the case of deliberate boycott of rival schools of thought by scholars engaged in evaluation procedures, it is well established that human beings are prone to many implicit biases. Why then should we suppose that scientists are immune to biases when assessing their peers and their work? And what happens if such biases are embedded into large-scale formal evaluation with profound implications for a national scientific environment?[11]

Some reflections on biases that may harm epistemic pluralism can be found in Donald Gillies's 2008 book. The book is a critical assessment of the Research Assessment Exercise (RAE, recently replaced by Research Excellence Framework, REF), whose results have been used to allocate public funding for research in the UK. Since RAE was based on peer review, Gillies asks: is peer review able to predict which research projects are going to bear fruitful results? Were his answer positive, it could be reasonable to concentrate many resources to that project, even at the expense of other strands. But that seems not to be the case. As he summarized in a later article,

> the root of the problem is what I will call *researcher narcissism*. This is a condition, which affects nearly all researchers (including the author of the present paper). It consists in an individual researcher believing quite strongly that his or her approach to research in the field is the best one, and most likely to produce good results, while the other approaches are less good and less likely to produce any good results (Gillies 2014, 8).

Gillies adopts a counterfactual strategy to substantiate his scepticism: he discusses various historical cases from many fields where peer review would have failed to foresee ground-breaking scientific advances: Frege's invention of mathematical logic, Copernicus's theorisation of heliocentrism, Semmelweis's invention of antisepsis (Gillies 2008, ch. 3), James Black's Nobel-awarding inventions of two important drugs (Gillies 2014).

[11] Feminist philosopher Jennifer Saul (2012) addressed the issue of implicit biases in two evaluative exercise undergone in the UK: the Philosophical Gourmet Report and the Research Excellence Framework. However, since she is only concerned with what I called sociological pluralism (§1), I do not address her work here.

4.2. Is there evidence for Epistemic Bias?

Despite its intuitive appeal, Gillies's discussion on researcher narcissism is purely speculative. Is there any evidence that such a kind of *epistemic bias* is at play in evaluation? Is it stronger in formal evaluation? To address these question, I browsed the literature in several social sciences that deal with the presence of biases across three different kinds of evaluative practice: peer review, bibliometric evaluations and hiring procedures.

Peer review

Peer review has been compared to democracy; both have been described as "a system full of problems but the least worst we have" (Smith 2006). Among these problems, many researchers highlighted many biases that compromise the alleged impartiality of the process (see Langfeldt 2006; Lee, Sugimoto, Zhang, and Cronin 2013). Nonetheless, few scholars specifically addressed the issue of epistemic bias as distinct from other biases, also because of the difficulty to disentangle them.[12]

If such epistemic bias applies to peer reviewing of scientific articles, referees might simply reject papers from rival schools of thought they disagree with, or even steer the author toward their own theoretical perspective. Might this be the case?

Some evidence in the literature suggests that the answer might be 'yes'. Mahoney (1977) asked 75 (unaware) referees in experimental psychology to review a given manuscript, whose results he slighlty modified, along with their interpretation. He found that referees tend to judge more positively the manuscripts that show positive results and/or that are in line with the theoretical perspectives of the referees.

However, whereas epistemic bias can exert a significant effect on scientific careers by influencing the fate of articles submitted in prestigious journals, its role is even more direct when it comes to allocate research fund. It is difficult to disagree that funding agencies may affect "the cognitive development of science by the structuring of the way in which research is done" (Braun 1998, 810; see also Goldman 1999, 257). Sadly (for epistemic pluralism), in doing that, they foster conservative researches over innovative ones (Braun 1998; see also Berezn 1998).

Having attended some meetings of panels that ought to adjudicate grants on behalf of the *Science and Engineering Research Council* (SERC), Travis and Collins (1991) observed that "committee members sometimes

[12] Wang & Sandström (2015) tried to overcome this problem by developing a data mining technique to measure the 'cognitive distance' between researchers, comparing how much both their references and some key term from their abstract overlap. However, their measurement conflates the proximity of school of thought with that of topic.

make decisions based upon their membership in scientific schools of thought" (323).

More recently, Luukkonen (2012) wondered whether ERC panels were able to ensure that funds are channelled into "new and promising areas of research with more flexibility" (http://erc.europa.eu/mission). Her answer was negative: she declared that "despite the ERC's aims, the peer review process in some ways constrains the promotion of truly innovative research" (Luukkonen 2012, 11), and she further observes that "[t]hese constraints arise from the very essence of peer review, namely, its basic function of judging the value of proposed research against current knowledge boundaries" (*ibid.*).

Bibliometrics

Prima facie, due to their mathematical format, bibliometric indicators might seem good candidates for providing *objective* measures of research quality. They also might be tempting due to their relative inexpensiveness, especially in large scale formal evaluation exercises where the number of research products to evaluate is high. However, it is worth remembering that since most widespread bibliometric indicators (e.g. *impact factor* and *h-index*, respectively measuring the impact of journals and of researchers) are based upon counting citations within peer reviewed articles indexed by a given database, they embed and aggregate the prejudices of both the referees and the editors of the journals, plus the indexing criteria of the database owners. Moreover, bibliometric indicators are meant to represent *impact,* not *quality*. Whereas sometimes impact is considered as a reliable proxy of quality (e.g. in the Italian VQR it is done for many scientific fields, especially in hard and life sciences), this identification is problematic, as it provides strong disincentives to work on mainstream problems and within heterodox schools of thought (as documented for instance by Castellani, Pontecorvo, and Valente 2016). To understand why this happens, consider the following scenario: two papers of comparable quality, P1 and P2, provide a relevant insight over a same issue. However, P1 does so from the standpoint[13] of a mainstream school of thought, with huge number of followers, whereas P2 from that of a minor (or yet to exist) school of thought, with far less followers. Given the reasonable assumption that the segregation between school of thoughts make it relatively less likely that some scholar reads (and thus cite) a paper from a rival school of thought, all else being equal, the wider audience would boost P1 impact far over P2's – irrespectively of their quality. In Muldoon and Weisberg's (2009)

[13] Here the notion of 'standpoint' should be interpreted according to either a cognitive dimension (e.g. the lexicon used to frame a problem, the method used to test it) or a sociological one (e.g. the paper being discussed in some conferences and published in journals that are commonly associated with one school of thought).

terms, the 'citation economy' discourages people from behaving like mavericks, because mavericks are arguably less likely to get cited.

For these and other reasons, many institutions and scientists subscribed the San Francisco Declaration on Research Assessment, which prescribe "not [to] use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions" (http://www.ascb.org/dora/).

Hiring

Among the various evaluative procedures, comparative evaluations of candidates for academic recruitment, as well as procedures assessing their eligibility (e.g. the abovementioned ASN in Italy), are possibly the more relevant as for epistemic bias due to a path-dependent reinforcing loops. In fact, it is very likely that the promoted candidate will oversee judging some future candidates, and any epistemic bias will be transmitted to the next generation of evaluators – and thus amplified.

