Croatian Journal of Philosophy

Articles

Are We Causally Redundant? Eliminativism and the no-Self View
JIRI BENOVSKY

Imagination: A Sine Qua Non of Science
MICHAEL T. STUART

Bayesianism and the Idea of Scientific Rationality
JEREMIAH JOVEN JOAQUIN

The Grounding Problem for Panpsychism and the Identity Theory of Powers
NINO KADIĆ

The Philosophical Critique of Concept of Miracle as "a Supernatural Event"
ADAM ŚWIEŻYŃSKI

Maximization, Slotean Satisficing, and Theories of Sufficientarian Justice
ALEXANDRU VOLACU

Self-deception and Selectivity: Reply to Jurjako
JOSÉ LUIS BERMÚDEZ

Book Discussion

Possible Uses of Tennant's Methodology in Secondary Education
RUDI KOTNIK

Vol. XVII · No. 49 · 2017

Croatian Journal of Philosophy

1333-1108 (Print)
1847-6139 (Online)

Co-published by
"Kruzak d.o.o."
Naserov trg 6, 10020 Zagreb, Croatia
fax: + 385 1 65 90 416, e-mail: kruzak@kruzak.hr
www.kruzak.hr

and

Institute of Philosophy
Ulica grada Vukovara 54/IV, 10000 Zagreb, Croatia
fax: + 385 1 61 50 338, e-mail: filozof@ifzg.hr
www.ifzg.hr

Available online at http://www.ceeol.com and www.pdcnet.org

CROATIAN
JOURNAL
OF PHILOSOPHY

Vol. XVII · No. 49 · 2017

Articles

Are We Causally Redundant?
Eliminativism and the no-Self View
JIRI BENOVSKY 1

Imagination: A Sine Qua Non of Science
MICHAEL T. STUART 9

Bayesianism and the Idea of Scientific Rationality
JEREMIAH JOVEN JOAQUIN 33

The Grounding Problem for Panpsychism
and the Identity Theory of Powers
NINO KADIĆ 45

The Philosophical Critique of Concept
of Miracle as "a Supernatural Event"
ADAM ŚWIEŻYŃSKI 57

Maximization, Slotean Satisficing,
and Theories of Sufficientarian Justice
ALEXANDRU VOLACU 73

Self-deception and Selectivity: Reply to Jurjako
JOSÉ LUIS BERMÚDEZ 91

Book Discussion

Possible Uses of Tennant's Methodology
in Secondary Education
RUDI KOTNIK 97

Croatian Journal of Philosophy
Vol. XVII, No. 49, 2017

Are We Causally Redundant? Eliminativism and the no-Self View

JIRI BENOVSKY
Department of Philosophy,
University of Fribourg, Fribourg, Switzerland

Some friends of eliminativism about ordinary material objects such as tables or statues think that we need to make exceptions. In this article, I am interested in Trenton Merricks' claim that we need to make an exception for us, conscious beings, and that we are something over and above simples arranged in suitable ways, unlike tables or statues. I resist this need for making an exception, using the resources of four-dimensionalism.

Keywords: Eliminitivism, material objects, Trenton Merricks, consciousness, causality.

The eliminativist view about ordinary macroscopic objects like chairs or statues suits the taste of those who prefer desert landscapes to baroque complexity, and it nicely solves a number of problems (composition, vagueness, material constitution, coincidence, causal overdetermination,…).[1] It elegantly avoids these problems with ordinary objects since if there are no such objects, there are no worries concerning them. The issue I will be interested in this article is to see whether eliminativists need to make an exception—for us. Indeed, eliminativists such as Peter Van Inwagen or Trenton Merricks famously argued (for different reasons and in different ways) that while eliminativism is the best theory around when it comes to tables, planets, or statues, it is not to be endorsed in the case of humans—an exception is to be made. Van Inwagen focuses on *living* entities, and Merricks focuses on *conscious* organisms. In this article, I will examine Merricks' reason to make such an exception, and I will argue that it can be resisted.

The question is: can we eliminate the Self in the same—or similar—way we can eliminate tables and statues, without losing some-

[1] See Unger (1979), Van Inwagen (1990), Merricks (2001), and Heller (1990, 2008).

thing important? Under ordinary-objects eliminativism, we can without any loss eliminate tables and statues because there are simples arranged tablewise or statuewise and those can play the same practical and theoretical roles that tables or statues could play if they existed. (Eliminativism is also compatible with ontologies that do not postulate the existence of simples, but I will leave this issue aside here; see Benovsky (2016) for a detailed discussion.) In this view, chairs understood as *single* objects, can be eliminated because there is a *plurality* of objects that takes their place, namely, simples arranged chairwise. I believe that the same strategy can be applied to the case of the Self, although it needs to be articulated in a way that suits such a special case. The basic idea is the same: a single entity such as *the* Self can be eliminated because there exists a plurality of other entities, namely, successive impermanent psychological states/experiences arranged 'Selfwise'. I have articulated in detail and defended this view in Benovsky (manuscript); here, my aim is to defend it against an objection raised by Merricks (2001).

The idea of such an eliminativism about the Self is that we can be eliminativists about *us*, understood in a reified and ontologically committing sense of Selves, but that we don't lose anything—I can still say "I am drinking a beer " in a sense understood in terms of simples arranged my-body-wise and beerwise, and in terms of the existence of a succession of impermanent psychological states. This is how we can hold a unified and complete eliminativist view, with no exceptions. In Merricks' (2001) view, however, there is a *disanalogy* since entities like tables or statues are *causally irrelevant*—whatever they can cause, can be caused entirely by the simples that compose them. So, in his line of thought, this is one of the good reasons to say that tables do *not* exist. But, he adds that "we humans—in virtue of causing things by having conscious mental properties—are causally non-redundant" (Merricks 2001: 114). When I decide to run, there is a cause to be understood in terms of microphysics or microbiology, but there *also* is a cause to be understood in terms of my *decision*. It is my decision, Merricks says, that causes the simples to move as they do. This is why we cannot be eliminated in the same way tables can be, since we are causally relevant—we have causal powers over and above the causal powers of simples that compose us. Merricks' argument to the effect that we, human organisms, are causally non-redundant in virtue of having conscious mental properties is a complex and a very long one—indeed, it stretches on almost thirty pages (see Merricks 2001: Chap. 4). In what follows, let me focus on a (rather self-standing) part of his argument in detail and see how this step can be resisted—if it can, the overall argument will then not go through.

The main point of the argument is to show that we humans are not causally redundant, and that we have conscious mental properties that do "not supervene on what our parts are like" (Merricks 2001: 88). We

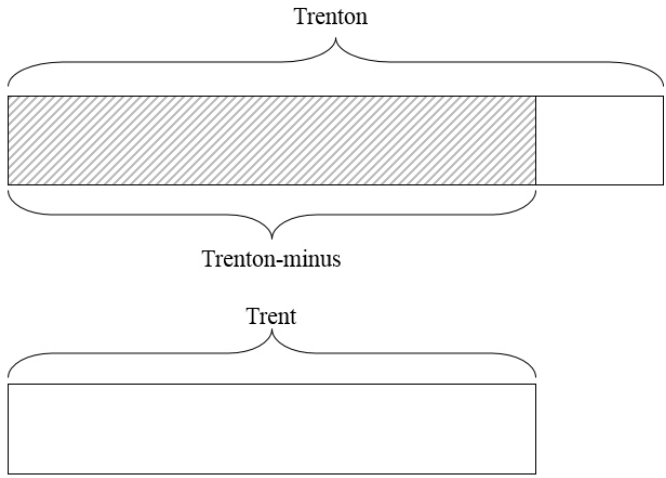
cause things in virtue of having these properties and we are therefore not causally redundant. (Thus, an argument based on the idea that we should eliminate anything that is causally redundant cannot be used to eliminate us—this is Merrick's overall main point.) On the way of defending this claim, Merricks argues for the rejection of:

> *Consciousness (C).* Necessarily, if some atoms $A_1 \ldots A_n$ compose a conscious object, then any atoms intrinsically like $A_1 \ldots A_n$, interrelated by all the same spatiotemporal and causal interrelations as $A_1 \ldots A_n$ compose a conscious object. (Merricks 2001: 94)

Here is, in short, Merricks's argument against (C)—it is a variant of the 'undetached parts argument'.[2] Suppose a small part of you is annihilated (say that your finger, or perhaps just an atom composing your finger, is cut away). Right after the amputation there is a conscious object—you—composed of some atoms. But these atoms existed in exactly the same way just before the amputation—indeed, they composed a (big majority) part of you. But, if (C) is true, this means that even before the amputation there was an object that was part of you—let's call it "you-minus"—that was a conscious object. So, it seems that before the amputation there were *two* non-identical conscious objects, namely you *and* you-minus. (By the same reasoning, there actually were *many* you-minus-like objects before the time of the amputation.) But this is false, since there was only one conscious object before the amputation. Thus, Merricks concludes, by *reductio* (C) is false.

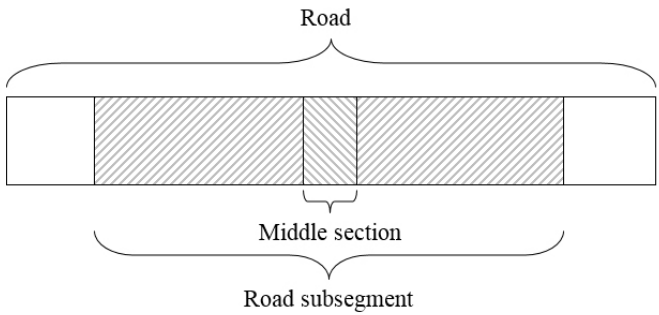
In case one would be tempted to answer the objection by appealing to four-dimensionalism and using talk about temporal parts to escape the unwelcome consequence that there were two conscious objects before the amputation, Merricks provides a temporal version of the objection as well, which can be formulated as follows. Take a four-dimensional person named "Trenton" who lives for 100 years. Take also another four-dimensional person, inhabiting the same possible world, who lives for only 80 years and is named "Trent". Suppose that Trent is microphysically intrinsically exactly like the temporal part of Trenton who lives for the first 80 years of Trenton's existence, and let us call this temporal part of Trenton "Trenton-minus", where Trenton-minus and Trent thus have atomic temporal parts exactly similar in intrinsic features and causal and spatiotemporal interrelations.

[2] See Van Inwagen (1981); for a discussion see *inter alia* Heller (1990) and Benovsky (2006).



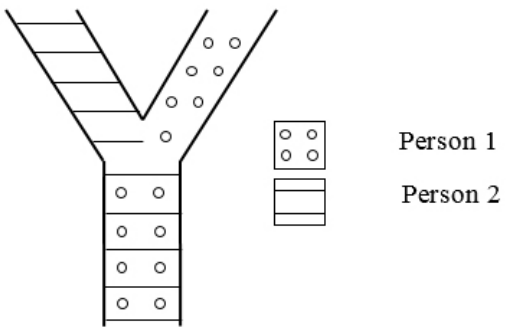
According to Merricks, according to the four-dimensionalist, assuming (C) for *reductio*, when it comes to Trenton between the age of 0 and 80, we then have a case where there are *two* coincident persons: Trenton *and* Trenton-minus. Trenton-minus is a conscious person in virtue of the existence of Trent and in virtue of the truth of (C). But such coincidence is unacceptable, and as a consequence, by *reductio*, (C) is false.

But the way Merricks presents the case here can be resisted. Indeed, this is not how four-dimensionalists typically describe the situation. Here is Sider (2001: 6), about the Statue and Lump famous case of coincidence: "At any given time it is only a temporal part of a spacetime worm that is wholly present. Thus it is only temporal parts of Statue and Lump that are wholly present at the time of coincidence. How can these temporal parts both fit into a single region of space? Because 'they' are *identical*." Sider then compares this to a case of a road:



There is a road that has a subsegment. In the very middle of the road, should we say that there are *two* entities, namely, the road and the road subsegment? Of course not. There is only one entity—the middle section—common to both the road and the road segment.

Similarly for Trenton and Trenton-minus. At a time where 'both' Trenton and Trenton-minus exist, should we say that they are two persons? Of course not. Before Trenton's 80th birthday, there always was only one person, exactly as in the middle section of the road there is only one road. It's just that this middle section is part of a road and of a road subsegment, and in the same way Trenton-minus is part of Trenton. This does not prevent Trenton-minus to be a person, as for instance David Lewis insists upon: "A person-stage is a physical object, just as a person is. (If persons had a ghostly part as well, so would person-stages.)" Lewis (1983: Postscript B). Typical examples of the way four-dimensionalism deals with such situations also involve cases of fission.[3] Let us say, for the sake of brevity, that for some reason a person undergoes fission, perhaps using a transporting device such as the one commonly used on the *USS Enterprise*, where due to a malfunction of the device, instead of simply transporting one person from one place to another, the device *also* leaves the original person behind. (You can replace this example with any other case of fission, if you don't like Star Trek stories.) Thus, after the fission, there are two persons, exactly alike. According to four-dimensionalism, we then have a situation where there are two four-dimensional persons, sharing an initial segment:



[3] I discuss one such case in detail in Benovsky (2013: 162–164).

Under a typical *endurantist* reading, this situation is one where the threat of coincidence is real: Person1-after-fission is not identical to Person2-after-fission, but Person1-after-fission is identical to Person1-before-fission, and Person2-after-fission is identical to Person2-before-fission—we then seem to have a situation where Person1-before-fission *is* identical to Person2-before-fission, that is, where these two persons seem to coincide in an unpalatable sense. But not so under the four-dimensionalist view where Person1-after-fission is *not* identical to Person1-before-fission, since these are two different temporal parts, numerically distinct, and similarly for Person2. In this way, (*i*) four-dimensionalists do not have to face the threat of coincident entities,[4] and (*ii*) they can say, relevantly to our present discussion, that there is only one person before the fission, in the same sense that there is only one person in the case of Trenton and Trenton-minus, and similarly in the spatial case of you and you-minus (in the finger amputation case). Metaphysically speaking, in all such cases where there seem to be two objects competing for the same space (and time), there really is only one, it's just that it's also part of other, spatially and/or temporally bigger, objects. It's like a wall that's common to two houses: if you need to repair it, you'll only need bricks to repair *one* wall, not two. In the way Merricks describes the situations he uses in his argument, there seems to be something like the principle that only 'the biggest' object is the one that counts. Thus, only Trenton, but not Trenton-minus is a conscious object. In his argument, appealing to the existence of Trent and to the truth of (C), Merricks then wants to force his opponent to recognize, for *reductio*, that Trenton-minus *is* a conscious object as well, thus creating a situation where there apparently are two distinct coincident objects that crowd each other out. But, as we have seen, four-dimensionalists do not, and do not have to, understand this situation (as well as the other similar situations) in this way.

One issue still remains. Who is doing the thinking, in the four-dimensionalist view? Is it the whole worm, or is it only a (rather short-lived) temporal part of the worm? If it were the whole worm, then—and only then—Merrick's objection above would go through. So, in order for the reply to work, we have to say that it is not the whole worm that has thoughts but that it has them only in virtue of having temporal parts that have them. A possible objection arises here:[5] say that a temporal part that lasts for only one minute thinks the thought "I have lived for 50 years". This is true, in a sense, because the whole worm did live for 50 years. But it is false, when thought by the temporal part, since the temporal part only lived for one minute. So, the same thought seems to be both true and false—how messy!

[4] I simply assume here, in agreement with Merricks, that such coincident entities are not acceptable.

[5] I would like to thank Trenton Merricks for raising this point in a discussion.

But this only points to a specific feature of four-dimensionalism (i.e. the worm view[6]). In this view, worms have most of their properties in virtue of the having of those properties by their temporal parts. The temporal parts can overlap, in the sense we have just seen, or in a fission scenario. Thus, what is being thought or said at some point by some temporal part is *ambiguous*. Take the case of fission we have seen above. Let us say that, *before* the fission, the person that is there is called "Jean-Luc Picard". From an atemporal standpoint such a name is then ambiguous, since it refers both to Person1 and Person2, and since these two overlap at the time before fission, but not at the time after the fission. But as David Lewis points out, such an ambiguity is perfectly harmless as long as the two bearers of the name "Jean-Luc Picard" are indiscernible—that is, precisely, before the fission (see Lewis 1983: 64–65). The need to distinguish the two persons arises only *after* the fission, and there is no ambiguity there, since there clearly are two persons, and we will use two different names to refer to them (even perhaps in a homonymic way). Similarly, what the one-minute-long temporal part thinks or says is ambiguous, and it is true under one disambiguation and false under another. The problem then easily dissolves as a mere case of ambiguity.

As a consequence, using four-dimensionalism to answer Merricks's objection, we can say that (C) is *true*, and thus one cannot use the alleged falsity of (C) to argue to the effect that we humans are not causally redundant because our conscious mental properties do not supervene on what our parts are like. And one cannot then use this as a reason to make an exception for us when it comes to eliminativism.[7]

References

Benovsky, J. 2006. "Four-dimensionalism and modal perdurants." In P. Valore (ed.). *Topics on General and Formal Ontology*. Polimetrica Publisher.

Benovsky, J. 2013. "Branching and (in)determinism." *Philosophical Papers* 42 (2): 151–173.

Benovsky, J. 2016. "Eliminativism and gunk." *Teorema* 35 (1): 59–66.

Benovsky, J. manuscript (draft). "Subjectivity and the no-Self view: How to survive eliminativism about the Self."

[6] In this article, I have focused on four-dimensionalism understood as the view that objects like tables or people are temporally extended entities, composed of temporal parts, since this is the view that Merricks appeals to in his argument, and since it is, to my mind, the best available version of four-dimensionalism. But note that another variant of four-dimensionalism, namely *the stage view* (see Sider (2001) and Varzi (2003)), can accommodate the truth of (C) even more easily, since according to this view conscious objects like persons are only instantaneous entities (that persist by having temporal counterparts at different times), and there is then no worry concerning amputation or fission cases since the person after the amputation or after a fission simply is a different person.

[7] I would like to thank Baptiste Le Bihan, Damiano Costa, Mark Heller, and Trenton Merricks for very helpful comments on an earlier version of this article.

Heller, M. 1990. *The ontology of physical objects: four-dimensional hunks of matter*. Cambridge: Cambridge University Press.

Heller, M. 2008. "The Donkey Problem." *Philosophical Studies* 140 (1): 83–101.

Lewis, D. 1983. "Survival and Identity." In Lewis, D. *Philosophical Papers*, vol. I. Oxford: Oxford University Press.

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

Sider, T. 2001. *Four-dimensionalism*. Oxford: Clarendon Press.

Unger, P. 1979. "There are no ordinary things." *Synthese* 41 (2): 117–154.

Van Inwagen. 1981. "The doctrine of arbitrary undetached parts." In Van Inwagen, P. 2001. *Ontology, Identity and Modality*. Cambridge: Cambridge University Press.

Van Inwagen, P. 1990. *Material Beings*. Ithaca: Cornell University Press.

Varzi, A. 2003. "Naming the stages." *Dialectica* 57: 387–412.

Croatian Journal of Philosophy
Vol. XVII, No. 49, 2017

Imagination: A Sine Qua Non of Science

MICHAEL T. STUART
Centre for Philosophy of Natural and Social Science
London School of Economics, UK

What role does the imagination play in scientific progress? After examining several studies in cognitive science, I argue that one thing the imagination does is help to increase scientific understanding, which is itself indispensable for scientific progress. Then, I sketch a transcendental justification of the role of imagination in this process.

Keywords: Imagination, understanding, thought experiments, scientific progress, schema, problem of coordination.

Blaise Pascal called the imagination that "deceitful part in man, that mistress of error and falsity." He said it was "all-powerful," and the "enemy of reason." Malebranche referred to imagination as the "madwoman in the house," and many fictional and historical catastrophes can indeed cite specific over-active imaginations at their roots. It is imagination that leads Goethe's young Werther to his infamous sorrows, and it is behind the ambition of Macbeth. Chapter eleven of *Mein Kampf* provides an actual and far more horrifying instance of the imagination being used to justify evil actions. According to George Orwell, Hitler saw himself as "the martyr, the victim, Prometheus chained to the rock, the self-sacrificing hero who fights single-handed against impossible odds. If he were killing a mouse he would know how to make it seem like a dragon" (Orwell 1940).

Yet, to reverse the sexist skepticism of Pascal and Malebranche, without imagination we could have no goals, no ethics, no knowledge. In the reflections of scientists we see tribute paid to the imagination quite regularly. Francis Jacob, a Nobel Prize winning biologist, recently wrote:

> It was not a simple accumulation of facts that led Newton, in his mother's garden one day, suddenly to see the moon as a ball thrown far enough to fall exactly at the speed of the horizon, all around the earth. Or that led Planck to compare the radiation of heat to a hail of quanta. Or William Harvey to

> see in the bared heart of a fish the thudding of a mechanical pump. In each case they perceived an analogy unnoticed up till then. As Arthur Koestler pointed out, everything in this way of thinking seems different from that of King Solomon when he compares the beasts of his beloved Shulamite to a pair of fawns, or that of Shakespeare's Macbeth, when he sees life as "a tale told by an idiot, full of sound and fury." And yet, despite the very different means of expression used by the poet and the scientist, imagination works in the same way. It is often the idea of a new metaphor that guides the scientist. An object, an event, is suddenly perceived in an unusual and revealing light, as if someone abruptly tore off a veil that, till then, had covered our eyes. (Jacob 2001: 119)

Jacob reminds us that no agglomeration of facts can give us the power over nature that science seeks, or the beauty and novelty of art. Dustin Stokes (2014) argues that even if Bach had known all there was to know about musical relationships, this still would not have been sufficient to compose *The Well-Tempered Clavier* (159–160). This resonates with Jacob's claim above; whatever is going on in scientific discovery, it is not merely the collection of facts. Other Nobel Prize winning scientists gesture to similar senses of imaginative artistry and its necessity in their work (e.g. Einstein 1931, 97, 1934, 163; c.f. Holton 1996, Hadamard 1996).