Available literature shows that some biases are indeed at play during hiring procedures: for instance, candidates who are someway connected with the examiners (e.g. they are co-authors of some articles, or come from a same department), are more likely to be hired (e.g. for France see Combes, Linnemer, and Visser 2008 and Godechot 2016; for Spain, see Zynovieva and Bagues 2015). However, these authors stress that they are unable to judge whether this advantage was due to epistemic bias or rather to social particularism (such as nepotism).[14]

Nonetheless, even social biases may have important epistemic consequences. Studying the hiring networks of American research institutions in computer sciences, business, and history, Clauset, Arbesman, and Larremore (2015) highlighted a very 'endogamous' and hierarchical structure: on the one hand, most professors (about four out of five) obtained their PhD in one of the 'top' 25% departments – among which further hierarchical layers could be distinguished. On the other hand, almost none of those who obtained their PhD in less prestigious institutions managed to be hired at the higher-levels. Therefore, they conclude that

> the centralized and highly connected positions of higher-prestige institutions enable substantial influence, via doctoral placement, over the research agendas, research communities, and departmental norms throughout a discipline . . . . The close proximity of the core to the entire network implies that ideas originating in the high-prestige core, regardless of their merit, spread more easily through-

[14] I speculate that in future works this confound may be at least partially addressed by applying formalized measures of cognitive distance such as that developed by Wang and Sandström (2015; see fn. 12).

> -out the discipline, whereas ideas originating from low-prestige institutions must filter through many more intermediaries (Clauset et al. 2015, 4).

To sum up, there is moderate evidence that epistemic bias is at play within each of the three evaluative activities I have discussed, i.e. peer review, bibliometrics, and hiring procedures. A sceptic may still argue that this evidence is far from being conclusive. I concede that. However, since I think that most researchers would take for granted that epistemic bias is at play and significantly distorts evaluations, I claim that the burden of disproof is up to the sceptics. Moreover, even if the pluralism-reducing effect of epistemic bias were moderate in each of the above-mentioned fields, the cumulative impact may be significant: even though being hired were only *slightly* more difficult if the members of the panel are hostile to your school of thought, it does become *considerably* harder if due to the unpopularity of your school of thought you had an hard time publishing your papers in high-ranked journals and to get them cited. Such a self-reinforcing loop has been reported to affect economics schools of thought in the UK: according to Lee, Pham, and Gu (2013), twenty years of RAE resulted in the disappearance of heterodox rival schools of thought in favour of mainstream economics.

All things considered, unless and until sceptics succeed in demonstrating that epistemic bias is negligible or harmless, there are strong reasons for worrying about epistemic bias and for trying to mitigate it.

5. Formal evaluation enhances epistemic bias

In §3 I have defined epistemic pluralism and argued in favour of at least a minimal form of pluralism. Then, in §4 I have introduced and substantiated the hypothesis that any kind of evaluative process is prone to be affected by epistemic bias, i.e. the evaluators might favour those who pursue their same school of thought over those who do not.

The simplest and most radical solution would be to cease any evaluation. Yet, this is hardly a viable option: as far as some finite amount of public resources must be allocated, we need some criteria for choosing how to allocate them. However, while *some* form of evaluation is mandatory, these forms need not be strong evaluation systems that (a) are steered by some rather restricted scientific elite, (b) follow highly formalized rules and procedures, and (c) have a straightforward impact on affect funding and careers. These are the characteristics of the systems described by Hicks (2012), and especially of the Italian systems described in §2. Due to their often wide-scope, they make a large use of bibliometry (e.g., in Italy bibliometric index have been employed for hard and life sciences) or to other highly standardised index and rankings (e.g., in Italy journals in social and human sciences have been classified in hierarchical rankings

for the ASN). According to Whitley (2007, 10), such systems tend to impose a standardization and a institutionalization of goals and values to a scientific discipline, so that "the diversity of intellectual goals and approaches within sciences should decline over time, especially where they challenge current orthodoxies". Eventually, "such reinforcement of disciplinary standards and objectives is likely to inhibit the development of new fields and goals that transcend current intellectual and organisational boundaries by increasing the risks of investing in research projects that do not fit within them" (*ibid.*).

My hypothesis is that this happens because formal evaluation amplifies the epistemic bias already existing in weaker evaluation practices, as well as accelerating their pluralism-dampening effects. This is consistent with the claim of Bonaccorsi (who recently governed the implementation of such procedures in Italy) that formal evaluation systems "ha[ve] the effect to foster and catalyze the epistemic reflection of the community" (2015, 88, translation is mine). However, Bonaccorsi does not side against epistemic pluralism, that he recognizes as a genuine (and perhaps even beneficial) feature of social sciences and humanities. He is optimistic that epistemic pluralism might be preserved notwithstanding epistemic bias, because he deems possible that schools of thought find some common ground for bias-free evaluation. Though, on the light of the evidence of epistemic bias discussed in §4, it seems much more likely that a dominant school of thought will impose its evaluative criteria as a common ground, promoting the extinction of scientific minorities (or preventing the birth of new ones). This evidence is not conclusive, but is likely sufficiently strong to put the burden of proof upon the shoulders of those who deem, like Bonaccorsi, that epistemic bias can be made consistent with a common ground for evaluation.

6. Some hypotheses for protecting epistemic pluralism from epistemic bias

Possibly, epistemic bias cannot be completely counteracted. However, some strategies have been proposed in order to reduce it. Bonaccorsi (2015) concedes that if (and only if) a common ground cannot be found (and there are reasons to suspect that this will be the norm, rather than the exception), members of the evaluative panels must be selected with the aim to represent (m)any school(s) of thought. He also stresses the importance of a rapid turnover of these panels.

Drwaing upon research in management studies, Osterloh and Frey (2015) endorse a more radical answer to the question "what kind of control is suited for science?" They think that both output control, i.e. bibliometrics, and process control, i.e. peer review, have too many flaws, and produce too many distortions. The only opinion left is input control:

in their opinion, candidate researchers should undergo a thorough hiring procedure, but then, if they get hired, they should be left free to determine their agenda by themselves (cf. Gillies 2008).

As for research grants, it has been proposed to supersede epistemic bias by picking at least some of them at random, through a lottery. This proposal has been recently detailed by Avin (2015a, 2015b, 2018; see also Gillies 2014), who also explained its rationale. To put it shortly: research project for grant allocation should be kept short, since long projects absorb plenty of time from both those who write and those who read them, and yet they fail to overcome the intrinsic unpredictability of the projects' outcome. All proposals of high merit should be funded, just as all proposals of low merit should be discarded. This, however, leaves out a wide number of proposals of medium merit. Given that noise and unpredictability would render finer-grained assessment useless, these medium-merit proposals should be placed in a lottery, and the winners should be funded. According to Avin, this method might lower the costs (especially of time), increase fairness and even make unorthodox ideas more easily funded.[15]

Other thinkers have recommended to fund people, rather than projects (e.g. Berezin 1998) – a proposal that has been considered by many institution of the NIH in the US (Kaiser 2014). To begin with, it could be wise not to concentrate all the funding into a single agency (Travis and Collins 1991): as reported by Whitley (2007), diversification of funding sources might soften the effects of formal evaluation systems.

Be as it may, the arguably most efficient strategy for slowing down the effects of epistemic bias, thereby preserving epistemic pluralism, is that of inverting the actual trend of concentrating resources in the hands of few researchers at the expenses of the many (Sylos Labini 2016). This might also be achieved by mitigating the use of formal evaluation in allocating funds, or simply by doing without them altogether.[16]

The assessment of the merits and flaws of these and other proposals would require a thoroughly discussion based on empirical analyses that also takes into account contextual information. Obviously, such an endeavour lies beyond the purpose of the present article. Hopefully, such an assessment would take benefit from a careful and well informed public

[15] As reported by Avin (2018), some lotteries have been implemented already for some grants (the Health Research Council of New Zealand's "Explorer Grants", New Zealand's Science for Technology Innovation "Seed Projects" and the Volkswagen Foundation's "Experiment!" grants). However, their implementation is too recent to assess the efficacy of the policy.

[16] This might also lead to considerable savings: Geuna and Piolatto (2016) estimate that, all things considered, UK spent around 88.5 million € for RAE 2008, and are spending from 130 up to 164 million € for REF 2014, while Italy will spend about 70 million € for the second edition of VQR.

debate. Prompting it is the aim of this article.

Acknowledgments

The author thanks the many people who provided insightful comments on earlier versions of this paper: Selene Arfini, Alberto Baccini, Raffaele Caterina, Carlo Debernardi, Giuseppe De Nicolao, Paola Galimberti, Alex Gillett, Eugenio Petrovich, Francesco Sylos Labini, Silvia Tossut, and the anonymous reviewers. The author is also grateful to Matteo Pinna Pintor for the proofreading, as well as for some interesting comments.

REFERENCES

Anderson, E. 2015. Feminist Epistemology and Philosophy of Science. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. http://plato.stanford.edu/entries/feminism-epistemology/.

Arfini, S., Bertolotti, T. and Magnani, L. 2018. *The Antinomies of Serendipity. How to Cognitively Frame Serendipity for Scientific Discoveries,* Topoi, DOI: 10.1007/s11245-018-9571-3

Avin, S. 2015a. *Breaking the Grant Cycle: On the Rational Allocation of Public Resources to Scientific Research Projects* (Doctoral dissertation, University of Cambridge). Accessed December 15, 2017. https://www.repository.cam.ac.uk/bitstream/handle/1810/247434/phd_dissertation_final_for_print.pdf?sequence=1&isAllowed=y.

_____ 2015b. Funding Science by Lottery. In *Recent Developments in the Philosophy of Science: EPSA13 Helsinki*, eds. U. Mäki, I. Votsis, S. Ruphy and G. Schurz, 111-126. Basel: Springer International Publishing.

_____ 2018. Policy Considerations for Random Allocation of Research Funds. RT. *A Journal on Research Policy and Evaluation* 6(1). https://doi.org/10.13130/2282-5398/8626.

Baccini, A. 2016. Napoléon et l'évaluation Bibliométrique de la Recherche: Considérations sur la Réforme de l'Université et sur l'Action de l'Agence Nationale d'évaluation en Italie. *Canadian Journal of Information and Library Science* 40(1): 37-57.

Balietti, S., Mäs, M. and Helbing, D. 2015. On Disciplinary Fragmentation and Scientific progress. *PloS one* 10(3): e0118747. https://doi.org/10.1371/journal.pone.0118747

Banfi, A. and De Nicolao, G. 2013. Valutare Senza Sapere. Come Salvare la Valutazione della Ricerca in Italia da chi Pretende di Usarla Senza Conoscerla. *Aut Aut* 360: 43-68.

Berezin, A. 1998. The Perils of Centralized Research Funding Systems. *Knowledge, Technology & Policy* 11(3): 5-26.

Bonaccorsi, A. 2015. *La Valutazione Possibile. Teoria e Pratica nel Mondo della Ricerca*. Bologna: Il Mulino.

_____ 2018. Towards an Epistemic Approach to Evaluation in SSH. *The Evaluation of Research in Social Sciences and Humanities,* ed. A. Bonaccorsi*,* 1-29. Cham: Springer.

Braun, D. 1998. The Role of Funding Agencies in the Cognitive Development of Science. *Research policy* 27(8): 807-821.

Castellani, T., Pontecorvo, E. and Valente, A. 2016. Epistemic Consequences of Bibliometrics-based Evaluation: Insights from the Scientific Community. *Social Epistemology* 30(4): 1-22.

Cimini, G., Gabrielli, A. and Labini, F. S. 2014. The Scientific Competitiveness of Nations. *PloS one* 9(12), e113470. https://doi.org/10.1371/journal.pone.0113470

Clauset, A., Arbesman, S. and Larremore, D. B. 2015. Systematic Inequality and Hierarchy in Faculty Hiring Networks. *Science Advances* 1(1), e1400005. https://doi.org/10.1126/sciadv.1400005

Collins, H. 1992. *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press.

Combes, P. P., Linnemer, L., and Visser, M. 2008. Publish or peer-rich? The Role of Skills and Networks in Hiring Economics Professors. *Labour Economics* 15(3): 423-441.

Dupré, J. 1993. *The Disorder of Things. Metaphysical Foundations of the Disunity of Science*. Cambridge and London: Harvard Press.

Fernàndez Pinto, M. 2016. Economics Imperialism in Social Epistemology: A Critical Assessment. *Philosophy of the Social Sciences* 46(5): 443-472.

Feyerabend, P. 1975. *Against Method: Outline of an Anarchistic Theory of Knowledge*. Bristol: New Left Books.

Fleck, L. 1979. *Genesis and Development of a Scientific Fact.* Chicago: University of Chicago Press. (Original: *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkolektiv*, 1935)

Fodor, J. A. 1974. Special Sciences (or: The Disunity of Science as a Working Hypothesis). *Synthese* 28(2): 97-115.

Fricker, M. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. New York: Oxford University Press.

Geuna, A. and Piolatto, M. 2016. Research Assessment in the UK and Italy: Costly and Difficult, but Probably Worth It (at Least for a While). *Research Policy* 45(1): 260-271.

Gillies, D. 2008. *How should research be organised*? London: College Publications.

_____ 2014. Selecting Applications for Funding: Why Random Choice is Better than Peer Review. *RT. A Journal on Research Policy and Evaluation* 2(1). https://doi.org/10.13130/2282-5398/3834

Godechot, O. 2016. The Chance of Influence: A Natural Experiment on the Role of Social Capital in Faculty Recruitment. *Social Networks* 46: 60-75.

Goldman, A. I. 2011. A Guide to Social Epistemology. *Social Epistemology: Essential Readings*, eds. A. I. Goldman and D. Whitcomb, 11-37. New York: Oxford University Press.

Hands, D. W. 1997. Caveat Emptor: Economics and Contemporary Philosophy of Science. *Philosophy of Science* 64: S107-S116.

Hicks, D. 2012. Performance-based University Research Funding Systems. *Research Policy* 41(2): 251-261.