However, it was common in the philosophy of science for a long time to hold that the imagination was not epistemologically relevant other than in the context of discovery. Partially thanks to the growing influence of science studies since the 1960s, many philosophers and cognitive scientists have reversed this trend, and now see the imagination as an important factor in the production of knowledge and other epistemological desiderata. One reason for this change was the dissolution of an absolute distinction between the contexts of discovery and justification. Another is the recently emphasized role of the imagination in scientific thought experiments. René van Woudenberg, in his introduction to a special issue of *Metaphilosophy* on thought experiments, claims that "the imagination, perhaps surprisingly, plays an important role in the process of obtaining knowledge: knowledge of certain normative issues, of possibilities, of moral truths, of certain physical matters, of one's self, and more" (van Woudenberg 2006: 160; see also Byrne 2005, Currie and Ravenscroft 2002, Kind 2016, Kind and Kung 2016, McGinn 2004, Salis and Frigg forthcoming, Stuart et al. 2017).

To support such a claim, some philosophers have argued that because normative, modal and ethical truths are not accessible to empirical investigation, they must be the result of *mental* investigation (whether rational, as in Brown 2012, or naturalistic as in Nersessian 2007 or Miščević 2007). Considering possible worlds is one way the imagination might play a role in the divination of such truths. For example, the imagination is crucial in making the inference from conceivability to possibility, which is attacked and defended as a means (or mere guide) to modal knowledge (see Gendler and Hawthorne 2002).

I would like to look again at the epistemological role of the imagination in science, specifically through the use of thought experiments. Assuming that thought experiments play some role in scientific progress, I want to find out the nature of that role, and the nature of the epistemic good produced. To do this, I am going to present some results from cognitive science that ask what scientists and students of science learn from thought experiments, and how.

One problem with discussing the role of the imagination is that cognitive science studies rarely refer to the imagination in a general way. Instead they refer to mental images, analogy, metaphor, counterfactual reasoning, mental models, and so on. We find something similar in mainstream analytic philosophy, which deals with the imagination as something that tests modal propositions by seeing whether they are conceivable, or produces psychological states which obey special norms, and much else. (See Gendler 2013 for a sample of ways philosophers characterize imagination). In order to connect empirical and epistemological issues, then, I maintain an inclusive reading of the imagination, delimiting not much more than the mental ability to interact cognitively with things that are not now present via the senses. These cognitive interactions need not be propositional or static (like images), and to allow conceptual space for rationalism, their content need not consist entirely of permutations of previous experience. If we like, we could add the requirement that the cognitive interactions depart from the truth (following Stokes 2014), which is a reasonable requirement if we want to define the sort of imagination that goes into creating something truly novel, but I do not think it is necessary at this level of investigation. Imagining a Boeing 747 at the bottom of the Mariana trench is no less an imagining if there is in fact a Boeing 747 there.

One preliminary conclusion after looking at results in cognitive science is that an important and mostly overlooked use of scientific thought experiments is to create *understanding* as opposed to knowledge. Even though explaining how thought experiments increase scientific understanding would partially address the "primary philosophical challenge" of thought experiments (see Brown and Fehige 2011), many writers focus on the ability of thought experiments to provide new knowledge, empirical evidence or empirical information. Still, increasing understanding is just as epistemologically interesting as providing new knowledge, and in the second half of this paper I will investigate this use of thought experiments.

Let us now turn to results in cognitive science. Kosem and Özdemir have recently claimed that imagination "is an indispensible component of scientific reasoning" (2014: 887), and many others agree (e.g., Brown 2006; Clement 1993, 2008, 2009; Gilbert and Reiner 2000; Klassen 2006, Lattery 2001; Reiner and Burko 2003; Reiner and Gilbert 2004). Still, it is not immediately obvious how we should go about investigating scientific imagination. One way is to consider historical cases.

Stephens and Clement (2012) argue that even though such an exercise may be helpful, it is not enough to discern the cognitive mechanisms that underlie imagistic mental reasoning of the type we find in scientific thought experiments. They write:

> It is difficult to analyse the mental processes that allow a scientist to generate and run a thought experiment during an investigation by using historical data because the original thought process can easily be buried under many changes and refinements the author carries out before publishing a thought experiment. Also, for many thought experiments it is hard to know whether they were originally part of a discovery process or created after the investigation to convince others. (Stephens and Clement 2012: 160)

Historical details can only take us so far; we must also study thinking agents in real-time. I will summarize the results of several such studies here. First I will look at studies done on thought experiments in science education, and then I will consider studies of the way thought experiments are spontaneously invented in scientific problem-solving, both by students and experts.

Reiner and Gilbert (2000) discuss thought experiments in textbooks by first cataloguing which thought experiments appear where, and for what purpose. Then they compare the original and textbook presentations of famous thought experiments. They conclude that thought experiments help students and scientists understand scientific concepts. What does it mean for a thought experiment to help us understand something? They cite Stephen Toulmin (1972) who explicates understanding a concept in terms of being able to use it. A concept of any kind is capable of use, and therefore understood, if two criteria are met: it is meaningful, in that the user knows what it means; and it is fruitful, in that it enables the user to achieve a goal or to identify new possibilities. (For an extended discussion of these criteria as evidence of the achievement of understanding in science, see Stuart 2016).

I highlight this characterization of understanding because most of the below-mentioned studies are easily brought under its framework – especially if we include not just concepts but what can be called *theoretical structures*, a term I use to refer to concepts, models, theories, principles, laws, etc. Scientists and their students must be able to use new theoretical structures, otherwise they serve no purpose. And one cannot use a structure without knowing what it means, or in other words, without the structure being meaingful. Meaningfulness is not always so easy to achieve, especially in science, and we will see that thought experiments can sometimes assist in affording this desideratum. Also, if one understands a theoretical structure, one can usually achieve something with it. Thought experiments help us explore the consequences of adopting certain structures, and see how conceptualized phenomena interrelate, and this opens up new possibilities for theorizing, modeling, and constructing experiments.

Building on this framework, Reiner and Gilbert argue that thought experiments in science textbooks (as opposed to those in scientific jour-

nals), are not used as effectively as they could be. In scientific literature, most thought experiments are presented in the following way: We begin with a scenario or problem-statement. We create an imaginary world to help us explore the scenario or problem. We "set up" or "design" a thought experiment in this imaginary world, which we then "run" and "observe." Finally, we draw a conclusion about the initial problem or scenario. This presentation-style spurs members of its audience to make new connections on their own. Textbooks, on the other hand, often present the conclusion of the thought experiment first, and then the imagined scenario is introduced, which lends credence to the conclusion. In this style of presentation, students do not vary variables in their minds; they simply follow along a text (Reiner and Gilbert 2000). This is suboptimal for achieving the conditions of meaningfulness and fruitfulness. If you do not perform the thought experiment or otherwise establish the semantic connections for yourself, a theoretical structure will have diminished meaning for you. It is also less likely that you will see all the ways to make the structure fruitful. (For other ways of making theoretical structures meaningful and fruitful see Stuart 2017).

Velentzas, Halkia and Skordoulis (2007) look at textbooks as well, and they show that what James R. Brown calls "constructive" thought experiments (Brown 1991: 36), i.e., those that provide evidence for or establish a theory, are preferred by textbook authors to what Brown calls "destructive" thought experiments (Brown 1991: 34), which function as counterexamples. The thought experiments used most commonly in physics textbooks are Einstein's train, Einstein's Elevator, and Heisenberg's Microscope, which the authors classify as constructive. Perhaps these are so popular because thought experiments like these show students how their everyday experiences relate to modern day physical theory (Velentzas, Halkia and Skordoulis 2007: 365ff.). In other words, such thought experiments might "build bridges between students' knowledge and everyday experience and the new or modified concepts and principles which have to be learned" (359). Building such bridges would certainly help to make new concepts meaningful and fruitful for students.

This study inspires several more by Velentzas and Halkia. In the first (2011), they discuss Heisenberg's Microscope "as an example of using thought experiments in teaching physics theories to students." They begin by citing Alexander Koyré, who claims that thought experiments "help scientists to bridge the gap between empirical facts and theoretical concepts" (Koyré 1968). Agreeing, they argue that while Heisenberg's microscope thought experiment is not generally well regarded by physicists (either at Heisenberg's time or now), the thought experiment is still quite useful for introducing the uncertainty principle in quantum mechanics, which they taught to 40 high school students in grade 11 using the thought experiment. First, they introduced some important concepts from quantum mechanics, and then let the students work through the thought experiment mostly on their own.

That is, through Socratic question and answer, the students were allowed to work through their guesses, and if they went too far off track, they were gently guided back. Velentzas and Halkia recorded the sessions in order to code and analyze them, and two weeks later administered a test for comprehension. They concluded that many students did come to understand the uncertainty principle from the thought experiment. And not merely for the case of gamma rays and microscopes; they appreciated the principle independently of any considerations of specific measuring apparatuses.

Next they turned to special and general relativity (2013a). Again the authors found that thought experiments in relativity make it possible for students to "grasp physical laws and principles which demand a high degree of abstract thinking, such as the principle of equivalence and the consequences of the constancy of the speed of light to concepts of time and space" (3026). They found this achievement more surprising than in the case of the uncertainty principle, because students have very strong folk intuitions which interfere with understanding General relativity theory. Students generally did not understand the concept of inertia and they assumed that their intuitive concept of simultaneity could not be wrong, that space is empty and separate from time, and that an observer's point of view has no bearing on physical laws as there is always an encompassing frame from which an objective state can be observed (Arriassecq and Greca 2012).

However the authors did manage to convey the concepts of relativity theory to the students successfully, letting them work through Einstein's elevator and train. They recorded the sessions and analyzed them, and then administered a test two weeks later for comprehension. From their success they concluded that thought experiments are used "both for clarifying the consequences of physics theories and for bridging the gap between the abstract concepts inherent in the theories and everyday life experiences" (3027). Finally, in their (2013b) the authors turned to Newton's Cannon. As in the above two cases, the authors got a group of students to work through the thought experiment on their own, and to see that projectile motion and orbital motion are governed by the same laws. The authors claim that Newton's thought experiment "can act as a bridge which enables students to correlate the idea of the 'downward' motion of objects drawn from their everyday experience with the same objects' motion 'to the center of the Earth'" (2623). To make this possible, students had to see the Earth as if from above, and extend their knowledge of regular projectile motion to a scale large enough to represent both suborbital and orbital motion. This allowed them "to link the motion of a projectile as it can be observed in everyday situations with the possible case of a projectile that can move continuously parallel to the ground in a context where the whole Earth is visible" (2623).

The metaphor of "bridging" is common to all of these studies, and continues to be invoked below. I think it is significant because it relates to both meaningfulness and fruitfulness. When a bridge opens, new ter-

ritory becomes accessible. The territory was already there, but we did not have access to it. A theoretical structure is not made *fruitful* by a thought experiment if that thought experiment does not make possible new and identifiable uses of the structure, and one way it might do this is by connecting the theoretical structure via "bridges" to existing concepts, background theoretical knowledge, experiences and skills. Such activity can also provide semantic content to the theoretical structure, rendering it (more) meaningful.

Velentzas and Halkia conclude that thought experiments are useful in education because they help students learn to apply difficult scientific concepts. But there are two other interesting conclusions they draw in their (2013b). One is that thought experiments are pedagogically superior to computer simulations, because only in a thought experiment is it completely up to the student to determine how the outcome of an imagined scenario results from the set-up. A computer simulation where the earth is seen from above and the student can program in different projectile velocities and see how these changes affect the motion of a projectile was useful, and certainly better than merely calculating consequences of Newton's laws. But in these cases the student takes a passive role by setting the parameters and waiting to see what happens. In a thought experiment, students mentally "set" the parameters, and then in addition have to figure out what will happen. And instead of trusting to the algorithms of a computer, students must provide some reason to believe the system will evolve as it does in their imaginations. Also, talking through imaginary scenarios enables teachers to see where a student stands with respect to their comprehension of the theory. Therefore the authors conclude that there is good evidence that thought experiments will not be replaced by computer simulations in the near future, at least in the classroom.

This is related to their second important conclusion, that "in any experiment, the manipulation of ideas is more important than the manipulation of materials" (2638). That is, "hands on is less important as compared to minds on" (Duit and Tesch 2010). Presumably the authors mean that manipulating laboratory equipment is pedagogically less useful to a student who does not grasp the deeper meaning behind these events. And with respect to the goal of increasing scientific understanding, this is something worth stressing.

Now that we have discussed some of the findings of thought experiments in science education, let us look at how thought experiments originate *in situ*.

In "The Symbiotic Roles of Empirical Experimentation and Thought Experimentation in the Learning of Physics," Reiner and Gilbert argue that in the course of solving empirical problems, subjects often construct and run thought experiments spontaneously. They conclude that "the process of alternating between these two modes—empirically experimenting and experimenting in thought—leads towards a con-

vergence on scientifically acceptable concepts" (2004: 1819). In other words, thought and empirical experiments appear in conjunction, and this is for the best, because together they enable us to go from "seeing" a physical phenomenon to "knowing" about it (1820). The evidence for this is the following.

Reiner and Gilbert asked students to analyze a physical mechanism that behaved in an unexpected way. Two heavy wheels were set next to one another into a base, and each was free to spin. If one was made to spin quickly, the other would do nothing. But as it slowed down, the other would begin to spin and speed up, until the first came to a complete stop. When the second wheel began to slow down, the first would start spinning again. The reason for this behaviour was a set of hidden magnets contained in the wheels. Given a list of the materials out of which the mechanism was built, the students were asked to figure out what was going on. Different sets of students were all observed to follow a similar methodology: they began by identifying the various physical mechanisms in a general way using concepts like force, acceleration, weight, direction, and so on. They used these to construct various (mental or physical) models to capture what they observed. Then they abstracted their models further into what the authors called a "representational space," where the relationships between features of the mechanism were represented, often with the help of pen and paper. Finally, the students created and used imaginary worlds to test their models using thought experiments.

The authors claim that instances like these show "how concepts emerge out of touching and seeing. A student forges links between the bodily and the mental, between the physical and the cognitive, faculties" (2004: 1831). Despite the reference to knowledge, the epistemological state in question is better described as understanding. Most of the knowledge discussed in traditional epistemology is propositional. And links between bodily, mental, physical and cognitive faculties, while they can be expressed in propositions, are not propositions themselves. Also, establishing connections between parts of theory and experience is typically referred to as "objectual understanding," which is grasping the "coherence-making relationships" in a comprehensive body of information (Kvanvig 2003: 192). And it can be produced by thought experiments (Stuart 2017).

In a different study, Kösem and Özdemir (2014) collected three groups of subjects, each with a different level of expertise in physics, and presented them with difficult problems drawn from dynamics or mechanics. The first group was made up of doctoral graduates, the second was university undergraduates, and the third was high school students in grade 12. The total number of thought experiments invented by each of the three groups was roughly equal.

In terms of the means of the thought experiments, each student either modified an object in an imaginary scenario (for example, the size

of a car), or a variable (its velocity). When they modified the object, they did so either to match a more familiar case with which they had previous experience, or a simpler case, for example, by dissecting a problem into several smaller, easier problems. When they modified a variable, they either eliminated or minimized the variable's value to eliminate its influence altogether, which helped them focus on the relationships of other variables, or they increased the value of a variable to make its effect on the system more obvious.

In general, changing the problem to a more familiar case by modifying the object was the most common type of thought experiment strategy used by the undergraduate and high school groups. Modifying the variables was used quite often by the doctoral group, and very seldom by the others. In terms of purpose, there were several. Sometimes a subject would have an intuition, which they explored with a thought experiment. This use was labelled "prediction." Other times a subject might have an independent reason for believing something, which they chose to illuminate with a thought experiment while trying to report or justify it. This was labelled "explanation." Other times the thought experiment played the role of a proof. The undergraduates used thought experiments as a proof more than any of the other groups. The high school students and doctoral graduates very rarely used thought experiments as proof. Across all three groups, however, by far, "the most frequently observed purpose of using a thought experiment is for 'explanation'" (882). That is, "to communicate ideas, or exemplify the solution" (879).

Finally, there are studies focused on the use of imagination by expert scientists in vivo. First, Trafton, Trickett and Minz (2005) ask if scientists use the imagination to manipulate mental representations. They conclude that they do. They argue that scientists create what Clement later calls "overlay simulations" (2009) between external and mental representations. That is, they compare and align mental and external representations, checking for fit or feature-similarity. The authors found that the scientists manipulated spatial representations more often in their heads than they did using their computers (2005: 97).

In a second study, Trickett and Trafton (2007) built on these results, arguing that scientists spontaneously invent "small-scale" or "local" thought experiments (867) in times of "informational uncertainty" (843). Scientists perform thought experiments in such conditions to "develop a general, or high-level, understanding of a system" (844). The authors focus on the data analysis phase of research, in which scientists must negotiate uncertainty to see what information the data presents, and interpret it. Employing "what if" reasoning helps scientists test out alternate interpretations of the facts, fill in holes in their data, and see how their data fit with existing research questions and background theories. They predict that thought experiments "will be used by experts when they are working either outside their immediate area of expertise or on their own cutting edge research—that is, in

situations that go beyond the limits of their current knowledge" (867; cf. Corcilius 2017 who argues that this is (roughly) how Aristotle used thought experiments)).

If the empirical results I have mentioned are on the right track, there is a great deal that is philosophically interesting here. In almost every one of the above studies, one of the main conclusions is that thought experiments are important because they bridge conceptual/theoretical knowledge to previous experience, existing knowledge and abilities.

What does this tell us about the epistemological role of thought experiments in science? If we separate the action of bridging existing instances of knowledge from the action of creating new instances of knowledge, we see that thought experiments are often instances of the first kind of action, whether or not they are instances of the second. Thought experiments are more often used to explore or interpret conceptual solutions to problems, communicate ideas, or model scenarios, than they are to provide solutions to problems. That is to say, the performance of a thought experiment usually increases understanding rather than producing new knowledge. In fact, Özdemir (2009) argues that students learn to shy away from using thought experiments as evidence in physics as they mature, although they do not shy away from using them to communicate and explore. It is possible that this trend maintains itself in the professional careers of scientists everywhere.

It is also important that all of the above studies produce results that support the idea that thought experiments create understanding in one of the two ways mentioned at the start. Velentzas and Halkia showed that students use thought experiments to bridge empirical knowledge and theoretical structures. Gilbert and Reiner saw a symbiotic relationship between thought and empirical experiments, which were performed in a way that "negotiated concepts" through communication and exploration, making a student's concepts and models intelligible to him or herself, and also to his or her peers. Stephens and Clement argued that thought experiments "appear to have considerable value as a sense-making strategy" (2006: 1). Kosem and Özdemir found that the most common use of thought experiments across different groups was to "communicate ideas or exemplify a solution." Trafton, Trickett and Mintz found scientists employing thought experiments to compare, align and manipulate representations, especially for communication. For each of these cases, the value of sense-making thought experiments derives at least partially from the fact that if we do not make sense of a theoretical structure we cannot make use of it.

Let us turn to some considerations of these results. First, these roles that we have just identified are epistemological. And since these roles produce understanding as opposed to knowledge, we are able to draw on the quickly expanding resources in the philosophy of understanding. Understanding, like imagination, was rejected as a topic of

serious study in the philosophy of science around the time of the logical positivists, because it was associated with a psychological and subjective *feeling* (especially by Hempel; see de Regt et al 2009: 3–5; de Regt 2009: 22–24). This feeling might be an outcome of good science and provide clues concerning what should be investigated next (see Lipton 2009, Grimm 2009, Thagard and Stewart 2011), but it might also be irrelevant or misleading (see Ylikoski 2009). Leaving the positivist- era characterization behind, philosophers now consider understanding in many different senses.

As with "thought experiment," a vague but useful term, we can say interesting things to differentiate understanding from other epistemological states in the absence of a necessary and sufficient definition. One kind of understanding is "mediated," that is, it comes by means of a model, an experiment, a theory, a thought experiment, an explanation, or something else. One way to know if such mediating entities provide increased understanding is to ask about abilities. When we understand something, we can use it in new ways. We can relate what is understood to new and old knowledge, and to abilities we already had. This is why I continuously return to the meaningfulness and fruitfulness of theoretical structures. If we look at the thought experiments used in the aforementioned studies and in the history of science, we see that very many make some concept(s) more meaningful and fruitful, and so increase (evidence of) understanding. In Stuart (2016) I showed that Maxwell's demon, Darwin's eye, and the clock-in-the-box all provide this sort of understanding by connecting theoretical structures to experience, existing knowledge or abilities. Others try and fail, including Heisenberg's microscope and Darwin's whale. I think we can extend the argument easily to many other thought experiments including Einstein's elevator and train, EPR, Galileo's falling bodies, and Stevin's prism. If thought experiments perform this function, this is no obstacle to their also serving as evidence for or against theoretical claims. That is, they could provide both understanding and knowledge, although it is understanding I am interested in here. How might thought experiments provide both knowledge and understanding?

First, I hope it is clear that the same thought experiment can have several different uses at different times or for different people. For example, Schrödinger's cat was once used to attack the Copenhagen interpretation of quantum mechanics, and now it is used to introduce physics students to superposition and entanglement. One might argue that we have here two different thought experiments, but it is the same imagined scenario drawing on similar underlying assumptions, even if it is used for a different purpose in the two cases. If it is the same thought experiment, then the same thought experiment is at one time used by experts as an argument against a theory, and later by teachers and students for pedagogical reasons (see Bokulich and Frappier 2017 for more on the identify conditions of thought experiments). Now, is

it possible that the same thought experiment can play more than one epistemic role, for the same person (or community), *at the same time*?

Yes: thought experiments like Heisenberg's microscope, Schrodinger's cat, Einstein's elevator and others, are simultaneously used by scientists to make sense of difficult new theoretical structures, which increases their scientific understanding by helping them connect abstract theoretical structures either to experience or to previously unconnected parts of theory. In addition to serving this purpose, many of these thought experiments simultaneously or derivatively use this new understanding to attack, subvert, popularize or explain a theory or theoretical interpretation. The application of new understanding often results in new knowledge.