Kaiser, J. 2014. NIH Institute Considers Broad Shift to 'People' Awards. *Science 345*(6195): 366-367.

Kellert, S. H., Longino, H. E. and Waters, C. K., eds. 2006. *Scientific Pluralism*. Minneapolis: Univeristy of Minnesota Press.

Kickert, W. 1995. Steering at a Distance: A New Paradigm of Public Governance in Dutch Higher Education. *Governance 8*(1): 135-157.

Kitcher, P. 1990. The Division of Cognitive Labor. *The Journal of Philosophy* 87(1): 5-22.

_____ 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusion*. Oxford: Oxford University Press.

Kosso, P. 1989. Science and Objectivity. *The Journal of Philosophy 86*(5): 245-257.

Kuhn, T. S. 1962/1970. The Structure of Scientific Revolutions. Chicago: University of Chicago Press.

Ioannidis, J. P. 2005. Why Most Published Research Findings are False. *PLos med 2*(8): e124. https://doi.org/10.1371/journal.pmed.0020124

Lakatos, I. 1970. Falsification and the Methodology of Scientific Research Programmes. In *Criticism and the Growth of Knowledge*, eds. I. Lakatos and A. Musgrave, 8-101. Cambridge: Cambridge University Press.

Langfeldt, L. 2006. The Policy Challenges of Peer Review: Managing Bias, Conflict of Interests and Interdisciplinary Assessments. *Research Evaluation 15*(1): 31-41.

Laudan, L. 1977. *Progress and its Problems: Towards a Theory of Scientific Growth*. Berkley and Los Angeles: University of California Press.

Lee, F. S., Pham, X. and Gu, G. 2013. The UK Research Assessment Exercise and the Narrowing of UK Economics. *Cambridge Journal of Economics* 37(4): 693-717.

Lee, C. J., Sugimoto, C. R., Zhang, G., and Cronin, B. 2013. Bias in Peer Review. *Journal of the American Society for Information Science and Technology* 64(1): 2-17.

Luukkonen, T. 2012. Conservatism and Risk-Taking in Peer Review: Emerging ERC Practices. *Research Evaluation* 21(1): 48-60.

Mahoney, M. J. 1977. Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System. *Cognitive Therapy and Research* 1(2): 161-175.

Mäki, U. 2005. Economic Epistemology: Hopes and Horrors. *Episteme* 1(03): 211-222.

Martini, C. and Pinto, M. F. 2017. *Modeling the Social Organization of Science*. European Journal for Philosophy of Science 7(2): 221-238.

Merton, R. K. 1957. Priorities in Scientific Discovery: A Chapter in the Sociology of Science. *American Sociological Review* 22(6): 635-659.

Muldoon, R. and Weisberg, M. 2011. Robustness and Idealization in Models of Cognitive Labor. *Synthese* 183(2): 161-174.

Oppenheim, P. and Putnam, H. 1958. Unity of Science as a Working Hypothesis. In C*oncepts, Theories and the Mind-Body Problem. Minnesota Studies in the Philosophy of Science. Vol. II*., eds. H. Feigl, M. Scriven and G. Maxwell, 3-36. Minneapolis: University of Minnesota Press.

Osterloh, M. and Frey, B. S. 2015. Ranking Games. *Evaluation Review 39*(1): 102-129.

Peirce, Charles S. 1879. Note on the Theory of the Economy of Research. In *Report of the Superintendent of the United States Coast Survey Showing the Progress of the Work for the Fiscal Year Ending with June 1876*, 197-201. Washington DC: US Government Printing Office.

Pinto, V. 2012. *Valutare e Punire. Una Critica alla Cultura della Valutazione*. Napoli: Cronopio.

Polanyi, M. 1966. *The Tacit Dimension*. London: Routledge.

Smith, R. 2006. Peer Review: A Flawed Process at the Heart of Science and Journals. *Journal of the Royal Society of Medicine* 99(4): 178-182.

Stanford, P. K. 2015. Unconceived Alternatives and Conservatism in Science: The Impact of Professionalization, Peer-Review, and Big Science. *Synthese*. https://doi.org/10.1007/s11229-015-0856-4

Strevens, M. 2003. The Role of the Priority rule in Science. *The Journal of Philosophy* 100(2): 55-79.

_____ 2013. Herding and the Quest for Credit. *Journal of Economic Methodology* 20(1): 19-34.

Suppes, P. 1978. The Plurality of Science. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Vol. 1978*, 3-16. Chicago: University of Chicago Press.

Sylos Labini, F. 2016. *Science and the Economic Crisis: Impact on Science, Lessons from Science*. Basel: Springer International Publishing.

Travis, G. D. L. and Collins, H. M. 1991. New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System. *Science, Technology & Human Values 16*(3): 322-341.

Viola, M. 2015. Some Remarks on the Division of Cognitive Labor. *RT. A Journal on Research Policy and Evaluation*, *3*(1). https://doi.org/10.13130/2282-5398/4768

Wang, Q. and Sandström, U. 2015. Defining the Role of Cognitive Distance in the Peer Review Process with an Explorative Study of a Grant Scheme in Infection Biology. *Research Evaluation 24*(3): 271-281.

Weisberg, M. 2013. Modeling Herding Behavior and its Risks. *Journal of Economic Methodology 20*(1): 6-18.

Weisberg, M. and Muldoon, R. 2009. Epistemic Landscapes and the Division of Cognitive Labor*. *Philosophy of Science 76*(2): 225-252.

Whitley, R. 2007. Changing Governance of the Public Sciences: The Consequences of Establishing Research Evaluation Systems for Knowledge Production in Different Countries and Scientific Fields. In *The changing Governance of the Sciences: The Advent of Research Evaluation Systems*, eds. R. Whitley and J. Gläser, 3-30. Dodrecht: Springer Science & Business Media.

Whitley, R. and Gläser, J., eds. 2007. *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*. Dodrecht: Springer Science & Business Media.

Zamora Bonilla, Jesus P. 2012. The Economics of Scientific Knowledge. In *Handbook of the Philosophy of Science. The Philosophy of Economics*, ed. U. Mäki, 823-862. Amsterdam: Elsevier.

Zollman, K. J. 2010. The Epistemic Benefit of Transient Diversity. *Erkenntnis* 72(1): 17-35.

Zinovyeva, N. and Bagues, M. 2015. The Role of Connections in Academic Promotions. *American Economic Journal: Applied Economics* 7(2): 264-292.

EuJAP | Vol. 13, No. 2, 2017
UDK
111.32(082)(049.3)
130.31(082)(049.3)

BOOK REVIEW

A REVIEW OF PERSPECTIVES ON THE SELF, edited by Boran Berčić, 2017.*

RADIM BĚLOHRAD
Masaryk University

The notion of self is one of the most elusive in contemporary philosophy. It is a concept with many layers and connotations and its analysis generates a vast amount of literature and a large number of disputes in philosophical, psychological as well as neuroscientific circles. What exactly is the self? What are its essential characteristics? Is it material? Is it identical to the body or could it move from one body to another? Is there anything like the self in the first place? Can we examine it by introspection? How do we reidentify selves in time? Is the self the bearer of agency and responsibility? Could I become someone else?