There is a complementary idea present in the work of Hans Radder on laboratory experiments (1996), which Sören Häggqvist (1996) and Tim de Mey (2003) applied to thought experiments. The idea is that the *performance* of an experiment is different from the *application* of the result of that experiment to theory. These two actions are often conflated in general discussions of scientific experiments. What I am suggesting is that sometimes the performance of a good thought experiment yields understanding, while the application of the result of that experiment yields knowledge.

What is novel here is that thought experiments are quite frequently significant for scientific understanding and not merely for knowledge. This idea has some nice consequences. For instance, it explains why many of the more famous thought experiments appeared in the later stages of their respective scientific revolutions. This is because they were meant to make sense of a new theoretical structure that had been introduced during the course of the revolution. If this is the case, the thought experiment could not have shown up earlier. The new quantum formalism was mathematically complete and empirically adequate by 1925, and Schrödinger's cat was not born until a decade later. Similar relationships obtain between Maxwell's demon and the statistical-mechanical interpretation of heat, Einstein's train and general relativity, the clock in the box and quantum mechanics, and many others.

This idea also helps to explain the role of thought experiments in the rhetoric of science. If you can provide an intuitive interpretation of a theory, this can be a way to get others to accept that interpretation, and therewith, the theory. If I am convinced of the Copenhagen interpretation of quantum mechanics, it is necessary that I am also convinced of quantum mechanics. Likewise, for those who oppose a new and competing theory, the first reaction is often to look for counterexamples, cases where the theory does not apply or that the theory cannot explain. And searching for counterexamples is itself an attempt to explore the connection between the new theory and the world (i.e., gain understanding of the theory's empirical content), and show that the proposed connection cannot be made (i.e., gain knowledge through falsification).

This also explains the prevalent place of thought experiments in science textbooks and websites which aim to describe in general outline how this or that modern scientific theory works or what its content is. Thought experiments help students take the steps their intellectual ancestors took in order to understand a theory. And even if someone does not understand all the difficult theoretical structures invoked by a theory, they might still grasp some of the relationships between those structures and their previous experiences and knowledge via a thought experiment.

This interpretation also sheds light on the role of thought experiments in scientific theory proliferation and "public marketing." If a theory has been developed in great theoretical or mathematical detail, but has not yet caught the eye of the greater scientific community, perhaps it is time to try some thought experiments. These may assist in securing funding and improving the theory's public image, since granting agencies and the public must be able to understand the theory to see it as pursuit-worthy. Late night infomercials on television encourage you to imagine yourself in some uncomfortable situation, from which only the Brand New Shining Product can save you. Thought experiments can also be powerful tools of advertisement that appeal to our intuitions and emotions via the imagination. Recognizing this power illuminates a new danger in thought experiments that was hidden until now: Since high-level understanding is one of the goals of science and thought experiments can provide it, they might be used (intentionally or not) to deliver merely apparent and not genuine understanding. Heisenberg's microscope is a potential example. While it does provide a way to visualize the uncertainty principle, it has been criticized harshly for doing so in a misleading way (see, e.g., Roychoudhuri 1978).

This is an interesting issue, because general understanding, while a desideratum, might not always be achievable. Our cognitive abilities might not always be sufficient for understanding our theoretical structures. Perhaps it has already happened in science that we have abandoned a good theory for a rival that was more easily intuited and understood, although false. Physicist Paul Dirac "regards models, images, pictures not only as redundant, but as dangerous. As long as the formalism and experimental results dovetail, theoretical physics has achieved its task" (quoted in Yourgrau 1967: 866). The Aristotelian theory of motion including natural places for the five elements strongly appeals to the imagination and is easy to understand, and this is surely one of the reasons it was dominant for so long. This is a problem that needs to be understood, and accounted for, although there is some reason for optimism. It is true that once we pass into the microscopic domain or higher dimensions we find it difficult to perform some kinds of imaginings, but this does not stop us from focusing on *aspects* of those systems that we *can* imagine. The entities that make up our world display a multitude of interesting properties, many of which stand in rela-

tions that can be visualized even if others cannot. The lesson is that, the more complicated our theories become, the more careful we must be with our imaginary examples.

Finally, if thought experiments provide understanding, they serve a function which is indispensable for the progress of human science. Without understanding, we cannot use our knowledge, and without knowledge there is less to understand. These two features of science can develop independently for a while, but not for long. Even with the greatest division of intellectual labour, one stagnates in the absence of the other. According to Peter Kosso: "knowledge of many facts does not amount to understanding unless one also has a sense of how the facts fit together" (2006: 173). He invites us to recall the Omniscienter from Pierre Dumal's novel *A Night of Serious Drinking*, over whose chair it reads "I know everything, but I do not understand any of it." Kosso suspects that "the Omniscienter has spent too much time gathering evidence and too little time thinking about it. He has taken the piecemeal empiricism too seriously and overwhelmed his science with observation. Too many data have left too little room for understanding. There are examples of knowledge without understanding in the physical sciences, and they are found in the most empirically dependent sciences or in any science at the time of new empirical discovery" (182).

To this end, Steven Weinberg remarks that general relativity offers more understanding than does quantum mechanics, because the latter cannot easily be bridged to our other stores of knowledge. He sees the Copenhagen interpretation as a surrender to the incomprehensibility of the theory, throwing up our hands and asking for empirical accuracy only (Weinberg 1992, Kosso 2006: 184). If it is true both that we need to understand our theories, and that quantum mechanics is inherently difficult to understand, then we should expect a great deal of thought experiments in quantum mechanics, especially in the first decades after the theory was introduced. And indeed, this is probably the period most replete with thought experiments in the history of science. (See Peacock 2017 for a detailed look at many of them).

Many scientists explicitly seek connections between their theories and the world or other pieces of knowledge, which I have characterized above as a search for understanding. Ernst Mach remarked that there has to be what he called "coordination" between the variables of a theory and the aspects of the world to which it refers (see van Fraassen 2008). The temperature reading taken from a thermometer must refer to something real, not to another conceptual entity. Reichenbach extended the problem, noticing that even the coordinating relation, if we could create one, would only be another abstract relation, which we would again need to coordinate (1965). Einstein remarked that if we want to talk about rigid bodies and their behaviour, we must first coordinate "experience[e]able objects of reality with the empty conceptual schemata of axiomatic geometry" (Einstein 1921). Einstein also spoke

of the "ever-widening logical gap between the basic concepts and laws on the one side and the consequences to be correlated with our experiences on the other—a gap which widens progressively with the developing unification of the logical structure, that is with the reduction in the number of the logically independent conceptual elements required for the basis of the whole system" (1934: 165). In other words, scientists recognize the need for something to bridge the gap between our theoretical structures, including laws, concepts, equations and mathematical models, and the world. Further, Einstein considers the possibility that as physics becomes more refined and united, it must make use of more and more abstract notions and relations to connect all its information to experience.

There is evidence that scientists have intentionally used thought experiments to solve this problem of coordination. Heisenberg showed in 1925 that the matrix and wave-mechanical formalisms of quantum mechanics were mathematically equivalent. Still, Schrödinger was set on the wave mechanical interpretation, and Heisenberg on the particle interpretation. According to Marten Van Dyck, Schrödinger called Heisenberg's theory a "formal theory of frightening, indeed repulsive, abstractness and lack of visualizability." And "'Heisenberg's theory in its present form is not capable of any physical interpretation at all,' was another claim made at the same time" (2003: 81). In response, Heisenberg began considering whether an interpretation focused on the particle nature of atomic elements could be visualized, and specifically whether in-principle observables could be simultaneously measured. "This was a turning point for Heisenberg's theory, because it led him to propose a visualizable interpretation of quantum mechanics *through thought experiments* based on the limits of measurement. Heisenberg wrote out all his ideas in a letter to Pauli at the end of February [1927], in an attempt, he said, to 'get some sense of his own considerations' as he groped towards a consistent theory" (Beller 1999: 105; emphasis added). Kristian Camilleri writes, "Heisenberg's introduction of the imaginary gamma-ray microscope was not intended primarily to demonstrate the limits of precision in measurement. Though it certainly did this, its real purpose was to define the concept of position through an operational analysis. This becomes evident once we situate Heisenberg's use of imaginary gamma-ray microscope within the context of his concerns over the meaning of concepts in quantum theory" (2007: 179). Heisenberg's thought experiment was therefore a way to link the new theoretical structure to some empirical content, whether through operationalization or visualization, *for Heisenberg*, in dialogue with his peers.

And this goes for many of the physicists of the period. Mara Beller writes, "most physicists, Bohr and Heisenberg included, wanted more: some feeling of understanding, of illuminating, or explaining the kind of world that quantum formalism describes. The need for this kind of

metaphysical grasp is not merely psychological but social as well—the power of a successful explanation and the power of the effective legitimation and dissemination of a theory are connected" (Beller 2002: 107).

This supports the notion that the understanding provided by thought experiments is important for many reasons, including pedagogy and popularization. But more importantly, it shows that scientists have been aware of this, and have used thought experiments for this purpose.

What have we learned so far? Thought experiments have many epistemological uses, many of which generate understanding as opposed to (or in addition to) knowledge. And the imagination plays some role in this. How does it work? Perhaps those who characterize thought experiments as mental models have an answer. Nenad Miščević (2007) argues that the power of the imagination results from its having evolved as a useful predictive tool with its roots in normal perception. Nancy Nersessian agrees, stating that "the perceptual system plays a significant role in imaginative thinking," which "makes sense from an evolutionary perspective" (2007: 136). While Nersessian does not claim that all the content that is manipulated by our mental models is perceptual or imagistic (142), she does "contend that a wide range of empirical evidence shows perceptual content is retained in all kinds of mental representations" (139). What grounds the epistemic use of thought experiments for Miščević and Nersessian is experience itself, and the usual cognitive and sensory faculties that provide empirical knowledge. Perhaps their justification of the outcome of thought experiments through mental or neural mechanisms can also be used to help explain the epistemic value of thought experiments conceived as producing understanding. Let us examine this claim.

The idea that we manufacture complex ideas from sensory experience via reason and imagination has its modern roots in British Empiricism. It is still well-supported empirically (see e.g., Prinz 2002) and is introspectively attractive. Nevertheless, there is something about the use of imagination in producing scientific understanding that seems left out of such a justification. The thought experiments discussed above do not succeed because the imagination has its roots in perception or other cognitive processes that evolved to represent the world accurately. We might be right to trust *knowledge* claims concerning the output of an imaginary scenario that accurately models a system with which we have relevant experience. But in producing new meaning or new abilities, we do not need representational accuracy; we only need to create the right bridges between theory and experience, however that is done. And sometimes increasing representational accuracy would hinder rather than help. It's hard to conceive of Maxwell's demon in a more realistic way doing the same job. Einstein's elevator succeeds in giving content to the equivalence principle because it takes us *away* from normal perception and gives us a new means of conceiving

the world. Since this use of the imagination is different from the one that generates new knowledge, we need a different justification for it.

To do this, I will set up an analogy. Just as imagination can help to determine the content of perception, thought experiments can help us to determine the content of theoretical structures.

There is support in cognitive science for the view that imagination can influence the content of perception. First, patients who have damaged parts of their neocortex sometimes cannot see conceptualized objects, like, e.g., ducks. They can only see lines, shapes and patches of colour (Thagard 2010: 70). One explanation for this phenomenon is the absence in these patients of "top-down processing," which is now "central in modern neuroscience" (Burchard 2011: 69). Top-down processing occurs when we first categorize or cognize things in broad strokes, and work through the details later. Those details are perceived as aspects of the more general conceptualized object, which means that the higher centers of our brain help to determine what we perceive. When top-down processing is operative, higher centers in the prefrontal cortex of the brain track and modify what happens in lower centers. When something new or difficult to identify is presented to a subject, top-down processing starts before recognition of the object is accomplished. According to Miller and Cohen (2001), the prefrontal cortex accomplishes this by providing bias signals to lower brain structures. These bias signals guide the flow of neural activity along certain pathways. In other words, when we see something new, parts of our brain normally associated with conscious thought are already involved in categorizing and making sense of the thing as it is presented to us (see also Buschman and Miller 2007). And imaginings can certainly be conscious. For example, it is well-known that if we approach an ambiguous figure (like the duck-rabbit of Gestalt psychology) with a certain mental image in mind, this can determine what we will see when we look at it. There is also support from the literature on "cognitive penetrability" (see e.g., Arstila 2017). I take all of this as evidence that the imagination, insofar as it is located in the higher centers of the brain, can play a constitutive role in determining the content of sense-experience.

Second, Mark Johnson writes in *the Body in the Mind* that we "connect up" (1987: 152) abstract mental structures with the contents of our sense perception using what he calls "schemata," which are "non-propositional structures of imagination" (19). He says "Even our most simple encounters with objects, such as the perception of a cup, involve schemata that make it possible for us to recognize different kinds of things and events as being different *kinds*" (20). Johnson's schemata have been very influential in cognitive science, and after the idea was re-expressed in Lakoff and Johnson (1999), it spawned what may be called a subfield of research. The basic idea is that through the imagination, we create schemata that give content to our beliefs, and structure perception and thought.

Nigel Thomas, a long time researcher of mental imagery, understands schemata as data structures in the brain that make possible our perceptual experience of the world (Thomas 1999). Thomas understands schemata slightly differently from the Johnson-Lakoff school, but he admits that the views are compatible, and again the imagination plays a crucial role. Thomas argues that schemata are not things that we experience, although they are necessary for experience in general.

Finally, there are sources of support for the fundamentality of imagination for sense experience that are more general. Stokes argues from a philosophical perspective that imagination is necessary (although not sufficient) for the formation of new beliefs, desires, intentions, as well as for learning new concepts and skills (Stokes 2014: 179–180). And Colin McGinn (2004) argues that imagination is necessary for all cognition, since it is necessary for grasping meaning.

If we grant the possibility that imagination can structure perception in a subconscious or non-occurent way, we can begin the analogy to thought experiments and the theoretical structures of science. Just as the imagination functions at the most fundamental level with respect to conceptual content, as it does for the Lakoff-Johnson school and Thomas, there is a sense in which we can understand the imagination playing this role at a conscious level to determine the semantic content of theoretical structures through thought experiments. Here, we occasionally use the imagination to settle on what a difficult new theoretical structure means, and in so doing, understand it by relating it to other concepts, increasing its empirical content, or becoming comfortable with it through repeated use. Instead of using the imagination to create a meaningful image of a duck from lines and colours and shapes, we use it purposely to assign new meaning to a theoretical structure via a thought experiment.

For Kant, the imagination is the link between the senses and the understanding. Every time we use a concept, we perform an action, or in Kant's words, create a schema, that links a specific experience to our concept. I think something like this becomes very plausible if instead of linking individual sense experiences to individual categories, we consider linking experience as a whole (or in swaths) to the partially-interpreted theoretical structures of scientific theories via uses of the imagination, whether consciously or unconsciously. In this case, an action is performed, which may sometimes take the form of a thought experiment, which connects theoretical structures to experience. The thought experiment can make these structures, which are often developed in a formal or mathematical way, meaningful and fruitful. No amount of mathematics, laboratory experimentation or computer simulation will establish for us the semantic content of the principle of equivalence, the uncertainty principle, or Newton's laws, because grasping semantic content is something we must do for ourselves, not something that can be done to or for us. The imagination is useful here

because through it we forge new connections between affective, sensorial, memorial and rational elements. All high level theoretical structures will require some act of semantic comprehension on our part if we are to make scientific progress by means of them, whether that act is prompted by a thought experiment, a simpler act of imagination, or an automatic act of imaginative association. And this act of schematization, which would be described by Kant, Johnson, Lakoff and Thomas as an act of imagination, cannot be justified by cognitive science or by philosophy. It is only justified in a transcendental sense because it is always necessarily presupposed by both. That is why it is a *sine qua non* of scientific understanding.

I hope this characterization of the role of the imagination in thought experiments sheds some light on the common conclusions of the empirical studies I considered above, namely, that thought experiments increase scientific understanding by bridging theoretical structures with existing knowledge or experience. Of course, there are different kinds of understanding thought experiments can produce. These are distinguished elsewhere (see Stuart 2017). There are also different kinds of imagination that are important for the discussion. In this paper, I considered imagination as an ability (or faculty, capacity); in future work I will turn to imaginative *processes* (actions or practices), which can be thought of as exercises of our ability to imagine. Unlike the imaginative faculty, imaginative acts can be discussed in a non-transcendental way. That is, we can say directly what makes different imaginative acts epistemically valuable. But at the level of generality I've taken up in this paper by discussing the faculty of imagination, we can only give something like a transcendental justification. In this, I follow Marco Buzzoni, on whose work I have drawn extensively (see Buzzoni 2008, 2013, 2016, 2017). Speaking in Buzzoni's terms, this paper takes up the transcendental perspective on thought experiments, where in future work I will be taking up what he calls the operational perspective.

To conclude, in the cases considered above where novel understanding is produced, it is often due to creating a connection between some theoretical structure(s) of science and existing knowledge, skills or experience, via an exercise of the imagination. We have substantiated this idea by considering the imagination as a key component in building these bridges. Thought experiments are instances of the sort of conceptual exploration that is needed to understand theoretical structures in science, which are themselves a necessary condition for the possibility of a working science. This argument, that thought experiments increase understanding by means of the imagination, which is fundamental to all theoretical understanding, suggests a novel way to justify the role of the imagination in creating scientific understanding, one that does not conflict with any of the existing accounts that aim to justify empirical knowledge produced by thought experiments.

Acknowledgements

I'd like to thank Marco Buzzoni, James R. Brown, Yiftach Fehige, Catherine Elgin, Kenneth Westphal, Hans Radder, Martin Carrier, Nenad Miščević, Marcel Weber, Aaron Wright, Greg Lusk, Cory Lewis, Hans Radder, Tobias Rosefeldt, audiences at the Universities of Toronto, Macerata, Bielefeld, Konstanz, and Dubrovnik. This research was supported by an Ontario Graduate Research scholarship, the Germany/ Europe fund from the University of Toronto.

References

Arriassecq, I., and Greca, M.I. 2012. "A Teaching–Learning Sequence for the Special Relativity Theory at High School Level Historically and Epistemologically Contextualized." *Science and Education* 21: 827–851.

Arstila, V. 2017. "Cognitive penetration, hypnosis and imagination." *Analysis* DOI: https://doi.org/10.1093/analys/anx048.

Beller, M. 2002. *Quantum Dialogue: The Making of a Revolution*. Chicago: University of Chicago Press.

Bokulich, A., and Frappier, M. 2017. "On the identity of thought experiments: Thought experiments rethought." In M. T. Stuart, Y. Fehige and J. R. Brown (eds.). *The Routledge Companion to Thought Experiments*. London: Routledge.

Brown, J. R. 2006. "The Promise and Perils of Thought Experiments." *Interchange* 37: 63–75.

Brown, J. R. 2011. *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London: Routledge.

Brown, J. R., and Fehige, Y. 2011. "Thought Experiments." *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), URL= <http://plato.stanford.edu/archives/fall2011/ entries/thought-experiment/>.

Buzzoni, M. 2008. *Thought Experiment in the Natural Sciences*. Würzburg: Königshausen and Neumann.

Buzzoni, M. 2010. "Empirical thought experiments: A transcendental-operational view." *Epistemologia* XXXIII: 5–26.

Buzzoni, M. 2013. "Thought experiments from a Kantian point of view." In M. Frappier, L. Meynell, and J. R. Brown (eds.). *Thought Experiments in Science, Philosophy, and the Arts*. London: Routledge.

Buzzoni, M. 2016. "Thought experiments in philosophy: A neo-Kantian and experimentalist point of view." *Topoi*. DOI: 10.1007/s11245-016-9436-6.

Buzzoni, M. 2017. "Kantian accounts of thought experiments." In M. T. Stuart, Y. Fehige and J. R. Brown (eds.). *The Routledge Companion to Thought Experiments*, London: Routledge.

Buschman, T. J., and Miller, E. K. 2009. "Top-down versus Bottom-up Control of Attention in the Prefrontal and Posterior Parietal Cortices." *Science* 315: 1860–1862.

Byrne, R. M. J. 2005. The Rational Imagination: How People Create Alternatives to Reality. Cambridge: MIT Press.

Burchard, H. G. W. 2011. "The Role of Conscious Attention in Perception." *Foundations of Science* 16: 67–99.

Camilleri, K. 2007. "Indeterminacy and the Limits of Classical Concepts: The Transformation of Heisenberg's Thought." *Perspectives on Science* 15: 178–201.

Clement, J. 1993. "Using Bridging Analogies and Anchoring Intuitions to Deal with Students' Preconceptions in Physics." *Journal of Research in Science Teaching* 30: 1241–1257.

Clement, J. 2008. *Creative Model Construction in Scientists and Students The Role of Imagery, Analogy, and Mental Simulation*. New York: Springer.

Clement, J. 2009. "Analogy Reasoning via Imagery: The Role of Transformations and Simulations." In B. Kokinov, K. Holyoak, and D. Gentner. *New Frontiers in Analogy Research*. New Bulgarian University Press.

Corcilius, K. 2017. "Aristotle and thought experiments." In M. T. Stuart, Y. Fehige and J. R. Brown (eds.). *The Routledge Companion to Thought Experiments*, London: Routledge.

Currie, G. and Ravenscroft, I. 2002. *Recreative Minds*. Oxford: OUP.

De Mey, T. 2003. "The Dual Nature View of Thought Experiments." *Philosophica* 72: 61–78.

de Regt, H. 2009. "Understanding and Scientific Explanation." In H. de Regt, S. Leonelli and K. Eigner (eds.). *Scientific Understanding: Philosophical Perspectives*. Pittsburgh: University of Pittsburgh Press.

de Regt, H., Leonelli, S., and Eigner, K. 2009. "Focusing on Scientific Understanding." In H. de Regt, S. Leonelli and K. Eigner (eds.). *Scientific Understanding: Philosophical Perspectives*. Pittsburgh: University of Pittsburgh Press.

Duit, R. And Tesch, M. 2010. "On the Role of Experiment in Science Teaching and Learning – Visions and the Reality of Instructional Practice." In M. Kaloglannakis, D. Stavrou, and P. G. Michaelides P.G. (eds.). *Proceedings of the 7th international Conferences Hands-on-Science. Bridging the Science and Socity Gap*. University of Crete: 17–30.

Einstein, A. 1921/1954. "Geometrie und Erfahrung." Berlin: Julius Springer. English translation: "Geometry and Experience." In Albert Einstein *Ideas and Opinions*. New York: Bonanza Books.

Einstein, A. 1931. *Einstein on Cosmic Religion: With Other Opinions and Aphorisms*. Mineola, NY: Dover Publications.

Einstein, A. 1934. "On the Method of Theoretical Physics." *Philosophy of Science* 1: 163–169.

Gendler, T. S. 2013. "Imagination." *The Stanford Encyclopedia of Philosophy*. E, N. Zalta (ed.). URL = <http://plato.stanford.edu/archives/fall2013/entries/imagination/>.

Gendler, T. S. and Hawthorne, J. (eds.). 2002. *Conceivability and Possibility*. Oxford: Oxford University Press.

Gilbert, J. and Reiner, M. 2000. "Thought Experiments in Science Education: Potential and Current Realization." *International Journal of Science Education* 22: 265–283.

Grimm, Stephen R. 2009. "Reliability and the Sense of Understanding." In H. de Regt, S. Leonelli and K. Eigner (eds.). *Scientific Understanding: Philosophical Perspectives*. Pittsburgh: University of Pittsburgh Press.

Hadamard, J. 1996. *The Mathematician's Mind: The Psychology of Invention in the Mathematical Field*. Princeton: Princeton University Press.

Häggqvist, S. 1996. *Thought Experiments in Philosophy,* Stockholm: Almqvist and Wiksell International.

Holton, G. 1996. "On the Art of the Scientific Imagination." *Daedalus* 125: 183–208.

Jacob, F. 2001. "Imagination in Art and Science." *Kenyon Review* 23: 113–121.

Johnson, M. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination and Reason*. Chicago: Chicago University Press.

Kind, A. 2016. Routledge Handbook of the Philosophy of Imagination. London: Routledge.

Kind, A. and Kung, P. (eds.). 2016. Knowledge through Imagination. Oxford: OUP.

Klassen, S. 2006. "The Science Thought Experiment: How Might it be Used Profitably in the Classroom?" *Interchange* 37: 77–96.

Kosem, S.D., and Özdemir, Ö. F. 2014. "The Nature and Function of Thought Experiments in Solving Conceptual Problems." *Science and Education* 23: 865–895.

Kosso, P. 2006. "Scientific Understanding." *Foundations of Science* 12: 173–188.

Koyré, A. 1968. *Metaphysics and Measurement*. Cambridge: Harvard University Press.

Kvanvig, J. 2003. The Value of Knowledge and the Pursuit of Understanding. Cambridge: Cambridge University Press.

Lakoff, G., and Johnson, M. 1999. *Philosophy in the Flesh: the Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Lattery, M. 2001. "Creative Model Construction in Scientists and Students The Role of Imagery, Analogy, and Mental Simulation." *Science and Education* 10: 485–492.

Lipton, Peter. 2009. "Understanding without Explanation." In H. de Regt, S. Leonelli and K. Eigner (eds.). *Scientific Understanding: Philosophical Perspectives*. Pittsburgh: University of Pittsburgh Press.

McGinn, C. 2004. *Mindsight: Image, Dream, Meaning*. Cambridge: Harvard University Press.

Miller, E. K., and Cohen, J. D. 2001. "An Integrative Theory of Prefrontal Cortex Function." *Annual Review of Neuroscience* 24: 167–202.

Miščević, N. 2007. "Modelling Intuitions and Thought Experiments." *Croatian Journal of Philosophy* 7: 181–214.

Nersessian, N. 2007. "Thought Experiments as Mental Modelling: Empiricism without Logic." *Croatian Journal of Philosophy* 7: 125–161.

Orwell, G. 1940. "Review of *Mein Kampf* by Adolf Hitler." Reprinted in S. Orwell and I. Angus (eds.). 1968. *The Collected Essays, Journalism and Letters of George Orwell, Volume 2*. Boston: David R. Godine Publishers.

Özdemir, Ö. F. 2009. "Avoidance from Thought Experiments: Fear of Misconception. *International Journal of Science Education* 31: 1–20.

Peacock, K. A. 2017. "Happiest Thoughts: Great Thought Experiments of Modern Physics." In Stuart et al. (eds.). The Routledge Companion to Thought Experiments. London: Routledge.

Prinz, J. 2002. *Furnishing the Mind: Concepts and their Perceptual Basis*. Cambridge: MIT Press.

Radder, H. 1996. *In and About the World: Philosophical Studies of Science and Technology,* Albany: State University of New York Press.

Reiner, M. and Burko, L. 2003. "On the Limitations of Thought Experiments in Physics and the Consequences for Physics Education." *Science and Education* 13: 365–385.

Reiner, M. and Gilbert, J. 2004. "The Symbiotic Roles of Empirical Experimentation and Thought Experimentation in the Learning of Physics." *International Journal of Science Education* 26: 1819–1834.

Reichenbach, H. 1965. *The Theory of Relativity and A Priori Knowledge*. Berkeley: University of California Press.

Roychoudhuri, C. 1978. "Heisenberg's Microscope—A Misleading Illustration." *Foundations of Physics* 6: 845–849.

Salis, F. and Frigg, R. (forthcoming) "Capturing the scientific imagination." In P. Godfrey-Smith and A. Levy (eds.). The Scientific Imagination. Oxford: OUP.

Stephens, L. A., and Clement, J. J. 2006. "Designing Classroom Thought Experiments: What we Can Learn from Imagery Indicators and Expert Protocols." *Proceedings of the 2006 Annual Meeting of the National Association for Research in Science Teaching,* San Francisco.

Stephens, L. A., and Clement, J. J. 2012. "The Role of Thought Experiments in Science and Science Learning." In B. Fraser, K. Tobin and C. McRobbie (eds.). *Second International Handbook of Science Education*. Dordrecht: Springer: 157–175.

Stokes, D. 2014. "The Role of Imagination in Creativity." In E. S. Paul and S. B. Kauffman (eds.). *The Philosophy of Creativity: New Essays*. Oxford: Oxford University Press.

Stuart, M. T. 2016. "Taming Theory with Thought Experiments: Understanding and scientific progress." *Studies in the History and Philosophy of Science* 58: 24–33.

Stuart, M. T. 2017. "How Thought Experiments Produce Understanding." In Stuart et al. (eds.). *The Routledge Companion to Thought Experiments*. London: Routledge.

Stuart, M. T., Fehige, Y. and Brown, J. R. 2017. *The Routledge Companion to Thought Experiments*. London: Routledge.

Thagard, P. 2010. *The Brain and the Meaning of Life*. Princeton: Princeton University Press.

Thagard, P. and Stewart, T. 2011. "The AHA! Experience: Creativity Through Emergent Binding in Neural Networks." *Cognitive Science* 35: 1–33.

Thomas, N. 1999. "Are Theories of Imagery Theories of Imagination? An Active Perception Approach to Conscious Mental Content." *Cognitive Science* 23: 207–245.

Toulmin, S. 1972. *Human Understanding*. Princeton: Princeton University Press.

Trafton, J. G., Trickett, S. B., and Mintz, F. E. 2005. "Connecting Internal and External Representations: Spatial Transformations of Scientific Visualizations." *Foundations of Science* 10: 89–106.

Trickett, S. B., and Trafton, J. G. 2007. ""What if...": The Use of Conceptual Simulations in Scientific Reasoning." *Cognitive Science* 31: 843–875.

Van Dyck, Maarten. 2003. "The Roles of One Thought Experiment in Interpreting Quantum Mechanics: Werner Heisenberg Meets Thomas Kuhn." *Philosophica* 72: 79–103.

van Fraassen, B. 2008. *Scientifc Representation: Paradox of Perspective*. Oxford: Oxford University Press.

van Woudenberg, R. 2006. "Introduction." *Metaphilosophy* 37: 151–161.

Velentzas, A. and Halkia, K. 2011. "The 'Heisenberg's Microscope' as an Example of Using Thought Experiments in Teaching Physics Theories to Students of the Upper Secondary School." *Research in Science Education* 41: 525–539.

Velentzas, A. and Halkia, K. 2013a. "The Use of Thought Experiments in Teaching Physics to Upper Secondary-Level Students: Two Examples from the Theory of Relativity." *International Journal of Science Education* 35: 3026–3049.

Velentzas, A. and Halkia, K. 2013b. "From Earth to Heaven: Using 'Newton's Cannon' Thought Experiment for Teaching Satellite Physics." *Science and Education* 22: 2621–2640.

Velentzas, A., Halkia, K. and Skordoulis, C. 2007. "Thought Experiments in the Theory of Relativity and in Quantum Mechanics: Their Presence in Textbooks and in Popular Science Books." *Science and Education* 16: 353–370.

Weinberg, S. 1992. *Dreams of a Final Theory*. New York: Random House.

Ylikoski, Petri. 2009. "The Illusion of Depth of Understanding in Science." In H. de Regt, S. Leonelli and K. Eigner (eds.). *Scientific Understanding: Philosophical Perspectives*. Pittsburgh: University of Pittsburgh Press.

Yourgrau, Wolfgang. 1967. "On Models and Thought Experiments in Quantum Theory." *Monatsberichte der Deutschen Akademie der Wissenschaften zu Berlin* 9: 886–874.

Croatian Journal of Philosophy
Vol. XVII, No. 49, 2017

Bayesianism and the Idea of Scientific Rationality

JEREMIAH JOVEN JOAQUIN
De La Salle University, Manila, Philippines

Bayesianism has been dubbed as the most adequate and successful theory of scientific rationality. Its success mainly lies in its ability to combine two mutually exclusive elements involved in the process of theory-selection in science, viz.: the subjective and objective elements. My aim in this paper is to explain and evaluate Bayesianism's account of scientific rationality by contrasting it with two other accounts.

Keywords: Bayesianism, historiographical theory of science, scientific rationality, rational reconstruction program.

1. *The Problem of Scientific Rationality*

The problem of scientific rationality is one the most important problems in philosophy of science. Throughout its long and colorful history, the problem has seen many formulations. However, there seems to be an essential theme that remains the same in all those varying formulations. This can be formulated as follows: "Does the choice of a particular scientific theory over another involve rationality?" Notice that the concept of rationality figures prominently here. It is, thus, important to show what it means.

When we talk about rationality in the problem of scientific rationality we are talking about the rationality involved in choosing one theory over another; i.e. we are talking about the conditions that constitute the reasonableness of such a choice. Let me elaborate on this further. Suppose that we have two rival theories, X and Y, trying to describe the *same* phenomenon. X and Y are not reducible to one another, since the set of statements, which makes up X, could not be subsumed to Y, and vice-versa. Suppose further that the scientific community chooses X over Y. The issue here is whether those scientists really have *good* reasons to choose X over Y, and if they have, what conditions then would constitute the choice's reasonableness.

On the one hand, some philosophers hold that a choice's reasonableness is simply determined by a strict methodological process. They claim that there are procedures and criteria in determining whether a theory is better than another. Others claim, on the other hand, that the reasonability of a choice is more complex than that. They claim that scientific rationality can only be explained by looking at arbitrary elements present in the processes involved in scientific enterprise as a whole. The issue about scientific rationality, therefore, is concerned with explaining the conditions that constitute the rationality of the theory-selection process in science. Before discussing this further, there is a much pressing matter that I need to address first.

Some philosophers have objected to the idea of characterizing the problem of scientific rationality only in terms of the rationality of the theory-selection process. They claim that this idea is founded on a faulty assumption. They contend that since the process of theory-selection is only one of the activities done in the sciences, it would not follow that if this were irrational, the whole scientific enterprise would then be irrational. For them, the whole debate about scientific rationality falsely assumes that the rationality of science as a whole is seen only in the theory-selection process.[1]

Like many other philosophers dealing with scientific rationality, I do not deny that the theory-selection process is just another activity done in the sciences. I need to emphasize, however, that the epitome of the scientific enterprise is seen in this process. The rationality of the whole enterprise is best seen in the manner by which the scientific community decides what theories to accept or reject. If their choice were made unreasonably, it puts into question the entire scientific enterprise. On the other hand, if it was proven otherwise, then it reassures us of the confidence that we give to science. The reason why the problem of scientific rationality, as is characterized here, focuses on the debate concerning the rationality of the theory-selection process is not only because it is the epitome of the scientific enterprise, but also because it assures us of the confidence that we give to science as a whole.

2. *Two Alternative Solutions*

There are two very influential solutions to the problem of scientific rationality. There are those who claim that the choice of a scientific theory is determined by strictly following a method. For others, such a choice is ultimately determined by reasons external to science itself—be it personal, social, or political. Adherents of the former solution are influenced by logical empiricism's rational reconstruction program; adherents of the latter are influenced by Kuhn's historiographical theory of science.

[1] For example, Siegel (1985) has argued that the issue concerning the rationality of the process of theory-selection presupposes an answer to the question, "In what constitutes rationality in science?" He claims that this question is prior to the question formulated in this paper. I shall argue against this claim.

For a long time, the rational reconstruction program has been the standard conception of scientific rationality. Adherents of this program not only include the logical empiricists, like Schlick and Carnap, but also the Popperians—supporters of Popper—and the later neo-pragmatists, like Quine and van Fraasen. By considering the following theses, we could have an idea of the rational reconstructionist's solution to the problem of scientific rationality:[2]

(1) the thesis of the unified method of science
(2) the thesis of formalizability of this method
(3) the demarcation thesis; and,
(4) the thesis of scientific rationality.

The first thesis tells us that by looking at the history of science, one would find some semblance of a unified scientific method. The second states that it is possible to formalize or systematize this method, and it is the philosopher's task to do so. Through the second thesis, the third thesis states that this formalized scientific method differentiates science from the non-sciences and the pseudo-sciences. Still via the second thesis, the fourth thesis tells us that such a method could show how science really works. That is, how scientific theories are made, how they are accepted, and whatnot.

From the four theses, we could already have an idea how the rational reconstructionist would answer the problem of scientific rationality. The solution is roughly this. Given two opposing theories, X and Y, scientists *would* choose X over Y if and only if (iff) using the formalized scientific method, X is shown to be better than Y. The idea is that this formalized scientific method would give adequate reasons to prefer one theory over the other. The process of theory-selection, therefore, would only be a matter of following the rules set by this method. But what is this method?

There are two competing "methods" available for the rational reconstructionists: the method of confirmation and the method of falsification.[3] The method of confirmation works as follows. Scientific inquiry usually starts with a theory. If predictions or descriptions made using this theory were shown to be true by some (either observational or experimental) evidence, then such a theory would thus be confirmed, or at least shown to be empirically adequate. In light of the problem of scientific rationality, this method works as follows. Given two opposing theories, X and Y, if the gathered evidence shows that X's descriptions are true, and shows Y's to be false, it would then warrant the choice of X over Y. Because of this simple formula for theory-selection, many

[2] I am following the discussion of these four theses in (Jiang 1985).

[3] It should be noted, however, that there is a deep tension between these "methods" of science. Proponents of the method of confirmation, like Hempel, claim that this method is a more powerful method than the method of falsification. Proponents of the other camp, like Popper, make the same claim in favor of their preferred "method".

rational reconstructionists were led to believe that the method of confirmation is the best method for science. Others, like Popper, were not quite impressed by this.

Popper has showed that if the method of confirmation were the real method of science, then theories like astrology and alchemy would have to be accepted as scientific theories, since this method could easily be applied to them. Of course, rational reconstructionists would repudiate this idea because, for them, these "sciences" are *not* really scientific. Since the method of confirmation would consider such theories as scientific, Popper claims that it is the wrong method of science. What he proposes as an alternative is the method of falsification.

The method of falsification assumes that theories can never really be confirmed; rather, they can only be *temporarily* corroborated by certain evidence. Like the method of confirmation, the method of falsification sees that scientific inquiry begins with a theory. However, unlike the former, the latter obliges scientists to look for evidence that could show that their theory is false—since the true mark of a scientific theory is its falsifiability (possibility to be false). If the lot of evidence were to show that the theory is false, then it would have to be rejected. If otherwise, then it is said to be corroborated by such evidence. Nothing is final here. Some accepted theory might eventually be rejected—due perhaps to some new evidence against it. But this should not cause dismay, for this process is the mark of "scientific progress." In light of the problem of scientific rationality, the method of falsification works as follows. Given two opposing theories, X and Y, X is chosen over Y iff X and Y are falsifiable and X is corroborated by certain evidence, while Y is not. If X is later shown to be false by some new evidence, and another theory, Z, which is falsifiable but is now corroborated by that new evidence, Z should be chosen over X.

For rational reconstructionists, therefore, scientific rationality is determined solely by the method of science. On the basis of evidence, theories are accepted or rejected. The manner by which theories are accepted or rejected depends either on how evidence corroborates or confirms them. There are three important elements in this account of scientific rationality. First, theories should be *about* something empirically testable. Second, evidence that confirms or corroborates *should* be external to the theory. Third, confirmation (or corroboration) determines the acceptance or rejection of a theory. Only by following the method of science could we show how scientific rationality is possible. Friends of the rational reconstruction program have thus shown that there can only be an *objective* way of answering the problem of scientific rationality.

Kuhn, a leading proponent of the historiographical theory of science, has raised crucial objections against the rational reconstruction program's solution to the problem of scientific rationality. First, he sees that the picture of the history of science proposed by the rational re-

constructionist is normative rather than descriptive. He argues that if we were to look at the actual history of science, we would not see a unified method that governs scientific growth; what we would see, rather, are "non-cumulative developmental episodes in which an older paradigm is replaced…by an incompatible new one" (Kuhn 1996, 92). Since the first thesis espoused by rational reconstructionists tells us that a unified method of science can be seen in the history of science, and if Kuhn's observations are correct, then this rational reconstructionist thesis would be false. What then is the intellectual force of such a thesis? For Kuhn, since this thesis is false, it would mean that the rational reconstructionist's insistence for a method of science would merely be an imposition of a dogma. If this were the case, it would then follow that their clamor for a method of science would be circular, thus making the "method" of science questionable.

Second, Kuhn points out that the rational reconstructionist's depiction of scientific rationality is limited. That is, their thesis of scientific rationality does not provide a complete description of the scientific process. For rational reconstructionists, the process of choosing a *theory* over another would simply be a matter of strictly following the method of science. However, Kuhn points out that this view only applies to a specific period in the history of science, which he calls "normal science", and not to whole history of science. Kuhn defines "normal science" as "research firmly based upon one or more past scientific achievements, achievements that some particular scientific community acknowledges for a time as supplying the foundation for its further practice" (1996: 10). Since normal science is based on the scientific community's acknowledgment of these past achievements, a particular way of doing science is thus born. This way of doing science is what Kuhn calls a "paradigm".

For Kuhn, a paradigm functions like the rational reconstructionist's view of the method of science. It determines what evidence would be acceptable in confirming a theory, or what research topic should be undertaken in perfecting a theory. This determination, however, is only made within this dominant paradigm. Kuhn further points out that in the actual history of science there were episodes where this paradigm breaks down due to some anomalies that could not be accounted by the dominant paradigm. The break down of a paradigm is what he calls, "crisis science". In crisis science, the scientific community suffers a terrible fate because the dominant paradigm is put into question. Without this paradigm, normal science would cease its activities. Kuhn argues that although some scientists would try to save the old paradigm, in the time of crisis many would offer new paradigms to account for the anomalies that the old paradigm could not. In this period, the whole scientific enterprise would have many different paradigms. But crisis science would eventually end. Its end is marked by the "emergence of a new candidate for paradigm and with the ensuing battle over its acceptance." (Kuhn 1996: 84). The problem, then, is to determine the con-

ditions and processes involved in choosing one paradigm over another.

To paraphrase Kuhn, it is impossible to use the rational reconstructionist's idea of the method of science as a standard of rationality of choosing one paradigm over another "for these (methods) depend in part upon a particular paradigm, and that paradigm is at issue." (1996: 94). Kuhn further claims that to resort to a "method" in choosing between paradigms is circular since "[e]ach group uses its own paradigm to argue in that paradigm's defense" (1996: 94). Thus, Kuhn shows that the rational reconstructionist's main theses are problematic. And since they are problematic, their solution to the problem of scientific rationality would be problematic as well.

Kuhn's alternative account of scientific rationality is somewhat controversial. He sees scientific rationality not as a matter of simple rule-governed processes, but a more complex one.[4] For Kuhn, science is a *human* endeavor. As such, there are elements in it that color the way science is conducted. Since science is a human endeavor, it follows that scientific rationality is also marked by these *humanistic* elements. For Kuhn, "[a]n apparently arbitrary element, compounded of personal and historical accident, is always a formative ingredient of the beliefs espoused by a given scientific community in a given time" (1996: 4). The combination of this arbitrary element and the personal niceties of scientists make the problem of scientific rationality a human issue. In this picture, scientific rationality is an ongoing process that starts from the formative years of the members of a scientific community, up to their activities in specific fields, then to their decision to accept or reject theories, and then to the process of relearning or unlearning old ways of thinking. Furthermore, this process informs the way that a scientist chooses anything. A scientist's background would influence his preferred area of research. A group of scientists' shared commitments would determine their choice of accepting the results of an experiment.

In general, for Kuhnian historiographers of science, scientific rationality—and the rationality of choosing one theory over another—depends on arbitrary elements external to the logic and method of science, or even to the facts observed. Thus, these apparent arbitrary elements also determine the theory-selection process. Such a process is founded upon certain value-laden reasons and commitments shared by the members of a scientific community. These reasons are not derived from any method of science, but are more political or social in nature. To put it roughly, the Kuhnian idea of scientific rationality with regard to theory-selection is this. Given two opposing theories, X and Y, X is chosen over Y iff a *consensus* to choose X over Y is reached by the members of the scientific community.