Boran Berčić has edited the collection *Perspectives on the Self* with the aim to shed light on some of the facets of the self. The collection, according to the editor, results from the activities of several philosophers at the Department of Philosophy in Rijeka and a conference on the self, which was held in Rijeka the spring of 2016.

There are 17 papers in the collection, 12 are authored by philosophers associated with the Department of Philosophy in Rijeka or other Croatian philosophy departments, and 5 by authors from UK, Serbia, Finland, Hungary and USA.

The collection is divided into 6 sections with an extensive introduction by the editor, who provides a brief overview of the claims and arguments made in each paper. The sections focus on the relationship between the self and the body, self-knowledge, the history of the concept of self, self as an agent, the very existence of self and general metaphysical and linguistic issues involved in the concept. There are two to four papers in each section.

The first section contains papers by Eric Olson, Miljana Milojević and Zdenka Brzović.

In his paper *The Central Dogma of Transhumanism* Olson argues against a claim frequently made by transhumanists, according to which it is possible to scan the synaptic matrix of the human brain and upload the information to a computer, thus guaranteeing the survival of the person.

*Received: 27.11.2017.
Accepted: 28.11.2017.

Olson argues that this supposition leads directly to the *branching problem*, which is well-known from classical objections to the psychological continuity theory of human persistence. Also, he shows that the transhumanist claim makes unintelligible the difference between a person being uploaded to a computer and a new person indistinguishable from the original one being programmed in the computer. Olson then identifies a deeper problem with the transhumanist claim: people are animals and the process of scanning and uploading will simply not move an animal to a computer, that is, guarantee the survival of the animal. Finally, Olson assesses three alternative accounts of human identity – the pattern view, the constitution view and the temporal-parts view. He argues that the first one is inconsistent with the fact that we change, the second one does not actually deal with the above problems, and the last one does not give us what we want from survival.

Milojević holds a very different theory of human identity in her *Embodied and Extended Self*. Assuming functionalism about mental states, the psychological view of the self and the extended mind thesis, she argues that the self can actually be more extended than the body, even including objects external to the body, such as a notebook. First, she uses standard thought experiments to show that neither animalism, nor the soul theory is the correct account of human identity. Second, she considers arguments against the psychological theory and concludes that it has fewer ontological commitments and explains more of our intuitions. After that she defends a form of realizer functionalism, according to which mental states are the typical realizers of the functions identified with the mental states. However, she further argues that unlike in the brain theory, the realization base of the functions can include entities other than brain tissue and that the self can extend beyond the human brain and even the human body. This conclusion is reached by considering Andy Clark and David Chalmers' supposition that cognitive processes can be realized in non-neural realizers (the extended cognition view) and that some of these processes constitute narrative autobiographical memories, which, according to Milojević, track personal identity.

Brzović returns to the biological account of the self in her paper *The Immunological Self* and attempts to define an individual organism. She starts with several implausible definitions, such as ones based on functional integration or autonomy. The main part of the paper is spent over immunological definitions. Brzović critically assesses the self-nonself theory, according to which an organism is everything that is tolerated by the immune system. She shows that there are counterexamples to this theory and that it cannot account for the phenomena of autoreactivity, immune tolerance and symbiosis. Brzović also considers other theories. The systemic theory of immunity is rejected as too vague to be useful. Polly Matzinger's danger theory lacks precision

in the definition of danger. And finally, the continuity theory by Thomas Pradeu embodies a problem common to all of the immunological theories: it actually assumes that we can already identify an organism to be able to delineate those factors that trigger an immune response.

The next part of the collection focuses on epistemological issues.

Nenad Miščević ponders over the question whether self-knowledge has any value. In his *The Value of Self-Knowledge* he distinguishes between the direct knowledge of inner phenomenal states and the, mostly inferential, knowledge of one's causal powers, and assesses these from the perspective of instrumental value as well as intrinsic value. He believes that both the knowledge of phenomenal states and the knowledge of causal powers clearly have instrumental value. Insensitivity to pain and thirst or complete lack of knowledge of what one can cause and what one's reactions to external causes are is evidently inconsistent with the survival of such a deprived individual. But both types of knowledge also have intrinsic value. Drawing from John McDowell, Miščević maintains that if we did not have epistemic access to our own mental states we would turn into zombies and stop being who we are. Knowledge of phenomenal states is thus constitutive of our selves. Knowledge of our causal powers is also intrinsically valuable, according to Miščević, because such form of self-inquisitiveness is a virtue that contributes to the authentic self.

Luca Malatesti has contributed with a paper titled *The Self-Ascription of Conscious Experience*. As the title suggests he is interested in the process in which we ascribe conscious experiences to ourselves, and he takes as his model examples the experiences of color. He argues that a necessary condition for a thinker to be able to attribute conscious experience to herself is that she have the relevant concept, and having the concept is preconditioned by actually having the relevant experience. Malatesti thus rejects several theories in philosophy of mind, such as behaviorism, which are inconsistent with this supposition. Malatesti then analyzes what he calls the central transition, that is, the transition from having a certain experience to the knowledge that one is having that experience. He rejects the inner sense theory, because it relies on what he takes to be an implausible assumption that we have direct awareness of the self and the conscious experiences. He also questions the idea that the capacities that we employ in having conscious experience are the same as those that enable us to formulate self-ascriptions of conscious experience. Malatesti then formulates at least a necessary condition for the transition: the transition from judgments about how things appear to be to self-ascriptions of the relevant experiences requires the capacities that are involved in the mastery of observational concepts. Finally, he turns to the concept of self that is implied in such self-ascriptions. Although the concept seems to be implicitly present in each conscious experience, Malatesti argues that a richer notion is necessary. He sides with Alan

Millar and claims that the mastery of the concept of conscious experience involves the capacity to think about ourselves as entities that have sense organs and internal mental states that are determined by interactions with certain sorts of stimulation of these sense organs.

The next part of the collection maps certain historical views of the notion of self, offering papers on the accounts of the notion in logical positivism, Brentano, Buddhist Philosophy and Ancient Philosophy.

Boran Berčić in his *Logical Positivists on the Self* analyzes two broad questions: 1) What is the prominent logical positivists' opinion of Descartes' notion of *Cogito*? 2) What positive accounts of the self did they develop? First, Berčić offers various interpretations of *Cogito* and finds most plausible the interpretation according to which it is an inference from an attribute (thinking) to a substance (the thinker). Then he documents the key logical positivists' arguments rejecting the inference. He shows that Moritz Schlick considered it to be a mere stipulative definition, Rudolf Carnap dismissed it as a meaningless claim, because it cannot be formulated in classical logic which does not contain the predicate for existence, Weinberg considered the inference a tautology, and Alfred Ayer dismissed it as invalid, because, in his view, the presence of thinking does not necessarily imply the existence of the subject of this event. In the second half of the paper Berčić analyzes the theories of the self that the logical positivists offered. He shows that for Carnap it was a class of elementary experiences, for Ayer it was a logical construction out of sense-experiences and for Reichenbach an abstractum composed of concreta and illata. The logical positivists were thus reductionists about the self. In the final part Berčić shows how they dealt with a classical objection to reductionism according to which it is a circular theory.

In his *Brentano on Self-Consciousness*, Ljudevit Hanžek brings an overview of Franz Brentano's thoughts and arguments on the nature of consciousness and self-consciousness. Hanžek claims that Brentano's thoughts on self-consciousness were related to his ideas about introspection. Brentano maintained that introspection of our mental states is an impossible process, because it would require the division of the subject into an observer and the observed, which he believed to be impossible. However, there is a mechanism by which the subject can become acquainted with her own mental states. Brentano calls it inner consciousness and contrasts it with inner observation, that is, introspection. Inner consciousness is a process in which the subject is aware of an object and simultaneously also peripherally aware of the mental state of awareness of this object. Hanžek then shows how Brentano refuted the Regress Argument threatening his position, and lists several arguments that Brentano used to support his position. Finally, Hanžek questions an alternative interpretation of Brentano's position by Amie Thomasson and dismisses it as lacking support by textual evidence.

In the following paper, *The No-Self View in Buddhist Philosophy*, Goran Kardaš brings insight into the Buddhist perspective on the self, that is, more precisely, the idea that there is no self as we typically think. Kardaš begins by characterizing Buddha as what we would call today a conventionalist about language and empiricist and verificationist in epistemology. Then he shows how this philosophical background led Buddha to see through the illusion that language items and syntactical relations correspond to objectively existing entities and events. A special instance of this illusion is the belief that the expression "I" actually denotes an entity – my self. Next Kardaš walks the reader through the process of reduction and elimination of the concept of self in Buddhist thought. The paper concludes by a brief overview of later developments in Buddhist philosophy.

The last paper of this section focuses on the self in Ancient Philosophy, as the title suggests. The author Ana Gavran Miloš attempts to refute a line of thought in contemporary history of philosophy, according to which the Ancient Greeks did not have a concept of self equivalent to the modern, post-Cartesian notion of self predominant today. First, she characterizes the Cartesian notion of self as a self-conscious individual with a privileged access to her own mental states endowed with epistemic certainty about them. Next, she formulates a challenge according to which the Greeks did not have such a subjective-individualistic concept of self, because they discussed the notion of self under a wider problem of what it takes to be a human being objectively. Gavran Miloš then analyses the work of Plato, Aristotle and Epicurus to show that none of them considers the self, or the soul, in their terminology, to be such a subjective, self-conscious and epistemically privileged entity. However, the picture is different if we consider the purpose for which Ancient Greeks employed the concept of soul, namely for practical concerns in the pursuit of happiness. Here, the author claims, the notion of individuality finds its application, because the notion of eudaimonia is always a notion of individual eudaimonia, that is, one's own happiness.

Part IV of the collection focuses on the notion of agency and its relation to the central concept of the book. It consists of papers by Matej Sušnik, Filip Čeč and Marko Jurjako.

In the paper *Ideal Self in Non-Ideal Circumstances*, Sušnik focuses on the complex relationship between reasons, motivation and justification of our actions. He adopts the Advice model of internalism about reasons in claiming that one's reasons for action are dependent on the advice of one's ideal self. A challenge to this view is that the ideal self will sometimes be a markedly different being, and, as a result, a normal agent will not be able to do what the ideal agent would advise her to do. Further, in such a case it seems a mystery, according to Sušnik, why we should seek advice from our ideal selves, rather than just anyone ideally placed, which would undermine one tenet of internalism.

In the end, Sušnik defends a solution developed by Williams, which does not depend on the concept of ideal self. Williams claims that an agent has a reason to do x only if there is a sound deliberative route from the agent's motivational set to the agent's doing x, regardless of the fact whether the agent is actually able to do x. Sušnik then applies this theory to several problematic cases to show its explanatory power.

Čeč in his paper *The Disappearing Agent* takes the reader to the debate surrounding the notion of free will. He defends a form of event causal libertarianism, according to which a free action is the product of indeterministic, agent involving mental events or states, which do not rely on any specific form of selfhood (in contrast to agent causal libertarianism, which presupposes the self as an ontologically irreducible entity that has the capacity to cause free choices). Čeč develops the notion of torn decision – a decision in which the agent has two equally justified options and decides on one of them without resolving the conflict – to demonstrate an objection to event causal libertarianism. According to the disappearing agent objection, if the torn decision is not resolved by the agent, but by an indeterministic event, then the presence of the agent is quite irrelevant in the decision-making process, the agent disappears. In the second half of the paper Čeč lists and assesses the strengths and weaknesses of several possible answers to the objection. He favors the response that the causal libertarian should acknowledge that something gets lost in the decision-making process and that there will be some residual arbitrariness, but that the agent will not disappear from the process, because the decision will still be attributable to her.

In *Agency and Reductionism about the Self*, Jurjako discusses the relevance of agency for personal identity. His aim is to show that agency based accounts of personal identity are not necessarily incompatible with classical psychological continuity accounts. Jurjako begins with a detailed exposition of the psychological theory of personal identity and then shows how the theory is committed to reductionism about personal identity, that is, the idea that personal identity is not a further fact over and above facts about bodies, brains and their functions. Further, he shows that reductionism entails the fact that sometimes questions about personal identity will have indeterminate answers. In part 4 Jurjako focuses on what's called the *Extreme Claim*, that is, the claim that if reductionism is correct, we have no reason to care about our own future, and presents Parfit's solution, according to which the preservation of personal identity is not necessary for survival as long as psychological continuity is preserved. In the next part Jurjako turns to agency accounts of personal identity, with an emphasis on Korsgaard's theory. He challenges the alleged incompatibility between Parfit's and Korsgaard's theories and shows that Parfit's theory has the resources to account for agency. In addition, he argues that the resulting theory can avoid the Extreme Claim that threatens reductionism.

Part V focuses on issues related to the existence of the self. In *On Never Been Born*, Marin Biondić analyzes value claims related to existence and non-existence, such as whether it is ever good to be brought into existence or whether not ever existing is preferable to existing at some time. Biondić employs the *Reference Argument*, according to which such claims only make sense if there is a referent of the subject of the value claim. If the claim does not refer, it is meaningless. As a result, it makes no sense to compare the value of existence versus non-existence for a person who never exists. Biondić further explores whether we could use an analogy of the comparative account of the badness of death to make value claims about prenatal existence and non-existence. He claims the analogical argument enables us to say that being brought into existence can be good or bad for an actual person, even if not being brought into existence could not possibly be bad *for her* (according to the Reference Argument). Biondić then discusses the arguments of two philosophers who challenge this reasoning. David Brenatar's general argument that it is never better to come into existence is found only partially successful and Palle Yourgrau's theory is rejected because it entails possibilism, which, according to Biondić, would be a high price to pay.