This does not mean that a theory is selected by mere majority vote

[4] In what will come next, I have refrained from articulating certain Kuhnian themes, like revolutionary science, changes of worldview, and incommensurability of theories as these are not deemed necessary to articulate the historiographical theory's main thesis.

or by a shared whim. Rather, members of the scientific community *eventually* arrive at a choice because of the values and commitments they share, like valuing the consistency and plausibility of the theory, or the commitment to scientific development, etc. It is not the case, however, that the sharing of values and commitments means that the assignment of importance of values or commitments is the *same* for each member of the community. This is not possible because each individual, informed by their personal backgrounds, would assign levels of value differently. Only by having these different subjective values meet could a consensus be produced. Kuhn's historiographical view of science gives much importance to these *subjective* values because these have "an important effect on scientific development" (ibid). But this emphasis on subjective values is not without problems.

Many philosophers have argued that Kuhn's emphasis on subjective values makes the whole theory-selection process a highly subjective affair. Kuhn does not deny this; in fact he embraces it. Subjectivity drives science to progress. Without it, science will be impossible. On the other hand, others have argued that if Kuhn's view is correct, then it would show that whole scientific enterprise would be irrational. Kuhn counters that this objection is only tenable if rationality *means* strictly following a rule or method; but as we have seen, he denies that there is such a method.

One very important objection against Kuhn's historiographical view is the fact that, contrary to Kuhn's point, the process of theory-selection involves *evidence*. Kuhn's account focuses too much attention on the historical aspects of science that the question of evidence has been overlooked. Why is it that although there are subjective elements that strongly influence a scientist's acceptance of a theory, the very same scientist would, more often than not, accept a theory on the basis of compelling evidence for it, even if such evidence is contrary to his personal beliefs? This is a feature of scientific rationality that Kuhn's view fails to give a judicious account.

3. *Bayesianism*

The main project of the Bayesian approach to scientific rationality is to combine the rational reconstructionist's insistence for an *objective* method of determining a choice's reasonableness with Kuhn's emphasis on the importance of *subjective* arbitrary elements that influence the members of the scientific community. There are three important elements here that need to be considered:

(1) a subjective interpretation of probability statements[5];
(2) the Dutch book argument; and,
(3) Bayesian thesis of rationality.[6]

With these three elements, Bayesianism does not only give an adequate theory of scientific rationality, but also restored the importance of evidence and confirmation in the theory-selection process.

Bayesianism begins with a subjective interpretation of probability. On this view, probability statements are statements about personal degrees of beliefs.[7] These degrees of beliefs are quantifiable according to a 0 to 1 scale, 0 being the lowest value and 1 being the highest. The assignment of these values is a highly personal, thus subjective, affair. A person can freely assign a value of .70 to his belief that he will win the lottery, regardless of whether he has strong grounds for it. The only restriction that Bayesianism imposes on the assignment of values is the coherence of this assignment with other beliefs. Since the assignment of values is too subjective, then there would be a problem of determining coherence, since we can have a coherent set of irrational beliefs. To answer this problem, Bayesianism has the Dutch book argument.

The Dutch book argument is a pragmatic test for the coherence of degrees of beliefs. In its simplest formulation, it states that if a person should be willing to act in accordance to his beliefs. However, if the result of his action would make him suffer more losses than receive more gains, then his beliefs are incoherent, and he is acting irrationally; otherwise they are coherent, and he is acting rationally. For Bayesians, the coherence of beliefs is a matter of a betting game.

[5] There are three dominant interpretations of probability statements: *a priori* (classical) interpretation, relative frequency interpretation, and the subjectivist interpretation. The classical interpretation, developed by the "fathers" of probability theory, Fermat and Pascal, tells us that probability statements are statements about the chances of some favorable outcome happening over the total number of *possible* outcomes. Thus, the statement, "There's a 25% chance that I'll get a clubs from a standard deck of cards" means that of the fifty-two cards, there are thirteen chances of having a favorable outcome. On the other hand, the relative frequency theory, developed by Keynes, claims that probability statements are statements about the number of instances that a favorable outcome happens over an observed period of time. Thus, the statement, "There's a 30% chance that I'll get six in a single roll of a loaded die" means that out ten times that I rolled that die, three turned up six. The subjectivist interpretation, developed by Ramsey, sees probability statements as statements about a person's partial beliefs. Thus, the statement, "There's a 20% chance that I'll get the job" means that the person who uttered the statement sees that there's a low chance for him to get the job. For further discussions on the interpretations of probability, see (Hajek 2012).

[6] Bayesianism is considered as a general theory of rationality, see (Joyce 2004). But although this is the case, it does not prohibit extending its use to account for scientific rationality.

[7] Ramsey is acknowledged as the first to discuss the philosophical underpinnings of a subjective interpretation of probability statements, see the collection of his works in (Ramsey 1996).

Suppose that person, A, assigns .51 to belief, B, and assigns the same value to a contrary belief, not-B. Suppose further that someone offered him a wager to the effect that A bets $6 on B and another $6 on not-B. If B obtains, A will win $10; if not-B obtains, he'll also receive $10. Suppose that either B or not-B will obtain, but not both. If A decides to bet on both B and not-B, then we will know that he has an incoherent set of beliefs, since he is willing to lose $12 only to gain $10.

The Dutch book argument shows that the coherence of a set of beliefs, hence also the rationality of the person having those beliefs, can be determined if that person is not willing to lose more than he could gain. If the person decides to act according to his incoherent beliefs, then he is acting irrationally. Notice here, that the argument works on two assumptions. First, rationality of choices involves coherence of beliefs, which in turn presupposes the notion of utility expectations; second, there are external elements that determine the coherence of beliefs. These two aspects are very important in Bayesianism's account of scientific rationality.

External elements, like observational or experimental evidence, are important in determining rationality of choices. Although, Bayesians are willing to grant that the assignment of values is subjective, they also believe that it is important to look at external objective factors that determine a choice's rationality. Objectivity is founded on a formalized notion of *confirmation*. It is formulated as follows. A certain evidence, E, *confirms* a person's subjective assignment of degrees of belief, P(B), just in case E raises P(B). That is, P(B/E) > P(B). Otherwise it is disconfirmed. Confirmation happens on the level of the subject involved. Via a subjectivist interpretation of probability, a person assigns a value to his belief. If some evidence confirms this belief, then this evidence raises his confidence to his belief. There is an implicit appeal to conditional probabilities here. That is, if E confirms P(B), then E *raises* P(B).

The formalized idea of confirmation has the Bayesian theory of rationality as a necessary consequence. This is formulated as follows: P(T/E) = P(E/T) x P(T)/P(E).[8] What this formula means is simply that a theory is more confirmed by unexpected evidence than expected ones.

[8] Where P(T/E) means that the degree of belief to a theory given the evidence; P(E/T) expresses *a measure* that that the evidence is unsuprising given the theory; P(T) means the degree of belief to a theory prior the evidence; and P(E) means the prior probability of evidence. Because of limited space, I could not unpack the niceties of this formula. However, I'll try to discuss the two principles involved here: (1) the prediction (expectation) principle; and (2) the surprise principle. The prediction principle states that if a person assigns a high value to the belief that some evidence, E, would occur because of a theory, T, then E *strongly* confirms T if E thus arise. The surprise principle, on the other hand, states that if a person is expecting two evidences: E and E* from T, if E is more surprising than E*, but would not be surprising if T were true, then E *strongly* confirms T than E* does. These two principles show that unexpected evidence that a theory predicts strongly confirms that theory than expected evidence could. For further details of the formulation, see (Joyce 2004).

Thus, if some evidence strongly confirms a theory, I *should* then assign a higher value to my theory given the evidence.[9] The importance of this result can be seen more clearly if we apply it in relation to the problem of scientific rationality.

The Bayesian account of scientific rationality—especially of the rationality of choice—amounts to the following. If two theories, X and Y, predict that some event, E, is *expected* to happen, and E does happen, then X and Y are *confirmed* by E. Of course confirmation here still relates to the raising of subjective degrees of beliefs. But if X predicts a further *unexpected* event F, which Y did not predict, and F does happen, given this *unexpected* evidence, one *should* raise the degree of belief to X than Y given F. As the Bayesian theory of rationality suggests, since some evidence raises our confidence to X than Y, then it should follow that we need to assign a higher value to X than Y. That is, X would be a reasonable choice than Y. Furthermore, given the Dutch book argument, if a person chooses Y over X given F, that person then is acting irrationally.

Bayesianism accounts for scientific rationality by considering two mutually exclusive elements in the theory-selection process: the subjective assignment of values to one's beliefs, and the objective confirming evidence of a theory. Bayesianism suggests that in choosing between two or more theories, it is always reasonable to choose the one which is *confirmed* by evidence. To choose otherwise is to succumb to the Dutch book argument.

4. *Conclusion*

I have discussed some of the intricacies of the philosophical debate about scientific rationality. I have shown that problem of scientific rationality is concerned with explaining the constitution of the rationality of choice in the sciences. Many philosophers have offered their solutions to it by maintaining either an extreme version of objectivism or subjectivism. The rational reconstructionists have espoused the former solution; while Kuhnians the latter. The rational reconstructionist's solution succumbs to Kuhn's historical critique. Kuhn's view, however, failed to recognize the importance of evidence in the theory-selection process. I have argued that Bayesianism offers a middle ground that reconciles both extreme positions. Armed with the subjective interpretation of probability, which highlights personal (subjective) assignment of values to beliefs, and the Dutch book argument, which is an objective test of the coherence of these assignments, Bayesians approached the problem of scientific rationality with a renewed interest on how evidence confirms a theory. As such, Bayesianism showed that although our beliefs are really subjective, we still have to choose the best theory

[9] Bayesianism is also characterized as a *normative* theory of rationality, see (Joyce 2004).

among other competing theories. And in having the notion of a "best" choice, we are already implying that we can have rational grounds for choosing one over the other. However, the rationality of this choice is not determined solely by a strict application of method or by mere personal arbitrary elements that surround our choices. The rationality of our choice of a theory is founded on evidence confirming that theory.

My main aim in this paper is to show that Bayesianism is indeed an adequate theory of scientific rationality. What I have discussed here are brief descriptions of the rational reconstruction program, Kuhn's historiographical view, and Bayesianism. Comparing the three, I have shown that Bayesianism reconciled the best aspects of the two other theories. As such, I can say that the Bayesian approach is indeed an adequate account of scientific rationality.

Acknowledgment

A version of this paper was delivered at the Osaka University-De La Salle University Joint Academic Research Workshops held at De La Salle University, Manila, Philippines, in September 2009. I wish to thank the participants and organizers of that conference for their positive feedback. I also would like to thank Robert James Boyles, Mark Anthony Dacela, Adrianne John Galang, Brian Garrett, James Franklin, Alan Hajek, Napoleon Mabaquiao, Raj Mansukhani, and this journal's anonymous referee for useful comments and suggestions. Finally, this paper draws some inspiration from Professor Jeffrey Kasser's The Great Courses Lecture Series on the Philosophy of Science.

References

Hájek, A. 2012. "Interpretations of Probability." *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/probability-interpret/>.

Jiang, T. 1985. "Scientific Rationality, Formal or Informal?" *The British Journal for the Philosophy of Science* 36 (4): 409–423.

Joyce, J. 2004. "Bayesianism." In A. R. Mele and P. Rawling (eds.). *The Oxford Handbook of Rationality*. Oxford: Oxford University Press.

Kuhn, T. 1996. *The Structure of Scientific Revolutions*, 3rd ed., Chicago: University of Chicago Press.

Ramsey, F. 1996. "Truth and Probability." In R. B. Braithwaite (ed.). *The Foundations of Mathematics and other Logical Essays.* London: Kegan, Paul, Trench, Trubner & Co.

Siegel, H. 1985. "What Is the Question concerning the Rationality of Science?" *Philosophy of Science* 52 (4): 517–537.

Croatian Journal of Philosophy
Vol. XVII, No. 49, 2017

The Grounding Problem for Panpsychism and the Identity Theory of Powers

NINO KADIĆ
Centraleuropean University, Budapest, Hungary

In this paper, I address the grounding problem for contemporary Russellian panpsychism, or the question of how consciousness as an intrinsic nature is connected to dispositions or powers of objects. I claim that Russellian panpsychists cannot offer an adequate solution to the grounding problem and that they should reject the claim that consciousness, as an intrinsic nature, grounds the powers of objects. Instead, I argue that they should favour the identity theory of powers, where categorical and dispositional properties are identified. I maintain that the identity theory serves as a better ontological basis for panpsychism since it avoids the grounding problem. Apart from that, I also argue that identity theory panpsychism is a position more parsimonious than Russellian panpsychism since it introduces fewer entities while successfully avoiding the grounding problem. Based on these considerations, I conclude that identity theory panpsychism is an option worth considering.

Keywords: Panpsychism, grounding problem, categorical properties, dispositional properties.

1. *Introduction*

Panpsychism is the view that consciousness is a fundamental and universal feature of reality. Though this view might seem odd and counterintuitive at first, there are many reasons for why we should take it seriously. These reasons largely follow the *ex nihilo nihil fit* principle—consciousness must come from *somewhere*, it cannot come into existence from nothing, such as when unconscious fundamental particles arrange themselves to form conscious brains. The solution panpsychism offers—maintaining that those particles are themselves conscious—raises a few eyebrows. Despite its initial strangeness, the view has been gaining traction in current philosophy of mind.

However, panpsychism faces a number of difficulties. The most discussed of those is the combination problem, the question of how small subjects come together to form big subjects. While this is and probably will remain a serious issue, I will instead focus on a less discussed but equally challenging problem for panpsychism: how is consciousness, as something present within all objects at the fundamental level of reality, connected to the dispositions or powers those objects exhibit?

In this paper, I argue that contemporary panpsychism, influenced by the ideas of Bertrand Russell, does not offer an adequate answer to this question. The reason for this is the ontological commitment of Russellian panpsychism to the idea that an object's intrinsic nature grounds or accounts for its dispositions or powers. It is hard to see how this grounding relation can be explained without invoking further problematic notions. This is the grounding problem for panpsychism. I then offer a way of avoiding this issue by arguing that the panpsychist is better off accepting the identity of intrinsic natures and powers rather than maintaining that they are ontologically different. Furthermore, through a discussion on intentionality and the directedness of powers, I demonstrate that panpsychism paired with the identity theory results in a more unified account of objects and properties than the identity theory alone. Finally, after addressing several objections to the identity view, I conclude that the panpsychist who accepts it is better equipped to handle the grounding problem than the Russellian panpsychist.

2. *The Russellian Motivation*

One of the main contemporary motivations for taking panpsychism seriously comes from Bertrand Russell, who argued (1927/1992) that observational science reveals only the mathematical structure of matter, without saying anything about what matter is intrinsically. Russell's approach has recently attracted renewed interest, forming a body of works which fall under the name of *Russellian monism*. Philosophers following Russell's line of reasoning think that consciousness is the best candidate to play the role of the intrinsic nature of matter, as it is the only such nature we know of. For example, William Seager has stated that consciousness is something we have "ready to hand" to play that role, and that nothing justifies positing additional intrinsic properties except the verbal demand that it be "non-mental" (Saeger 2006: 137).

In addition to Russell, this argument has historic roots in the work of Arthur Eddington who argued (1928: 259) that science cannot reveal the nature of the atom since it only describes it in terms of pointer readings on instrument dials. However, in the case of pointer readings regarding his own brain, it is clear to him that the readings are attached to a background of consciousness. If that is true, Eddington suggests that "the background of other pointer readings in physics is of a nature continuous with that revealed to me in this particular case"

(1928: 259). In other words, he argues that all physical facts should be attached to a background of consciousness:

> If we must embed our schedule of indicator readings in some kind of background, at least let us accept the only hint we have received as to the significance of the background—namely, that it has a nature capable of manifesting itself as mental activity (Eddington 1928: 260).

Eddington's argument appeals to parsimony, as it aims to show that it is more reasonable to presume that the background of pointer readings is consciousness rather than something inherently non-conscious. As he puts it, attaching pointer readings to "something of a so-called 'concrete' nature inconsistent with thought" would be "silly" if we are left wondering "where the thought comes from" (Eddington 1928: 259).

The form of Russellian panpsychism[1] arising from these considerations is committed to the claim that consciousness is the intrinsic nature of matter which grounds all physical facts of reality. This claim can be cast in terms of categorical and dispositional properties. Dispositional properties are commonly defined as the directedness of an object towards a certain kind of manifestation, under appropriate conditions (Jaworski 2016: 57). For instance, a vase has the disposition to shatter when struck. More broadly, a dispositional analysis describes how something behaves in space and time, under this or that condition. Categorical properties, in contrast, are defined as powerless or non-dispositional features of objects, such as their shape and size (Jaworski 2016: 55). A categorical analysis describes what an object is like "in itself", non-relationally. Russell's motivation for panpsychism can now be put as follows: observational science only reveals the *dispositional* properties of objects, but it is silent about the *categorical* properties that ground these dispositions. The idea that consciousness is the intrinsic nature of matter can be understood as the claim that consciousness is the *categorical* property of objects, while the physical facts it grounds are the *dispositions* of those objects. This view is a hybrid approach between *pandispositionalism*, the claim that all properties are fundamentally dispositional, and *categoricalism*, the claim that all properties are fundamentally categorical (Choi and Fara 2012). Russellian panpsychists, as described here, thus accept the existence of both categorical and dispositional properties but specify consciousness as the universal categorical property of objects at the fundamental level of reality (Pereboom 2015).

3. *The Grounding Problem*

Russellian panpsychists are faced with an objection raised by Karen Bennett[2] and further developed by Derk Pereboom (2015). Pereboom

[1] There are other forms of panpsychism, but I will focus my attention only on the particular form of Russellian panpsychism as described here.

[2] Karen Bennett, "Why I Am Not a Dualist", ms., as reported by Pereboom (2015).

argues that micropsychists[3] need to introduce brute laws[4] in order to explain how "microphenomenal absolutely intrinsic properties" are linked to microphysical properties (2015: 317). Otherwise, the connection between microphenomenal and microphysical properties would be unintelligible. Pereboom (2015: 317) further argues that micropsychism is ill-equipped to explain the properties revealed to us by current microphysics, considering that brute laws generally provide no adequate explanations. Panpsychists thus need to account for how consciousness, as a categorical property, grounds the dispositional properties of objects.[5] To do that, they can either concede and posit inexplicable laws, or say that there are no such laws.

The former option is equivalent to the necessitation relation discussed by David Armstrong (1978, 1983), Fred Dretske (1977) and Michael Tooley (1977) within the framework of their view on natural laws. Necessitation can be defined as the law-making universal N which holds between universals or natural properties F and G, so that if a possesses F, then a necessarily possesses G (Armstrong 1978, 1983). This form of necessitation is a brute law since it cannot be reduced to a more basic level, which is problematic because it fails to provide an explanation where there should be one, committing us instead to an ontology where the notions used are—by definition—unintelligible. This issue was clearly formulated by David Lewis, who has argued (1983: 366) that Armstrong fails to provide a transparent account of necessitation, and that a relation is not necessary simply in virtue of being called "necessary". Considering that there are rival theories which offer a coherent explanation of properties without invoking brute laws, the panpsychist is seemingly left without strong reasons to posit them. Naturally, a panpsychist could claim that other theories do *not* offer a satisfying explanation, arguing instead that we need to have brute laws. Without going further into this extensive debate, I limit myself to proposing a solution to the grounding problem which will not rely on brute and inexplicable laws.

The latter option results in several problems as well. If consciousness is the categorical property that grounds dispositions, there is an immediate worry of *how* it could ground them if there are no brute laws. Positing consciousness as a categorical property is problematic unless it is in some sense connected to dispositions or powers. Without this connection, we end up with *epiphenomenalism*—the view that consciousness lacks causal efficacy. For an epiphenomenalist, mental events (or tokens of conscious experience) are caused by physical brain

[3] Pereboom uses "micropsychism" for all views that see consciousness as present at the fundamental level of reality. This includes the form of panpsychism I am discussing.

[4] Laws are brute when they have no further explanation or when they cannot be explained by appealing to something more fundamental.

[5] In this context "categorical" and "dispositional" are interchangeable with "microphenomenal" and "microphysical".

states, but mental events themselves have no effects on physical events whatsoever (Robinson 2015). If the panpsychist were to choose this option, they would have to deal with the following difficult questions:

a. How does consciousness ground dispositions without brute laws?
b. How do they avoid epiphenomenalism, a largely unattractive view nowadays, or—alternatively—why should epiphenomenalism be accepted?
c. If there are no brute laws linking categorical and dispositional properties, then why do only certain types of physical systems result in consciousness?

Question c) might initially appear as a variation on the combination problem for panpsychism,[6] but it is not. Instead, it is the following query: why would minds be specifically tied to brains (or any particular form of matter) if there were no laws linking categorical and dispositional properties? We have good reasons to accept a sort of parallelism between complex physical states and complex mental states: an intuitive and empirically justifiable answer to the question of why only human brains are capable of abstract and higher-order thought, as opposed to other animals, is that human brains are more advanced. A panpsychist claiming that there are no brute laws of grounding would have to deny this parallelism. In order to explain why only some physical states result in complex conscious subjects, the panpsychist would need to offer an account *alternative* to the claim that consciousness, as a categorical nature, is linked to certain types of physical or dispositional systems resulting in complex consciousness. Unless they introduce (brute) laws of grounding which hold between physical and mental states, it is not clear how they could explain the existence of such a parallelism, which is a largely uncontroversial concept. This is a big bullet to bite. However, this is not a reason to straight out reject a non- brute-law version of panpsychism. My intention here is only to show that this version of panpsychism leads to us having to accept a wide array of unappealing views. Because of that, I will try to develop a solution to the grounding problem which avoids these issues.

4. *The Identity Theory of Powers*

There is a way for the panpsychist to avoid the problems stated above and to offer a promising solution to the grounding problem. The identity theory of powers, discussed by Charles B. Martin (1994, 1997), John Heil (2003) and William Jaworski (2016), is uncommitted to the bifurcation of categorical and dispositional properties. For the identity theorist, "categorical" and "dispositional" only describe the differing theoretical roles properties play. In reality, though, there is no such

[6] The combination problem, in its most common variant, is the difficulty of explaining how simple (or the simplest) subjects combine into more complex subjects (see Chalmers 2016 and Goff forthcoming).

division: categorical and dispositional properties are one and the same thing. Every property possessed by an object gives it the power to interact with other objects in various ways (Jaworski 2016: 57). To illustrate this, Heil (2003: 112) uses the example of a snowball, whose spherical shape is traditionally understood as a categorical property. He (Heil 2003: 112) argues that the shape of the snowball confers to it the power to roll on a flat surface. In other words, sphericity is a quality or categorical property possessed by the snowball, but at the same time its power. Jaworski (2016: 63) uses a clearer example—a diamond—and argues that the diamond's hardness empowers it to scratch glass. In contrast, proponents of the hybrid view of properties would argue that the diamond's tetrahedral arrangement of carbon atoms is a categorical property which grounds its powers. The identity theorist argues instead that these descriptions denote different theoretical roles that one property plays—the categorical "is made out of carbon atoms" and the dispositional "scratches glass" role (Jaworski 2016: 54). In reality, though, the diamond's structure simply *is* its power to scratch glass (Jaworski 2016: 54).

The panpsychist can accept the identity theory of powers as an ontological basis and so avoid the grounding problem. This move indicates a step away from the Russellian view of consciousness as a categorical property which grounds dispositions or powers. However, it remains loyal to the basic Russellian motivation for panpsychism—the idea that matter must have an intrinsic nature. For the identity theory panpsychist, consciousness is a fundamental and ubiquitous property which is at the same time categorical and dispositional; a quality and a power. Identity theory panpsychism solves the grounding problem by eliminating the grounding relation: consciousness is no longer an isolated intrinsic nature serving as the categorical basis for dispositions but a property which fulfils both the categorical and dispositional role.

One distinct advantage of identity theory panpsychism over Russellian panpsychism is that it normalises consciousness by giving it the same ontological status as it gives to every other fundamental property. In Russellian panpsychism, consciousness is *the* categorical property, the intrinsic nature of matter, given primacy over all other properties. In identity theory panpsychism, consciousness is a fundamental and ubiquitous property like every other such property (e.g. spin, mass, electric charge, colour charge). In other words, consciousness is *a* fundamental property whose existence we need to admit in order to explain how complex subjects come into being, not *the* fundamental property which grounds all others. The bifurcation of categorical and dispositional properties present in Russellian panpsychism is rejected here for a simpler model of powerful qualities or properties that are at the same time dispositional and categorical. Because of this, identity theory panpsychism is a more parsimonious view since it avoids introducing more than one type of property or "ultimate" categorical proper-

ties. These are the positive reasons for why we should consider identity theory panpsychism as a serious option.

5. *Objections to the Identity Theory of Powers*

Before adopting the identity theory of powers, the panpsychist must first address specific issues the theory faces in its own right. One very important issue was raised by David Armstrong, who argues (Armstrong, Martin and Place 1996: 95) that the categorical and dispositional roles of a property must be related either contingently or necessarily. He goes on to explain that if the relation were contingent, then it would be possible for the categorical side to have different powers "attached" to it, "or even with no powers at all" (Armstrong, Martin and Place 1996: 95). To turn back to an earlier example, this means that it would be possible for the diamond's hardness to be correlated with the disposition *not* to scratch glass (Jaworski 2016: 78). This is not compatible with the identity theorist's view that the diamond's hardness is *identical* to its power to scratch glass (Jaworski 2016: 78). Armstrong (Armstrong, Martin and Place 1996: 96–7; as reported by Jaworski 2016: 79) further argues that if the relation were necessary, then it would be unclear why the roles are necessarily related. Importantly, the proponent of the identity theory would have to introduce *brute laws* to explain why the relation between categorical and dispositional roles is necessary (as reported by Jaworski 2016: 79). This means that accepting the identity theory of powers does not avoid the brute laws issue raised against panpsychism in the form of the grounding problem. Both panpsychism and the identity theory thus suffer from a version of the brute laws problem. If this is true, it would be devastating for the aims of this paper.

Luckily, there is a way of responding to this objection. The identity theorist could provide the following account of the categorical-dispositional relation and argue that it *is* necessary:

> [T]he reason why the diamond's hardness is necessarily correlated with the diamond's power to scratch glass is that the diamond's hardness is identical to the diamond's power to scratch glass. (Jaworski 2016: 79)

Armstrong finds this proposition "totally incredible"; claiming that it is a category mistake to identify categorical properties with dispositional properties; and concluding that "they are just different" (2005: 315). Jaworski responds by saying that Armstrong is begging the question: "To assume at the outset that qualities and powers are 'just different', as he says, is simply to assume that the identity theory is false" (2016: 79). In other words, when Armstrong claims that identifying categorical and dispositional properties is a category mistake, he is assuming without arguing that they cannot be identified at all. For Jaworski (2016: 79), this alone is enough to reject Armstrong's objection. Thus, while Pereboom's grounding problem does raise a valid point about brute laws to

Russellian panpsychism, Armstrong does not raise a valid point about brute laws to identity theory panpsychism. In the former case, there *is* an ontological bifurcation of categorical and dispositional properties, so Pereboom is justified in demanding an explanation of the relation holding between those properties. In the latter case, there is no such *ontological* bifurcation—"categorical" and "dispositional" are merely ways of describing the different theoretical roles a property can play. Hence, there is no need for an explanation of how these roles are related, unless one assumes (like Armstrong does) that these roles *cannot* be identified. However, as was shown, this assumption is question-begging.

Armstrong raises one further important objection. He (Armstrong, Martin and Place 1996: 16) starts by explaining that it is not a necessary truth that every power of an object is always manifested at some point of the object's existence. If we imagine an object which has some power but never manifests it, then its power is directed towards a manifestation which does not actually exist (Armstrong, Martin and Place 1996: 16–7). For example, even in a world without water, sugar would still have the disposition to dissolve when put in water. This means that its disposition to dissolve is aimed at some non-existent manifestation—and Armstrong thinks that properties cannot "point beyond themselves to what does not exist" (Armstrong, Martin and Place 1996: 17). In other words, Armstrong (as reported by Jaworski 2016: 58) is implying that the identity theory of powers is committed to Meinongian[7] non-actual entities since it allows that dispositions or powers can be related to not-yet-existent manifestations. This is deeply problematic.

As a response, proponents of the identity theory can reject the claim that powers are *real* relations to their manifestations (Jaworski 2016: 58). Instead, as Jaworski argues (2016: 58), the directedness of powers towards their manifestations can be understood through an analogy with intentional mental states. For example, I have a desire to eat pizza, but my desire can remain unfulfilled. It is the same case with powers—salt has the disposition to dissolve, but its solubility does not stand in a real relation to its manifestation. It is directed towards it analogous to how my desire for pizza is directed towards pizza, even if all pizzerias in my town go bankrupt and close down (Jaworski 2016: 57). In other words, the directedness of powers does not depend upon the existence of the manifestations they are directed towards (Jaworski 2016: 58). If the directedness of powers can be conceived of as analogous to intentional mental states, then the identity theorist can avoid the charge of being committed to Meinongian non-actual entities (Jaworski 2016: 57).

[7] Alexius Meinong, an Austrian philosopher and psychologist, is known for introducing non-existent objects as part of his ontology (Marek 2013).

Armstrong is suspicious of this solution. He (Armstrong, Martin and Place 1996: 17) claims that mental states have the property of being intentional, but expresses hope that they will ultimately be logically or empirically analysable. He thinks it strange and objectionable to put intentionality, or something like it, into the "ultimate structure of the universe" (Armstrong, Martin and Place 1996: 16). Similarly, Jaworski (2016: 58) stresses that the analogy between intentional mental states and the directedness of powers is merely that—an analogy. He claims, without providing an argument, that "intentional mental states are powers and the directedness of those states is a species of the directedness of powers in general" (Jaworski 2016: 58).

However, hopes and claims are not convincing arguments, which brings us to the question: why do Armstrong and Jaworski put the directedness of powers over mental intentionality? Their insistence on the primacy of directedness appears to be ad hoc. Otherwise, it could be understood as an intuitive argument, based on current sentiments in metaphysics and philosophy of mind. Whatever the case, Armstrong and Jaworski did not extensively discuss reasons for why they give primacy to directedness of powers. They could argue that we have no reason to ascribe intentionality or any aspect of consciousness to non-living matter since it does not exhibit behaviour we would characteristically describe as conscious. However, what we first observe as human beings is the fact that we are conscious and, as part of that, our ability to have intentional mental states. Indeed, the first piece of knowledge we ever acquire is the knowledge of conscious experience. We know for certain that consciousness exists and that we are conscious, but we can never know for sure whether other living and non-living beings have conscious experiences. The solution to this was to ascribe consciousness based on behaviour: *x* is conscious because it behaves similarly enough to us, while *y* is not conscious because it does not behave similarly to us (or behave at all).

Is behaviour really a good criterion for ascribing consciousness? We could easily imagine a dormant super-intelligent being, or a being so advanced that we appear as non-conscious or barely conscious to it. It is a relative scale. Cats and dogs appear less conscious (or less complexly conscious) to us, while plants and rocks appear non-conscious, but we could be so low on this scale relative to some existing or hypothetical intelligence that *we* would then be the rocks. Less extravagant examples are comatose patients. While outwardly these people appear unconscious, brain scans strongly suggest that they retain *some* level of consciousness (Cyranoski 2012). Of course, we know that patients were fully conscious before they fell into a coma, but would not very simple conscious subjects, whose standard level of consciousness is very low, always appear comatose to us? We would have no way of detecting conscious activity in such subjects. Thus, behaviour seems more like a provisional and pragmatic criterion for consciousness rather than as

a certain nomic principle. Moreover, since we know that consciousness exists with more certainty than we know anything else, positing that there are things which are not conscious introduces a new kind of entity to our ontology—non-conscious existents.[8] A more parsimonious view is one where consciousness comes in degrees, from rocks to amoebas to dogs to humans. That way, we avoid introducing a new and unproven ontological entity. The view that there are non-conscious existents has been so deeply ingrained into us that we cannot even consider the possibility that it might be wrong (or at least less explanatorily powerful). Nonetheless, in conjunction with independent arguments for panpsychism, I believe that we have good reasons to doubt that there are non-conscious existents.

6. *Concluding Remarks*

It is important to note that I have not been arguing for panpsychism in this paper. The discussion presented is aimed at philosophers who are already sympathetic to panpsychism. Specifically, in view of the grounding problem, I have argued that panpsychists are better off rejecting the Russellian ontological commitment to a hybrid view of properties, where the categorical grounds the dispositional. Instead, as I have claimed, there are good reasons for why they should accept the identity theory of powers as their ontological basis. The first reason is that identity theory panpsychism avoids the grounding problem. The problem of needing to introduce brute laws between two things disappears when only one thing with differing roles exists. The second reason, more positive in nature, is that identity theory panpsychism normalises consciousness by giving it the same status it gives to other fundamental properties, thus eliminating the need for introducing an additional special type of property.

Apart from addressing objections to panpsychism, I have also demonstrated that the combination of the identity theory of powers and panpsychism successfully addresses objections raised to the identity theory in its own right. Most importantly, an identity theory panpsychist has independent reasons for thinking that mentality, especially intentionality, is part of the structure of reality. In contrast, at least in cases addressed by this paper, Armstrong and Jaworski seem to merely assume that intentionality cannot be a part of reality and that primacy should be given to the directedness of powers. They are introducing more entities than panpsychism does to explain the same thing. Considerations of parsimony thus push us to consider identity theory pan-

[8] As a side note: The idea of matter being directed towards manifestations in a way analogous to intentionality, but without intentionality, is more mysterious to me than simply saying that this directedness is a form of intentionality, considering that we already know what intentionality is but have no idea of what the directedness of powers is, apart from the technical definition of the term and the demand that it involves no intentionality.

psychism as the theoretically more adequate explanation. That is why I believe that the combination of panpsychism and the identity theory is indeed a powerful one, and that it could serve as a future starting point for many philosophers of mind and metaphysicians.

References

Armstrong, D. M. 1978. *A Theory of Universals*. Cambridge: Cambridge University Press.

Armstrong, D. M. 1983. *What Is a Law of Nature?*. Cambridge: Cambridge University Press.

Armstrong, D. M., Martin, Charles B., Place, Ullin T. 1996. *Dispositions: A Debate*. Edited by Tim Crane. London: Routledge.

Armstrong, D. M. 2005. "Four Disputes about Properties." *Synthese* 144: 309–20.

Chalmers, D. 2016. "The Combination Problem for Panpsychism." In G. Bruntrup and L. Jaskolla (eds.). *Panpsychism: Contemporary Perspectives*. Oxford: Oxford University Press: 179–214.

Choi, S. and Fara, M. 2012. "Dispositions." *Stanford Encyclopedia of Philosophy*. <URL= http://plato.stanford.edu/entries/dispositions/>.

Cyranoski, David. 2012. "Neuroscience: The Mind Reader." *Nature* 486: 178–80. <URL= http://www.nature.com/news/neuroscience-the-mind-reader-1.10816>.

Dretske, F. 1977. "Laws of Nature." *Philosophy of Science* 44: 248–268.

Eddington, A. 1928. *The Nature of the Physical World*. Cambridge: Cambridge University Press.

Goff, P. Forthcoming. *Consciousness and Fundamental Reality*. Oxford: Oxford University Press.

Heil, J. 2003. *From an Ontological Point of View*. Oxford: Clarendon Press.

Jaworski, W. 2016. *Structure and the Metaphysics of Mind*. Oxford: Oxford University Press.

Lewis, D. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61: 343–377.

Marek, J. 2013. "Alexius Meinong." *Stanford Encyclopedia of Philosophy*. <URL= http://plato.stanford.edu/entries/meinong/>.

Martin, C. B. 1994. "Dispositions and Conditionals." *The Philosophical Quarterly* 44: 1–8.

Martin, C. B. 1997. "On the Need for Properties: The Road to Pythagoreanism and Back." *Synthese* 112: 193–231.

Pereboom, D. 2015. "Consciousness, Physicalism, and Absolutely Intrinsic Properties." In T. Alter and S. Walter (eds.). *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press: 300–323.

Robinson, W. 2015. "Epiphenomenalism." *Stanford Encyclopedia of Philosophy*. <URL= http://plato.stanford.edu/entries/epiphenomenalism/>.

Russell, B. 1927/1992. *The Analysis of Matter*. London: Routledge.

Seager, W. 2006. "The 'Intrinsic Nature' Argument for Panpsychism." *Journal of Consciousness Studies* 13 (10–11): 129–145.

Tooley, M. 1977. "The Nature of Laws." *Canadian Journal of Philosophy* 7: 667–698.

Croatian Journal of Philosophy
Vol. XVII, No. 49, 2017

A Philosophical Critique of the Concept of Miracle as a "Supernatural Event"

ADAM ŚWIEŻYŃSKI

Institute of Philosophy, Cardinal Stefan Wyszynski University, Warsaw, Poland

The notion of the supernaturality of an event may be understood in various ways. Most frequently 'supernatural' means 'separated from nature', i.e. different from nature. Thus, what is meant here is the difference in ontological character. The definitions of miracle, present in literature, emphasize the fact that we may talk about a miracle only when the phenomenon takes place beyond the natural order or stands in opposition to it. The description of a miracle as a 'supernatural event' contains in itself the reference to that which is natural. The supernaturality of an event means that it surpasses (transcends) naturality. Additionally, this transcendence contains a kind of opposition to that which is natural. However, the miracle as a supernatural event takes place within the scope of that which is natural, although it takes place in a different way from natural events. It seems that this supernaturality may involve two things: (1) the course of the miraculous event; (2) the cause of the miraculous event. We should consider each of them separately and specify what we understand by the supernatural course of the event and by the supernatural cause of the event. If we could prove that we can talk about supernatural events at least in one of the two signaled aspects of supernaturality, then we would be able to define the miraculous event as a supernatural one. The analyses proposed in the paper allow us to formulate the following statement concerning the miraculous event, which is, to a great extent, a critical correction of the traditional way of understanding it: the miracle may be correctly understood as a supernatural event, only when this supernaturality concerns the personal cause of the event and not its course.

Keywords: Laws of nature, miracle, ontology, supernaturality.

1. *Introduction*

The notion of the supernaturality of an event may be understood in various ways (see Williams 1990 and Daston 1991). Most frequently 'the supernatural' means 'separated from the nature', i.e. different from the nature. Thus, what is meant here is the difference in ontological nature. Sometimes, the events understood as supernatural ones are those that belong to a certain part of nature inaccessible to human knowledge. In this case, the problem of supernaturality is reduced to the question of human cognitive limitations. Therefore, the supernatural thing is the one, which hasn't been known yet or which will never be known as natural.[1]

The definitions of miracle, we can encounter, emphasize the fact that we may talk about the miracle only when the phenomenon takes place beyond the natural order or stands in opposition to it.[2] As a result, the natural (scientific) explanation of the event is not possible and will never be so. It seems, therefore, that the attribute of supernaturality, which expresses the ontology of the miracle, is regarded as an irreducible base for asserting its absolute inexplicability in terms of nature.[3] Simultaneously, the miracle as a supernatural event is regarded as the act of exceeding the laws of natural sciences (scientific laws) as well as the laws of the nature itself.

The description of the miracle as a 'supernatural event' contains in itself the reference to that which is natural. The supernaturality of an event means that it surpasses (transcends) naturality. Additionally, this transcendence contains a kind of opposition to that which is natural. Although the supernaturality of the event is a kind of unnaturality, the natural element is not entirely annihilated by it. Rather, we should

[1] See Miller, Vandome and McBrewster (2009: 36–37). Such an approach to the supernaturality of the miracle is present e.g. in John Locke's works. For Locke, the violation of the established course of nature by a miracle involves merely the violation of the laws, causes and effects we know. Thus, the miracle, understood as a violation of the laws of nature involves, in fact, the conformity with laws that are unknown to us. These laws, together with the ones we know, constitute the full set of the 'laws of nature' (see Mooney 2005: 150).

[2] The notion of the miracle as an event that contradicts natural laws originates from the distinction between the natural and supernatural causes, introduced by Anselm of Canterbury. William of Auvergne, in turn, distinguished two elements within the notion of miracle: the Divine origin and the opposition to the forces of nature. The description of the miracle as a fact as opposed to nature, most probably, appeared for the first time in the work of medieval scholar, Alexander of Hales'. Yet, he noted that specifying the miracle as a 'contra naturam' event is insufficient, as strange and mysterious things may also take place that are inconsistent with nature or even in opposition to it and they are not miracles, because they arise from natural causes (see Grant 1952).

[3] Such an approach towards miraculous events is characteristic of apologetics (fundamental theology), and is manifested in numerous statements concerning miraculous events such as 'violating the laws of nature' (see Hesse 1965: 36; Walker 1982: 103–108; Basinger 1984: 1–8).

say that what we have in the case of a miraculous event is the metamorphosis of the natural into the supernatural.[4]

Although the supernatural event is usually regarded as being brought about in an unnatural way, it is not a necessary condition of the supernaturality of the event. The supernatural event may have no cause, and despite this fact, it may be the event going 'beyond' the causal force of nature. For instance, a cosmologist with purely materialistic views may say that the first natural phenomenon in the history of the cosmos was a supernatural event, which was not engendered by any previous natural cause.[5] Moreover, although it is necessary for the supernatural cause to be unnatural in character, the supernatural event may be both natural and unnatural in its course. The only requirement is that the supernatural event cannot be brought about in a natural way, i.e. by a natural cause. It may be useful at this point, to introduce the distinction between the permanently (unconditionally) supernatural event and the conditionally supernatural one. The former is the event, which may never be caused by a natural cause. The latter, however, could be caused by a natural cause on certain conditions, but in this particular case, these conditions are not met.[6]

Hence, it is sometimes suggested that the miracle should be described as the natural effect of the event which was brought about by an unnatural cause, and which couldn't be brought about in a natural way (see P. Dietl 1968: 130–134; Young 1972: 123; Ward 2002: 741–750). Such a definition doesn't contain the direct statement concerning the character of the unnatural cause. Hence, scholars claim that the miracle is the event, which remains beyond the capabilities of nature and its activities. They talk about miraculous events as being exclusively unnatural, and not as being merely supernatural.[7]

However, the question of the degree of transcendence, of that which is natural within supernatural events, is still a matter of debate among authors dealing with the problem of miracles.[8] They commonly agree

[4] For example, biblical miracles are supernatural events taking place within the natural world (Ex 14,1–30; 2Chr 5:1–14; Jn 2:1–11 and many more)

[5] Such a situation may take place in the case of cosmology of cyclic cosmos, in which we are unable to indicate the first natural event. For example, see Steinhardt and Turok 2001: 1436–1439.

[6] For instance, the virgin conception of a child is naturally possible with the use of so-called artificial insemination, yet, it wasn't so in the case of Christ's conception by the Holy Virgin, as the appropriate medical technique was unknown then. Yet, the very distinction between that which is 'natural' and that which is 'artificial' seems arguable in many cases (see Meller 2010: 191–199).

[7] Not every unnatural cause need be regarded as a supernatural one, although each supernatural cause would, at the same time, be an unnatural one. Thus, we may still distinguish the category of 'merely unnatural cause' (see Clarke 2007). It doesn't change the fundamental problem of the unambiguous determination of the different nature of these causes.