In the next paper in this section Iris Vidmar analyzes the notion of *fictional characters*. She sets out by exploring a puzzling feature of fictional characters: they do not exist, but we still treat them a real in a sense. Vidmar then discusses logical, metaphysical and semantic theories of the existence of fictional characters, more specifically the realist ones, according to which fictional characters are real entities of a sort, and argues that these theories ignore the fact that fictional characters are artistic creations. She prefers to analyze them from literary-aesthetics perspectives, according to which their identity is indeterminate, open to interpretations, imbued with properties we recognize as human as well as purely artistic qualities. The main body of the paper consists of the author's defense of a multi-layered account of the identity of fictional characters, according to which their identity consists of aspects related to the author's activities in creating them and those involved in readers' activities in responding to them. Along the way Vidmar discusses how fictional characters come into being and vanish, how they represent certain types and classes, and how their identity is relative to our interests.

The final part of the collection offers three papers on the metaphysics and philosophy of language of the self. In the first paper of this section, *Haecceity Today and with Duns Scotus: Property or Entity?,* Márta Ujvári compares the current notion of primitive thisness, that is, haecceity, with its original counterpart developed by Duns Scotus. Ujvári shows that today haecceity is considered to be a non-qualitative property whose function is to guarantee trans-world identity and possible world individuation in modal metaphysics.

Next, the author looks at the definitions of the concept of non-qualitative property and questions its identification with impure qualitative properties in the work of several authors. She also poses other challenges for the current concept: Can haecceities exist uninstantiated? Do they really guarantee trans-world identity? How do they connect to individual natures? Ujvári then presents an alternative understanding originating with Duns Scotus, who considered haecceity to be a principle of individuation. She shows that for Scotus, there was a difference between particularity and individuality, and that haecceities were the means of securing individuality. But since according to Scotus every unity presupposes a unity-maker with entitative status, haecceities must be entities, not properties, as it is claimed today.

The penultimate paper in the collection, by Arto Mutanen, is titled *Who am I ?* The author claims that this question is in fact a cluster of questions with a host of different answers. First, he ties the question to the notion of identification and shows how it is handled in referentially opaque contexts. Second, he turns to the mind-body problem to distinguish the notion of identification from the notion of identity. Identity, he argues, is an ontological notion while identification is a methodological one, comprising the methods and techniques used to define an individual. In the second half of the paper the author assesses the approaches to identification by Hintikka, Russell and Gleason.

The final piece is titled *Meta-Representational Me*. The author Takashi Yagisawa analyzes the notion of the first person singular, the notion of *me*. First, he argues that the concept of *me* is different from the concept of *self* and is not reducible to it. Then he inquires whether one could grasp the notion by means of the semantic analysis of "I". He outlines Kaplan's indexical theory of "I" and claims that in spite of its plausibility it fails to account for the notion of *me*. In particular, it cannot account for the fact that the notion of *me* only applies to me, while anyone can use "I" to refer to himself or herself. Also, Kaplan's theory does not explain why "I" is a rigid designator, especially since its referent is not fixed causally, as is the case in typical "Kripkean" rigid designators. Yagisawa then goes on to argue that the notion of *me* is not a linguistic notion and that we can grasp the concept more adequately if we assume that it has its conceptual origin in representation. He shows that representation with the same content, object and recipient may occur in different ways and one specific form of representation is the *me*-way. Then he argues that we can extract the notion of me from the *me*-way representation by means of the so-called *way-to-thing shift.* Yagisawa concludes by giving a distinctive account of the rigidity of "I".

The collection *Perspectives on the Self* brings a representative selection of topics related to the notion of self. The editor Boran Berčić has done a good job collecting quality authors with a shared interest, and writing an introduction with a careful exposition of the contents of each paper. The

level of the papers varies. Some are intended as an introduction to the subject, providing an overview of the various positions in the debate under discussion, thus being more suitable for undergraduate students. Others are more challenging and technical, with an intention to move the relevant debate forward. These will be appreciated by graduate students as well as academics. The collection is recommended for anyone who would like to get a quality exposition of the problems of the self in many of its various connotations.

ABSTRACTS IN CROATIAN

RAZLIČITI ARGUMENTI, ISTI PROBLEMI. MODALNA DVOZNAČNOST I VARLJIVE SUPSTITUCIJE

RAFAL URBANIAK
Ghent University
University of Gdansk

SAŽETAK

S tri klasična primjera prikazujem greške što proizlaze iz korištenja modalnog operatora istovremeno uzimajući u obzir višestruke interpretacije unutar istoga argumenta; nedostaci proizlaze posebnom lakoćom ako se ne usmjeri pažnja prema rasponu propozicijskih varijabli. Premise se, gledane odvojeno, mogu činiti uvjerljivima, a zamjena za propozicijsku varijablu u modalnom kontekstu može se činiti legitimnom. No ne postoji jedna interpretacija uključenih modalnih operatora kojom bi sve premise bile plauzibilne, a zamjena uspješna.

Ključne riječi: Church-Fitchov paradoks, futura contingentia, modalna logika, modalni operatori, propozicijski kvantifikator, Swinburneov modalni argument

OBJAŠNJENJE I INDIVIDUALNA ESENCIJA

MÁRTA UJVÁRI
Corvinus University of Budapest

SAŽETAK

U ovom radu prikazujem da novije ontičko razumijevanje objašnjenja, kojemu je namjera da obuhvati *de re* esencijalne značajke pojedinaca, može poduprijeti kvalitativno gledište individualnih esencija. Nadalje je argumentirano da navodne štetne posljedice Leibnizovog principa (***PII***) i njegova suprotnost kvalitativnom gledištu, može biti izbjegnuta, pod uvjetom da se pojedinačne esencije ne konstruiraju u stilu naivne teorije svežnja s uvjetima teorijskog identiteta za skupove. Prihvaćanje ili

sofisticiranije dvoslojne teorije svežnja ili, alternativno, neo-aristotelovske pozicije gledanja na esencije kao na prirodnu narav u aristotelovskom smislu, može pomoći da se izbjegnu ove glavne optužbe protiv kvalitativnog gledišta. Funkcionalne paralele s alternativnim hekceitetskim gledištem individuacije i individualne esencije također će biti razmatrane.

Ključne riječi: kvalitativna individualna esencija, teorija svežnja, identitet nerazlučivog, aristotelovski esencijalizam, ontičko objašnjenje, argument jaza u objašnjenju

SPAŠAVANJE BRODA

JOHN BIRO
University of Florida

SAŽETAK

Braneći početnu tvrdnju da artefakti ne postoje, zapravo, nikakvi neživi materijalni objekti slične vrste, Peer van Inwagen je argumentirao da se istine o takvim navodnim objektima mogu parafrazirati kao istine koje ne čine esencijalno referiranje na njih i da bismo trebali prihvatiti samo ontološke obaveze parafraziranja. U ovom zapisu argumentiram da parafraze koje van Inwagen preporučuje ne mogu ispuniti njegove uvjete. Čitane na jedan način, gube nam neke istine. Čitane na drugi način, podrazumijevaju postojanje tih istih objekata kojih bi nas trebale riješiti. Međutim, ne trebamo dijeliti van Inwagenovu nenaklonost za potonje: reći da postoje nije isto što i reći da bilo što postoji kao dodatak nedjeljivim jedinicama što ga sačinjavaju.