[8] "The fundamental problem is not about miracle, but about transcendence" (Hesse 1965: 42).

that the miracle is the effect of God's action, but they argue with regard to determining a sufficient basis for asserting God's intervention in nature. Some of them think that the miraculous phenomenon has to be one that has not been explained by science so far.[9] Others tend to be stricter and claim that in order for a given event to be classified as a miracle, it has to be proved that it is not only unexplained so far, but also can never be explained.[10] Still others express the opinion that even the phenomenon, for which there exists a natural explanation, a miraculous event has only occurred, provided we know for certain that it was actually performed by God (see Clarke 1997).

Thus, the miracle treated as a supernatural event should be regarded as transcending regularities that exist within nature and those attributed to it by natural scientists. Yet, in the case of the transcending regularities that exist within nature and those attributed to it by natural scientists. Yet, in the case of the aforementioned transcendence, we have not only insufficient human knowledge about the world and its processes, but also the transcendence of a certain state of nature—i.e. its internal regularities—independent of human knowledge. The supernatural event is, therefore, regarded as the event transcending the laws of nature, and constituting the ontological structure of material reality. Because of this transcendence, the miraculous event is also treated as inexplicable within the methods and explanations provided by natural sciences.

The discrepancies just signaled, in which there also appears the problem of a natural inexplicability of the miracle, make us reflect more deeply upon defining the miraculous event as a supernatural one. It seems that this supernaturality may involve two things: (1) the course of a miraculous event; (2) the cause of a miraculous event. We should consider each of them separately and specify what we understand by the supernatural course of event and by the supernatural cause of event. If we could prove that we can talk about supernatural events in at least one of the two signaled aspects of supernaturality, then we would be able to define the miraculous event as a supernatural one.

2. *A critique of the concept of miracle as an event with a supernatural course*

'Extraordinariness' of the course of event can be understood as being in the epistemological or ontological category. Thus, there are situations (at least potentially), in which our being surprised and astonished can-

[9] Yet, some scholars think that such an approach towards the miracle carries in itself the danger that a phenomenon in the current state of knowledge regarded as a miracle may turn out to be a natural one in the future.

[10] "We can only speak of a miracle when an event occurs outside and against the known order of nature. This event must not be open to any natural explanation whatsoever, and it must also never be capable of explanation in any natural way whatsoever" (Loos 1965: 46).

not be treated merely as the consequence of lack of knowledge of the nature of the world (a lack that may be overcome by gaining a more thorough knowledge of reality); it is rather, that our being surprised and astonished should be treated as something related to the irreducibility of the unpredictable character of natural processes, that follow from their functioning in a way that is different from the normal (natural) one. However, the ontological extraordinariness doesn't seem to be the necessary determinant of that which is miraculous. This is so, because the supernatural course of the event is the one that should differ significantly from the natural course. The supernatural course should mean violation, suspension, or surpassing the regularities of nature. Each of the situations just mentioned, concerns, in turn, the change within the metaphysical structure of material beings or imposing on them (from outside) a new way of acting and interacting. Yet, it seems that, in both cases, the way the world functions remain natural but different with respect to the phenomenal sphere.

So far, understanding miraculous events as the ones violating, suspending or surpassing the laws of nature in force is, to a great extent, the consequence of the picture of the world, which was provided by the emergence and development of the natural sciences. The period of looking at nature in a mechanistic and strictly deterministic way, especially in the 18th, and partly, in the 19th centuries, strengthened the conviction that events and processes inconsistent with the established regularities of nature violate its laws. Yet, further development of natural sciences questioned such an approach towards phenomena, which couldn't be explained by adopted scientific theories. The remarkable example of this change is the emergence of quantum mechanics in 20th century. The rules of quantum mechanics are not deterministic but statistical. The fact that contemporary natural sciences rejected the strictly deterministic picture of reality changed the status of these sciences as the one that determines accurately what is or is not possible within nature. Existing scientific theories turned out, and still turn out, to be susceptible either to partial modifications or to being totally questioned.[11] Yet, the switch from Newtonian to quantum physics, as well as the emergence of deterministic chaos theory and of other theories didn't significantly influence the way miraculous events are understood. They still are the events, which by their nature, fall beyond the regularities of the natural world. Because of the lack of any clearly formulated idea, the question of supernaturality of miraculous events still remains a matter of debate.

In considerations concerning the miracle being understood as the violation or suspension of regularities of nature, we may encounter the opinion that the very concept of suspension or violation of some

[11] The example of such changes in cosmology may be the theory of the Stationary State, which was refuted because of new empirical results concerning universe expansion (see Singh 2005).

regularity is internally contradictory. If the true event *Z* occurs inconsistently with the nomological principle *N* concerning the course of phenomena, it means that the principle *N* doesn't determine properly 'that which cannot happen', and for that reason, this principle can no longer be treated as nomological. Yet, if the principle *N* is really a nomological one, the event *Z* cannot be regarded as its actual violation. So, the event *Z* cannot be understood as being an 'actual' violation of any regularity. The nomological principle is regarded as the universal and necessary law (see McKinnon 1967: 309–312; Flew 1976: 28–30).

Other authors, who think that the fundamental problem connected with the concept of miracle as the event that breaks the regularities of nature involves the fact that this conception is used to defend the supranaturalistic approach within theistic apologetics, argue with the above opinion (see Corner 2007: 2; Byrne 1978: 166–169; Kellenberger 1979: 152–153). They claim that in the case of the natural functioning of nature, the laws of nature indicate that we have a situation, in which there is no intervention by God. But these laws do not inform us about the way the world functions in the case of divine intervention. When this intervention takes place, the laws of nature are violated and a miraculous event emerges (see Otte 1996: 155).

The treatment of miraculous events, which in their course, surpass the laws of nature, requires a more detailed description of the ontological structure of a supernatural event, and then, considering the validity of describing the miracle as a supernatural event. The miraculous event surpassing the laws of nature may be treated as the exception from these laws. We should then wonder whether such an event is supernatural or natural in character. The answer will depend on the adopted type of the cause of a given event. Let us suppose that the event *X* is inconsistent with the law of nature *P*, confirmed many times. There are three possible explanations of the occurrence of the event *X*: (1) some unknown (and perhaps inscrutable) natural cause brought about this event; (2) the event *X* was brought about by the action of the supernatural cause; (3) the event *X* doesn't have a natural or supernatural cause; it can be regarded as a single, unique anomaly.

In the case of first option, there is no reason for understanding the event as surpassing the law of nature and for treating it as a supernatural event. In the second case, however, the event is treated as surpassing the law of nature and hence it is a supernatural event. Yet, if the laws of nature determine what happens (or doesn't happen) in specific natural circumstances, they cannot be used to explain the event, which happens when the supernatural cause acts. Therefore, even if the event that took place is inconsistent with the law of nature and was brought about by a supernatural cause, we wouldn't be able to say that it surpasses the laws of nature and hence it is a supernatural event. The third option, in turn, assumes that the law of nature is adequately and empirically confirmed and the event, which takes

place, does so only once. Thus, we can say that the principle and exception from it are present simultaneously, namely, that the type *X* events both occur and do not occur in the same natural circumstances. Such a situation would mean that we wouldn't have to make a choice between the rejection of event *X* and the modification or rejection of the law *P*. Some authors express the opinion that only such events may be regarded as surpassing the laws of nature (see Basinger and Basinger 1986: 13–14). Thus, it would be a supernatural event, not because of its supernatural cause, but because it surpasses the laws of nature, i.e. its supernatural course. Nevertheless, such an event couldn't be described as a miracle, as it excludes the action of any cause, including God. Therefore, in the light of the options just considered, we have the alternatives: the event is supernatural either because of its course (it can be described as violating, suspending or surpassing the laws of nature), or because of the action by the supernatural cause, which brought it about. The third option, in which an event is supernatural, due to both its supernatural course and cause, turns out to be unnecessary, because the action of the supernatural cause doesn't necessarily have to generate the supernatural course of the event. In the case of the second element of the above alternative, the event is not supernatural in its course (it is not questioning the laws of nature), but is supernatural because of its supernatural cause.[12] Thus, the second element of the above alternative, i.e. the action of the supernatural cause, is sufficient to classify the event as the supernatural one, without deciding whether its course is, or is not, supernatural.

It is reasonable to present fundamental difficulties, which emerge when a supranaturalistic conception of the miracle is adopted, with regard to its supernatural course. The element, appearing within the conception just mentioned, is the attempt to define the miracle as the event that directly violates the laws of nature, or at least, the one that surpasses these laws or brings about any other form of intervention into the natural function of the world. Yet, there is no clear reason for accepting the view that the event, which cannot be subject to any natural regularity, has to be treated as the violation of this regularity. While analyzing the conception of miracle as the event violating the laws of nature, we have to note that within this framework, the miracle is treated as something, which 'tears apart' the structure of nature, and hence the miracle is possible only if we assume the existence of an efficient cause external to nature. Yet, the internal contradiction is not obvious within the very conception of violating the laws of nature, as contemporary writers want it. There is no inconsistency in the state-

[12] According to Mumford, the best way of understanding the miracle is to treat it as the event, which is natural with regard to its course, but having its supernatural cause. In Mumford's opinion, such conception of the miracle may (but doesn't have to) lead to the claim that its emergence is necessarily connected with breaking the laws of nature (see Mumford 2001: 191–202; cf. Clarke 2003: 459–463; Luck 2003: 465–469; Clarke 2003: 471–474).

ment that an event happened, which we cannot subordinate to the laws of nature, and that the laws of nature are understood as fully determined regularities.[13] But there is no reason to treat such an event as a violation, i.e. as something, which in some way, is inconsistent with the real structure of the natural world or as something that forces us to accept the existence of anything surpassing nature.[14] It is impossible to point to any empirical criteria when distinguishing the anomalies caused by supernatural intervention into nature from 'ordinary anomalies' or from spontaneous breakdowns of natural order. This is why supranaturalists have no reasons for claiming that a specific anomaly is the result of supernatural intervention into the natural order of things and that the emergence of this anomaly means the supernatural course of events. Let us emphasize here that there exists the possibility of proving the distinction just mentioned, in the case of capturing supernatural intervention in teleological terms.

It is worth noting once again, that the main problem connected with the conception of the miracle as the event that breaks the laws of nature involves the fact that this conception is used to defend the supranaturalistic approach. But the category of the supernatural course of a miraculous event turns out to be useless for an apologist, who seeks to persuade us that nature is not all that exists.[15] This is so because it is impossible to provide a way of distinguishing the event proceeding in the supernatural way, from the one being an ordinary natural anomaly. It seems, therefore, that we should search for other objective criteria in

[13] The law of nature is only conditionally (physically) necessary; it is not absolutely (metaphysically) necessary, as its negation leads to falseness and not to absurdity. If the laws of nature are not absolutely, but relatively necessary, miraculous events are not contradictions in themselves.

[14] We can also imagine the situation, in which the miracle means a natural effect caused by the supernatural cause, and this natural effect could also potentially be brought about in a natural way, by a natural cause. Miraculous events understood in this way can be divided into two categories: (1) 'replacement' miracles—when the natural cause, which could appear in a natural way, is actually brought about by a supernatural cause; (2) miracles 'through the natural non-determination of phenomena'—when the natural effect, which can appear in a natural way, is not caused by a natural cause, but, at the same time, this natural effect is different from the one, which would appear, if it was not caused by a supernatural cause. Scholars started talking about miracles of the second kind together with the emergence of quantum mechanics. These miracles became popular, because in their supporters' opinion, if at the atomic level nature is not determined, then God could intervene at this level, without causing the supernatural course of the event, and merely 'choosing' a specific quantum state of a physical system. Manipulating the initial conditions at the quantum level, God may bring about unusual events that are inconsistent with the regularities observed at present (see Murphy 1995: 112).

[15] "The fact that our senses and measuring apparatus are able to capture some of these things, while some others are not, is the epistemological not ontological problem. So if we want to adopt the ontological criterion, in spite of all, then, if we are unable to distinguish between the nature and non-nature, we have to assert that the nature includes all the things, including angels and miracles, if we believe in them" (Tałasiewicz 2007: 408).

determining that which is extraordinary-supernatural and that the extraordinariness of the event, understood in an ontological way, doesn't have to be identified with violating, suspending or surpassing the regularities of nature (see Adams 1992; Hanfield 2001; Larmer 2011).

3. *The critique of the concept of miracle as an event with a supernatural cause*

The conclusion to the previously made detailed considerations of the supernatural cause of an event is the rejection of these conceptions of miracle, which assume that a possible miraculous event can only be explained by pointing to the supernatural cause, as being the one that is responsible for its occurrence.[16] The supranaturalistic approach, which I'm criticizing, treats the supernatural cause as the hypothesis explaining the event, concurrent with the naturalistic attempts of explaining the event. Hence, if it is possible to point to natural causes being responsible for the event, referring to the supernatural cause no longer makes sense.

The fundamental problem connected with the notion of a supernatural cause is that supranaturalists treat the supernatural cause analogically to the natural one. Yet, such an analogy should be regarded as the empty one, because treating the supernatural cause similarly to the natural one changes, each time, our notion of the supernatural cause to that of a natural cause. Additionally, there exists no way of characterizing the supernatural cause without making an analogy with the natural one. But if we seek to preserve the fundamental distinctness of the character of supernatural and natural cause, then there would be the problem of determining the way the supernatural cause influences the natural elements of the world (see Miles 1966; Pratt 1968; Saler 1977).

Thus, those who defend the claim concerning the supernatural cause of some event, encounter a dilemma—two possible solutions both of which turn out to be unsatisfactory. A supranaturalist, willing to explain the conception of supernatural cause, characterizes it in a way similar to the natural one. In consequence, the difference between the two causes is obliterated, and the supernatural causality is reduced to the natural one. If the supporter of the existence of a supernatural cause wants to justify its distinct character, he may encounter another problem. When he accepts its distinctness from a natural cause and treats it as an unnatural cause, a doubt arises concerning the possibility of defining it as a cause as such, since the common basis for comparing both causes is removed. Moreover, the radical distinction between the natural and supernatural raise questions on the abilities of causal impact of that which is supernatural, on that which is natural.

[16] There is also the possibility of understanding the supernatural cause as the one cooperating with the natural ones. In this case, the supernatural cause doesn't exclude the operation of natural causes.

The analogy between the natural and supernatural cause turns out to be inadequate in the sense that the supernatural cause doesn't have in it a certain crucial feature, which the natural cause possesses, namely, the property of physical impact. Thus, it is unknown how the supernatural cause influences the natural world, and if it is impossible to explain, in what sense can we talk about the supernatural cause as the one analogous to the natural cause? Moreover, in order to use the analogy in question, we should assume that the action of the supernatural cause is subject to specific laws, as it is in the case of the natural causes operating inside the world. These laws should be distinguishable from the laws concerning the functioning of nature. Yet, we do not know the laws other than those functioning inside the universe. Thus, what we should do is either to assume that the interactions between nature and the supernatural are subject to the laws of the nature we know, or to speculate on the existence of some unknown laws governing these interactions. In the first case, that which we describe as the supernatural turns out only to be the continuation of that which is natural and the expansion of the applicability of natural laws. In the second case, however, we should assert that we can say nothing about these unknown laws. We may observe the cooperation of nature and the supernatural just from the viewpoint of the observer situated inside the natural universe and using its laws; and this doesn't give us the chance to reasonably use the analogy between the natural and supernatural laws, or even to say something positive about the existence of the latter. The laws concerning nature always operate together with the physical properties of bodies, e.g. their mass, momentum, electric charge, etc. Then what would the statement mean that the laws governing the interaction between the natural and supernatural being 'is similar' to the laws governing the interaction of material bodies, with the objection that, because one element of the interaction is supernatural, i.e. nonphysical, it is not the interaction between material bodies? Once again, we see that the analogy is inadequate (Corner 2015: 48–49).

Thus, the supernatural cause cannot possess any physical properties, and if such properties are attributed to it, it becomes the natural cause. If we treat both kinds of causes as totally distinct from each other, then, because we know only the natural causes, we may wonder if the supernatural action may still be treated as the cause.

A similar difficulty may be observed within the conception of a supernatural explanation, which is a further element of the supranaturalistic conception of the miracle. This explanation is reduced to approving the action of the supernatural cause. If it is applied in terms of being an analogy of scientific (natural) explanation, it should have the property of empirical verifiability, which obviously seems impossible, because of the total distinctness between the supernatural cause and the natural causes. If, in turn, empirical verification of the action of the cause, which remains beyond the set of causes known so far, the

conception of supernatural explanation would turn out to be unnecessary, because each explanation, which can be verified in an empirical way, loses the property of the supernatural explanation. So, we can assume that a given event is the miracle manifesting divine action, but we shouldn't explain this event by looking for a supernatural cause. If we search for the explanation of a miraculous event, this explanation is completely different in character from the one used within natural sciences. Such an explanation should not refer to pointing to the cause, but should be teleological in character. Particularly, if we agree that explaining the event is realized not only by referring to the laws of nature, but also by providing the meaning of a given fact.

There is still one more problem to be discussed here. It appears that when describing a natural anomaly such as the event with the supernatural cause, we gain nothing. Why would the reference to the supernatural cause be better than approving the action of some unknown natural cause or lack of any cause at all? The exception here is the situation in which we understand the supernatural cause as the personal one, which is identified with God's action. Yet, those two terms are not synonymous (although they are often used interchangeably). Thus, only if we treat the anomaly as a manifestation of personal divine action (analogical to human action), are we able to prove the significant contrast between an event of this sort and an 'ordinary' anomaly, i.e. a spontaneous break in natural order. The very assertion concerning the action of the supernatural cause changes nothing, because such cause, by its nature, cannot be connected with the space and time of our world. Its action cannot be transmitted by any physical interaction.

Let us apply here the comparison to hypothetical material objects with features that are impossible to recognize empirically. Even if a given object had an unrecognizable feature, it would contribute nothing to our knowledge of it in relation to our knowledge of the objects without this feature. By introducing the supernatural cause, and treating it, at the same time, as a special kind of natural one, we gain nothing. Because we cannot imagine the supernatural cause in any way other than as an analogy of the natural cause, we should propose, as a replacement, the conception of the supernatural-personal cause and, in consequence, the teleological approach towards the miracle as the manifestation of God's will and action, together with the context it is manifested in. Simultaneously, we should move away from capturing God's action in purely causal terms, particularly, when understood as having an outside (interventionist) impact on the world.

Thus, the basic mistake concerning the conception of a miraculous event is the application of an interventionist conception of God's action (breaking the laws of nature), as well as combining it with the notion of a supernatural cause and supernatural course of the event. It leads to the emergence of the opposition between God and nature, which is absolute, and impossible to overcome notionally; it also leads to a one-

sided way of looking at miraculous events as the effects of divine action understood in terms of the way an efficient cause operates.

4. *Conclusion*

David Hume, one of the most famous critics of the possibility of miraculous events, expressed the conviction that the accounts of miracles and prodigies will be found in all history, sacred and profane (D. Hume, *An Enquiry Concerning Human Understanding*, *Section X: Of Miracles*). The accuracy of the prediction made by the Scottish rationalist has been confirmed in the subsequent centuries (including the present one). This confirmation was made through the constant appearance of such accounts, and discussions, which concerned, and still concern, the possibility of the occurrence of events described as miraculous, and the nature of these events. Moreover, Hume's statement seems to reveal the element of human nature, which generates the human need of accepting new intellectual challenges in the face of such events, or at least, the theoretical possibility of their occurrence. Without judging at this point, how to classify the events described as miraculous, we should say that the miracle is a particularly interesting object of interest for the human mind. This is because of the mystery accompanying the miracle; because of the complexity of the problems considered with respect to the miracle; and because of the views revealed when discussing the miracle.

Yet, is the problem of the miracle important and interesting from a philosophical point of view? The views in this respect vary considerably, yet it seems that the notion of a miracle and its content should be interesting for those who attempt to know the nature of the reality around them, and the reality that they are an element of; and also for those who endeavor to understand the process of discovering the world and the existential experience of a human being. It appears that a miracle, and the considerations of it, exemplifies the content of these very fundamental questions stimulating everyone who tries to gain at least a slightly better understanding and at least a bit more wisdom. If a miracle itself is the peripheral problem for philosophy in its traditional sense, the problems it poses are certainly, very important for philosophers as the basis for genuine philosophical quests.

What we can also observe in contemporary philosophy of the miracle is the characteristic trend towards 'naturalizing' miraculous events. This tendency in philosophical quests takes two basic forms: (1) the tendency to explain miraculous events by suggesting the manner in which God would act within nature (i.e. explaining the 'mechanism' of God's action within nature); and (2) the tendency to reduce miraculous events to purely natural ones, the explanation of which should be sought within constantly developing natural sciences. Both the aforementioned ways of 'naturalizing' the miracle pose certain difficulties. The first could be described as a 'moderate naturalization'. Although

it preserves the notion of miracle, there are some objections against it, namely because, it imposes a certain vision of God's action within nature, while trying to negotiate this vision with the present state of natural knowledge about the world. The second, however, goes even further; it can be described as a 'radical naturalization', because it seems to lead straight towards questioning the traditional sense of miracle the possibility of its occurrence, and as a result, to classifying it as an ordinary natural phenomenon. Both forms of naturalizing miraculous events, present in literature, seem to be dead ends as far as their results are concerned. They lead either to endless speculations on God's interactions with nature, or to eliminating the miracle as such. If we want to avoid both dangerous situations and their results, we should take a fresh look at the problem of the miracle and we should find a new way of understanding it.

Understanding the miracle is closely connected with understanding it as an event caused by God.[17] It is usually assumed that if a miraculous event is the effect of God's intervention in the material world, it must be regarded as different from the ordinary (natural) phenomena of nature. In this case, the postulate of regarding the miracle as a supernatural event is the consequence of understanding the miraculous event, as the one, the efficient cause of which is God. Yet, we can adopt the reverse way of argumentation, namely, starting from the ontological extraordinariness of the event, understood as its supernaturality, we can search for an adequate cause for events of this type. This way of analyzing the notion of miraculous event has the philosophical advantage of not assuming a priori that this event was brought about by the actions of a transcendental being on nature.