Ključne riječi: van Invagen, parafraza, kompoziti, nedjeljive jedinice

EVALUACIJA ISTRAŽIVA(NJ/Č)A I NJENA PRIJETNJA EPISTEMIČKOM PLURALIZMU

MARCO VIOLA
Moscow State Pedagogical University, Russian Institute for Advanced Studies

SAŽETAKDok je jedan oblik evaluacije oduvijek bio upotrebljavan u znanosti (npr. recenzije kolega, zapošljavanje), formalni sustavi evaluacije istraživanja i istraživača nedavno su počeli igrati istaknutiju ulogu u mnogim zemljama zbog usvajanja novih modela upravljanja. Prema takvim modelima mjerena je kvaliteta učinka kako istraživača tako i njihovih institucija, a pitanja kao što su podobnost za održavanje ili raspodjela javnog financiranja istraživačkim institucijama, ključno ovise o ishodu takvih mjerenja. Međutim, porasla je zabrinutost zbog rizika da takva evaluacija može zaprijetiti epistemičkom pluralizmu tako što će kazniti postojeće heterodoksne škole misli i obeshrabriti težnju za novima. Predloženo je da bi se ovo moglo dogoditi zbog epistemičkih predrasuda što favoriziraju *mainstream* istraživačke programe. U ovom radu tvrdim da (1) je epistemički pluralizam poželjan i da bi ga trebalo sačuvati; (2) formalna provođenja evaluacija mogu ugroziti epistemički pluralizam jer mogu biti pogođena nekom vrtom epistemičke predrasude; dakle (3) da bismo sačuvali epistemički pluralizam, trebamo neku strategiju koja bi aktivno ublažila epistemičku predrasudu.

Keywords: ekonomska epistemologija, epistemički pluralizam, politika istraživanja, evaluacija istraživanja

Abstracts translated by:
Jelena Kopajtić, University of Rijeka, jelena.kopa@hotmail.com

INSTRUCTIONS FOR AUTHORS

Publication ethics

EuJAP subscribes to the publication principles and ethical guidelines of the Committee on Publication Ethics (COPE).

Submission

Submitted manuscripts ought to be:

- be unpublished, either completely or in their essential content, in English or other languages, and not under consideration for publication elsewhere;
- be approved by all co-Authors;
- contain citations and references to avoid plagiarism, self-plagiarism, and illegitimate duplication of texts, figures, etc. Moreover, Authors should obtain permission to use any third party images, figures and the like from the respective copyright holders. The pre-reviewing process includes screening for plagiarism and self-plagiarism by means of internet browsing;
- be sent exclusively electronically to the Editors (eujap@ffri.hr) (or to the Guest editors in the case of a special issue) in a Word compatible format;
- be prepared for blind refereeing: authors' names and their institutional affiliations should not appear on the manuscript. Moreover, "identifiers" in MS Word Properties should be removed;
- be accompanied by a separate file containing the title of the manuscript, a short abstract (not exceeding 250 words), keywords, academic affiliation and full address for correspondence including e-mail address, and, if needed, a disclosure of the Authors' potential conflict of interest that might affect the conclusions, interpretation, and evaluation of the relevant work under consideration;
- in American or British English
- no longer than 8000 words, including footnotes and references

The Editors reserve the right to reject submissions that do not satisfy any of the previous conditions.

If, due to the authors' failure to inform the Editors, already published material will appear in EuJAP, the Editors will report the authors'

unethical behaviour in the next issue and remove the publication from EuJAP web site and the repository HRČAK.

In any case, the Editors and the publisher will not be held legally responsible should there be any claims for compensation following from copyright infringements by the authors.

If the manuscript does not match the scope and aims of EuJAP, the Editors reserve the right to reject the manuscript without sending it out to external reviewers.

Style

Accepted manuscripts should:

- follow the guidelines of the most recent Chicago Manual of Style
- contain footnotes and no endnotes
- contain references in accordance with the author-date Chicago style, here illustrated for the main common types of publications (T = in text citation, R = reference list entry)

Book
T: (Nozick 1981, 203)
R: Nozick, R. 1981. *Philosophical Explanations.* Cambridge: Harvard University Press.

Chapter or other part of a book
T: (Fumerton 2006, 77-9)
R: Fumerton, R. 2006. The Epistemic Role of Testimony: Internalist and Externalist Perspectives. In *The Epistemology of Testimony*, ed. J. Lackey and E. Sosa, 77-92. Oxford: Oxford University Press.

Edited collections
T: (Lackey and Sosa 2006)
R: Lackey, J. and E. Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford: Oxford University Press.

Article in a print journal
T: (Broome 1999, 414-9)
R: Broome, J. 1999. Normative requirements. *Ratio* 12: 398-419.

Electronic books or journals
T: (Skorupski 2010)

R: Skorupski, J. 2010. Sentimentalism: Its Scope and Limits. *Ethical Theory and Moral Practice* 13: 125-136. http://www.jstor.org/stable/40602550

Website content
T: (Brandon 2008)
R: Brandon, R. 2008. Natural Selection. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 26, 2013. http://plato.stanford.edu/archives/fall2010/entries/natural-selection

Forthcoming
For all types of publications followed should be the above guideline style with exception of placing 'forthcoming' instead of date of publication. For example, in case of a book:
T: (Recanati forthcoming)
R: Recanati, F. forthcoming. *Mental Files*. Oxford: Oxford University Press.

Unpublished material
T: (Gödel 1951)
R: Gödel, K. 1951. *Some basic theorems on the foundations of mathematics and their philosophical implications*. Unpublished manuscript, last modified August 3, 1951.

Final proofreading

Authors are responsible for correcting proofs.

Copyrights

The journal allows the author(s) to hold the copyright without restrictions. In the reprints, the original publication of the text in EuJAP must be acknowledged by mentioning the name of the journal, the year of the publication, the volume and the issue numbers and the article pages.

EuJAP subscribes to Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). Users can freely copy and redistribute the material in any medium or format, remix, transform, and build upon the material for any purpose. Users must give appropriate credit, provide a link to the license, and indicate if changes were made. Users may do so in any reasonable manner, but not in any way that suggests the licensor endorses them or their use. Nonetheless, users must distribute their contributions under the same license as the original.

Archiving rights

The papers published in EuJAP can be deposited and self-archived in the institutional and thematic repositories providing the link to the journal's web pages and HRČAK.

Subscriptions

A subscription comprises two issues. All prices include postage.

Annual subscription:

International:

individuals € 10

institutions € 15

Croatia:

individuals 40 kn

institutions 70 kn

Bank: Zagrebačka banka d.d. Zagreb

SWIFT: ZABAHR 2X
IBAN: HR9123600001101536455

Only for subscribers from Croatia,

please add: “poziv na broj”: 0015-03368491

European Journal of Analytic Philosophy is published twice per year.

The articles published in the European Journal of Analytic Philosophy are indexed and abstracted in the Directory of Open Access Journals (DOAJ), The Philosopher's Index, European Reference Index for the Humanities – ERIH PLUS, PhilPapers, and Portal of Scientific Journals of Croatia (HRČAK).