If we accept the possibility of the existence of extraordinary-supernatural events we may (and even should) think of their cause. The potential occurrence of supernatural events, because of their being ontologically diverse from the natural ones, requires the appropriate justification. It means the necessity to point to the cause, which would be capable of bringing about a supernatural event. Because natural causes are capable of bringing about only natural effects, the cause, which would be responsible for the occurrence of a supernatural event, should also be supernatural in character. The supernatural character of the cause bringing about a supernatural event means that it cannot be any cause coming from the field of nature. It is the case with both the part of nature, which is already known to us, and the natural processes and phenomena, which are still cognitively inaccessible. We assume that both the field of known natural phenomena and the unknown ones, and probably, the inscrutable ones too, is governed by the internal principles characteristic of this field, and hence, on its own, it

[17] The authors dealing with the problem of miraculous events share the conviction that if there is no reason to regard a given event as caused by God, there is no reason either, to regard it as a miracle (see Corner 2015).

doesn't generate the events that can be regarded as supernatural ones. Thus, we should take into account that, the principle being the fundament of causality, the effects are of the same nature as their causes, i.e. the effects are proportional to their causes.

Thus, while searching for an adequate cause of supernatural events, we may determine it as the external cause, transcendental in relation to the material world. Within a strictly philosophical perspective, the absolute being is usually regarded as such a transcendental factor. Within a philosophical and religious perspective (e.g. Christianity, Judaism, Islam), however, the factor in question is called God and treated as the unique personal being. God, as a being, not belonging to nature, and His existence that is significantly different in character from the material beings, seems to be regarded as the main candidate for causing a supernatural event; this is because of the characteristics, which are attributed to Him.[18] Thus, the miracle, understood as a supernatural event, may be justified by the action of supernatural cause, which is seen to be God.

Some authors claim that all the adequate and complete explanations causal in character should be the scientific explanations, namely, they should determine empirically all the conditions, both necessary and sufficient, for the occurrence of a given phenomenon. Therefore, if God's action is, by its nature, non-empirical, any event caused directly by God contains in itself the efficient cause, empirically unverifiable. Thus, such an event is supernatural and it cannot be adequately explained within natural sciences. This is why such an explanation cannot be regarded as the one, which is causal in character (Nowell-Smith 1950). For instance, the prayer that precedes the sudden healing of an ill person may be regarded as the circumstance preceding the healing and directly connected with God's action, the result of which is the recovery. Yet, God and His actions are, by their nature, imperceptible to the human senses.

References

Adams R. M. 1992. "Miracles, Laws of Nature and Causation – II." *Proceedings of the Aristotelian Society. Supplementary Volume* 60: 207–224.

[18] Also other immaterial beings, spiritual beings, (both good and evil ones), who would be able to influence with their actions the course of the phenomena taking place in the world may be regarded as the agents of supernatural events. Another question is whether such an action may be called a miracle. This action should be considered within the context of their created nature. Good spirits only execute God's will; namely, they are merely the instruments of His actions. While evil spirits cannot, by their actions, realize the good intended by God. Thus, we cannot regard their activity as the miraculous one, since they are either the intermediate element of miracle, which is worked by God, or their action is not oriented towards the good, which, within Christian theology, contradicts the crucial characteristics of the miracle, i.e. the good purpose of the occurrence of miraculous event (see Lawton 1959: 33; Beaudoin 2007; Weddle 2010: 28–29).

Basinger D. 1984. "Miracles as Violations: Some Clarifications." *The Southern Journal of Philosophy* 22 (1): 1–8.

Basinger D. and Basinger R. 1986. *Philosophy and Miracle. The Contemporary Debate*. Lewiston – Queenston: The Edwin Mellen Press.

Beaudoin J. 2007. "The Devil's Lying Wonders." *Sophia* 46 (2): 111–126.

Byrne P. 1978. "Miracles and the Philosophy of Science." *Heythrop Journal* 19: 162–170.

Clarke S. 1997. "When to Believe in Miracles." *American Philosophical Quarterly* 34 (1): 95–102.

Clarke S. 2003. "Luck and Miracles." *Religious Studies* 39 (4): 471–474.

Clarke S. 2003. "Response to Mumford and Another Definition of Miracles." *Religious Studies* 39 (4:) 459–463.

Clarke S. 2007. "The Supernatural and Miraculous." *Sophia* 46 (3): 227–285.

Corner D. 2007. "The Philosophy of Miracles." London – New York: Continuum.

Corner D. 2015. "Miracles (The Definition of 'Miracle')." *Internet Encyclopedia of Philosophy,* http://www.iep.utm.edu/miracles/#H1; (download: 30.10.2015).

Daston L. 1991. "Marvelous Facts and Miraculous Evidence in Early Modern Europe." *Critical Inquiry* 18 (1): 93–124.

Dietl P. 1968. "On Miracles." *American Philosophical Quarterly* 5 (29): 130–134.

Flew A. 1976. "Parapsychology Revisited: Laws, Miracles and Repeatability." *The Humanist* 36: 28–30.

Grant R. M. 1952. *Miracle and Natural Law in Greco-Roman and Early Christian Thought*. Amsterdam: North Holland.

Hanfield T. 2001. "Dispositional Essentialism and the Possibility of a Law-Abiding Miracle." *The Philosophical Quarterly* 51: 484–494.

Hesse M. 1965. "Miracles and the Laws of Nature." In C. F. D. Moule (ed.). *Miracles. Cambridge Studies in their Philosophy and History*. London: 33–42.

Kellenberger J. 1979. "Miracles." *International Journal of Philosophy of Religion* 10 (3): 145–162.

Larmer R. A. 2011. "Miracles, Divine Agency, and the Laws of Nature." *Toronto Journal of Theology* 27 (2): 267–290.

Lawton J. S. 1959. *Miracles and Revelation*. London: Lutterworth Press.

Loos R. 1965. *The Miracles of Jesus*. Leiden: E. J. Brill.

Luck M. 2003. "In Defence of Mumford's Definition of a Miracle." *Religious Studies* 39 (4): 465–469.

McKinnon A. 1967. "'Miracles' and 'Paradox.'" *American Philosophical Quarterly* 4 (4): 308–314.

Meller J. 2010. "Naturalny czy nienaturalny początek życia człowieka?" In A. Lemańska and A. Świeżyński (eds.). *Filozoficzne i naukowo-przyrodnicze elementy obrazu świata*, vol. 8. Warszawa: Wydawnictwo UKSW: 191–199.

Miles T. R. 1966. "On Excluding the Supernatural." *Religious Studies* 1 (2): 141–150.

Miller F. P., Vandome A. F. and McBrewster J. (eds.). 2009. *Miracle*. Beau Bassim: Alphascript Publ.

Mooney T. B. and Imbrosciano A. 2005. "The Curious Case of Mr. Locke's Miracles." *International Journal for Philosophy of Religion* 57 (3): 147–168.

Mumford S. 2001. "Miracles: Metaphysics and Modality." *Religious Studies* 37 (2): 191–202.

Murphy N. 1995. "Divine Action and the Natural Order: Buridan's Ass and Schrödinger's Cat." In N. Murphy, A. Peacocke (eds.). *Chaos and Complexity: Scientific Perspectives on Divine Action*. Vatican City State – Berkeley: Vatican Observatory-The Center for Theology and the Natural Sciences: 325–357.

Nowell-Smith P. 1950. "Miracles – The Philosophical Approach." *Hibbert Journal* 48: 354–360.

Otte R. 1996. "Mackie's Treatment of Miracles." *International Journal of Philosophy of Religion* 39 (3): 151–158.

Pratt V. 1968. "The Inexplicable and the Supernatural." *Philosophy* 43: 248–257.

Saler B. 1977. "Supernatural as Western Category." *Ethos. Journal of the Society for Psychological Anthropology* 5 (1): 31–53.

Singh S. 2005. *Big Bang: The Origin of the Universe*. New York: Harper Perennial.

Steinhardt P. J. and Turok N. 2001. "A Cyclic Model of the Universe." *Science* 296 (5572): 1436–1439.

Tałasiewicz M. 2007. "Naturalizm ontologiczny a naturalizm metodologiczny (na marginesie artykułu Marcina Miłkowskiego „Naturalizm ontologiczny a argument Hume'a przeciwko wiarygodności cudów")." *Przegląd Filozoficzno-Literacki* 12 (2): 403–408.

Walker I. 1982. "Miracles and Violations." *International Journal for Philosophy of Religion* 13 (2): 103–108.

Ward K. 2002. "Believing in Miracles." *Zygon. Journal of Religion and Science* 37 (3): 741–750.

Weddle D. L. 2010. *Miracles. Wonder and Meaning in World Religions*. New York: New York University Press.

Williams T. C. 1990. *The Idea of the Miraculous. The Challenge to Science and Religion*. New York: St. Martin's Press.

Young R. 1972. "Miracles and Epistemology." *Religious Studies* 8 (2): 115–126.

Croatian Journal of Philosophy
Vol. XVII, No. 49, 2017

Maximization, Slotean Satisficing, and Theories of Sufficientarian Justice

ALEXANDRU VOLACU*
National University of Political Science and Public Administration (SNSPA Bucharest), Romania

In this paper I seek to assess the responses provided by several theories of sufficientarian justice in cases where individuals hold different conceptions of rationality. Towards this purpose, I build two test cases and study the normative prescriptions which various sufficiency views offer in each of them. I maintain that resource sufficientarianism does not provide a normatively plausible response to the first case, since its distributive prescriptions would violate the principle of personal good and that subjective-threshold welfare sufficientarianism as well as objective-threshold welfare sufficientarianism committed to the headcount claim do not provide normatively plausible responses to the second case, since their distributive prescriptions would violate the principle of equal importance. I then claim that an objective-threshold welfare sufficientarian view committed to prioritarianism under the threshold offers the normatively plausible response to both cases and therefore resists the challenge raised by scenarios that involve differential conceptions of rationality.

Keywords: Maximization, resources, satisficing, sufficientarianism, welfare.

1. *Introduction*

Sufficientarianism holds that distributive justice should primarily be concerned with providing individuals *enough* of some preferred conception of the proper currency of justice. This core idea embodies two central claims, termed by Paula Casal the *positive thesis* and the *negative thesis*, respectively. According to Casal, "the positive thesis stresses the

* I thank Adelin Dumitru, Adrian Miroiu, Tom Parr and two anonymous reviewers for comments on earlier drafts of this paper, as well as an audience at the University of Manchester, where some of the arguments developed here were originally presented.

importance of people living above a certain threshold, free from deprivation. The negative thesis denies the relevance of certain additional distributive requirements" (Casal 2007: 297–298).[1] The view has originally been developed by Frankfurt (1987) as a reaction to the pervasive egalitarian strand of thought characterizing contemporary analytical political philosophy and, independently, by Crisp (2003) as an alternative to both telic egalitarianism and prioritarianism. It has subsequently been extended by a number of authors (Orr 2005, Benbaji 2005, 2006, Casal 2007, Huseby 2010, Shields 2012, Axelsen and Nielsen 2015, 2016), who vary different components of the original theories and provide their own versions of the sufficiency view. In this article, I seek to explore the plausibility of a number of sufficientarian theories in light of their responses to cases in which individuals act on the basis of different conceptions of rationality.[2] There are a number of reasons why examining normative theories in light of such cases is important. First, case-based desirability critiques form a central part of the methodology of analytical political and moral philosophy, as they provide tools with which philosophers can submit theories to "normative tests" (McDermott 2008: 19). Thus, if the cases constructed are useful in illuminating the moral commitments of theories, and in particular, their counterintuitive and morally problematic consequences, they should be taken seriously by philosophers, regardless of their practical likelihood of occurrence. Second, taking such cases into account is useful in illuminating some of the ontological commitments of normative theories as well. In the particular context of sufficientarianism discussed here, the cases will show that some sufficiency views provide adequate responses only when all individuals are satisficers, while others provide adequate responses only when all individuals are maximizers. Third, a wide range of empirical evidence shows that individuals are not actually identical maximizing machines as the *homo economicus* model of neoclassical economics assumes for methodological purposes, but that they are distinctly rational (or, even irrational) on various dimensions. The differential nature of human reasoning should therefore be taken into account when we design normative theories in general, and theories concerning distributive justice in particular.

The paper is structured as follows: in section 2 I describe the constitutive elements of a sufficientarian theory of justice and show how they can be varied in order to obtain a number of different sufficiency views. In section 3 I describe the two conceptions of rationality used in

[1] Some sufficientarians replace the latter with a weaker, *shift thesis*, which only states that "once people have secured enough there is a discontinuity in the rate of change of the marginal weight of our reasons to benefit them further" (Shields 2012: 108).

[2] In particular, I am only concerned here with what Satz and Ferejohn call a "formal and thin conception of rationality" (Satz and Ferejohn 1994: 72), taking into account only the mathematical properties of individual preferences, not their content.

constructing the cases with which the paper is concerned. In section 4 I describe the first case, *Resource plenitude*, and argue that the response which resource sufficientarianism offers to such cases is morally objectionable. In section 5 I describe the second case, *Resource scarcity*, and argue that the responses which subjective-threshold welfare sufficientarianism and objective-threshold welfare sufficientarianism committed to the headcount claim offer to such cases are also morally objectionable. I then argue, in section 6, that objective-threshold welfare sufficientarianism committed to prioritarianism under the threshold offers the morally plausible response in both cases and is impervious to the challenge raised in this paper by weakening the standard assumption of maximizing rationality. Section 7 concludes.

2. *Theoretical background: Sufficientarianism*

Since sufficientarianism is a view of distributive justice, one of the key issues which it needs to address in order to be considered a complete normative theory is to specify a currency of justice, or otherwise stated, to answer the *equality of what* question. In this respect, classical sufficientarian theories (Frankfurt 1987,[3] Crisp 2003) as well as many recent developments (Benbaji 2005, Huseby 2010) standardly take welfare[4] as the currency of justice, while others endorse either resources (Orr 2005) or some conception of capabilities[5] (Anderson 1999, Axelsen and Nielsen 2015, 2016).

While the currency issue concerns all theories of distributive jus-

[3] Frankfurt's preferred currency is actually somewhat more difficult to ascertain, since his discussions on distributions are generally conducted only in terms of money. This has led Temkin (2003: 765) to suggest that Frankfurt is actually attacking a straw man, since egalitarians would agree that it is not simply the inequality of economic resources which we should aim to mitigate. But there are good grounds to claim that he does in fact employ welfare as currency, a position which we may infer from his operationalization of the threshold notion (see below), with economic resources exclusively playing the role of distribuendum of justice (see Gheaus 2016 for the distinction between distribuenda and currencies of justice). The idea that Frankfurt proposes a welfarist version of sufficientarianism is also suggested by Goodin (1987: 45–46), Nathanson (2005: 371) and Huseby (2010: 181).

[4] Following Arneson (2000), throughout this article I use the terms utility, welfare and well-being interchangeably. While the three concepts may not be, strictly speaking, identical under some interpretations, this terminological simplification is required in order to preserve a common language for the family of distributive justice theories with which I am concerned here, since various sufficientarians use all of them to denote the same idea (for instance Frankfurt (1987) uses the term utility, Crisp (2003) uses utility and welfare interchangeably, Huseby (2010) uses welfare and well-being interchangeably and Benbaji (2005) uses all three of them interchangeably).

[5] In this paper I will only be concerned with theories instantiating either welfare or resources as a currency, since the informational framework of the cases in which I am interested in is too parsimonious to adequately capture the demands of capability sufficientarianism. See, however, Arneson (2006) for a powerful criticism of this view.

tice, a complete sufficiency view needs to further address four other questions as well: (1) what the sufficiency threshold is, (2) how the currency is to be distributed below the threshold of sufficiency, (3) how the currency is to be distributed above the threshold of sufficiency and (4) how strict should the priority relation generated by the threshold be. Various sufficientarian theories offer different responses to the first question. Harry Frankfurt sets the sufficiency threshold at the level of *contentment*, understood in the sense that while an individual's marginal utility for gaining economic benefits above the threshold is not nullified, she does not have an *active interest* in obtaining more economic resources. In his own phrasing, "the fact that he is content is quite consistent with his recognizing that his economic circumstances could be improved and that his life might as a consequence become better than it is. But this possibility is not important to him" (Frankfurt 1987: 39). Roger Crisp offers a different answer. To illustrate the idea of a sufficiency threshold, he uses two elements: (1) the notion of impartial spectator and (2) the notion of compassion. According to him, "the spectator puts himself or herself into the shoes of all those affected and is concerned more to the extent that the individual in question is badly off. A spectator who shows no special concern for the badly off has a vice—he or she is uncompassionate" (Crisp 2003: 757). What results from the conjunction of the notion of an impartial spectator with that of compassion is the sufficiency threshold, one which is in his own terms "principled and nonarbitrary" (Crisp 2003: 757), and which is set at the highest level of welfare at which this impartial spectator still feels compassion for the individual in question. Since it is the spectator who evaluates the level of welfare, not each particular individual, the implication of the theory is that the level where compassion disappears is the same for all individuals. Other sufficientarians, such as Huseby (2010) or Benbaji (2005) use multi-level thresholds instead of single-level thresholds, as was the case with those proposed by Frankfurt (1987) and Crisp (2003). Huseby distinguishes between two different sufficientarian threshold levels, a minimal one and a maximal one, with the former being located at the level where basic means to subsistence, or the basic needs of the individual, are satisfied, and the latter being located at the level where the individual can be said to be *content*, understood here as "satisfaction with the overall quality of one's life" (Huseby 2010: 181). In contrast with the two-tiered sufficiency view proposed by Huseby, Benbaji's view recognizes three levels of sufficiency as morally salient: a *personhood* level, a *pain* level and a *luxury* level. The first of these is located just above the level where the life of the respective person is not worth living anymore,[6] the second one just above the level where individuals have negative welfare values

[6] Benbaji avoids the implication that non-human beings would therefore have lives not worth living, by specifying the additional condition that only the life of a being which *falls* below the threshold, after previously being above it would be subjected to the application of this principle (Benbaji 2006: 339).

and the third one is placed at the level where individuals "are so well off at that time that every small benefit to them would be a luxury" (Benbaji 2006: 339–342).

The four sufficiency views described in the previous paragraph give rise to a general and fundamental distinction in the operationalization of welfare sufficientarian thresholds, namely between *subjective* thresholds and *objective* thresholds. Theories that use subjective thresholds maintain that the level of welfare at which the threshold is placed is established by each individual through the means of particular perceptions of her own welfare level. In general, to operationalize this threshold, we assume that there is a point in the welfare functions of individuals where they will say that they are, in some specific sense, satisfied with their current welfare level. Frankfurt's (1987) theory as well as Huseby's view share this feature. In Frankfurt's theory, this point is represented by the level after which individuals would not have an active interest in pursuing the accumulation of further resources as means for welfare enhancements. In Huseby's proposal, this point would be represented by the level at which the individual would consider that he is content with his level of utility. By contrast, theories that use objective thresholds maintain that the welfare level at which the threshold is placed is set by an external source, without any input from the agent subjected to the distributive scheme. Crisp's account of the sufficiency threshold is paradigmatic for this view, since he builds his notion of a threshold in relation to an impartial spectator, who evaluates the distribution and establishes the threshold at the utility level where compassion on the part of the impartial spectator would enter. While Benbaji's account is not so explicit in this regard, the personhood and pain thresholds seem to be non-controversially objective, since decision-making capacities are not subjected to individual perception and a negative level of welfare is described in neutral terms to the perception of the agent subjected to it. Even though we have less information on the conceptualization of the luxury threshold, it seems plausible to also include it in the category of objective thresholds, since otherwise it might be claimed that examples such as the notorious *Beverly Hills* case (see Benbaji 2005: 314–315) would require some form of redistribution, if at least some of the individuals involved would not consider that they are at a luxury level of welfare.

The second distinction between welfare sufficiency views that is important for the purposes of this paper, concerns the second question raised earlier on in this section, i.e. how the currency is to be distributed below the threshold of sufficiency.[7] The main positions regarding distribution below the threshold are to either commit to the *headcount claim,* which states that "we should maximize the number of people who secure enough" (Shields 2012: 103) or to commit to prioritarianism,

[7] While the third and fourth questions raised above are important in their own rights, they have no bearing on the arguments in this article.

which in its canonical formulation states that "benefiting people matters more the worse off these people are" (Parfit 1997: 213). To briefly illustrate the difference between the two views, consider an example where we have two individuals with an identical sufficiency level of 20 units of welfare, in which the first one has in the current state of the world a level of 0 welfare and the second one is at a level of 15 units of welfare and in which we have to decide on how to distribute 5 extra units of welfare. While the sufficientarian committed to the headcount claim would give these 5 extra units to the second person, since it would enable one person to reach the sufficiency threshold, the sufficientarian committed to prioritarianism under the threshold would give more (or even all) units to the first person, since benefiting her has greater moral weight considering that she is worse-off. Both positions are defended by various sufficientarians, with the headcount claim being endorsed by Frankfurt (1987: 31) or Dorsey (2008: 437–438) and prioritarianism under the threshold by Crisp (2003: 758), Huseby (2010: 184) or Shields (2012: 111).

With these distinctions in mind, we can proceed with analysing the plausibility of various sufficiency views in light of the cases described to be in section 4 and 5. However, before advancing to this point it is worthwhile to describe the basic elements of a further distinction which is central to my cases, namely the distinction between maximizing views of rationality and satisficing views of rationality. I take up this task in the next section.

3. *Maximization and Slotean satisficing*

While discussions on the concepts of *maximization* and *satisficing* have occupied a significant place in economics ever since Simon's suggestion of the latter idea (Simon 1947),[8] in political and moral philosophy, the distinction between *maximizing* and *satisficing* types of rationality is usually traced back to Slote's (1984) restatement of the idea of satisficing as a permissible operationalization of act-consequentialism. The original development of the idea that people might act in a satisficing rather than maximizing manner was part of the wider project undertook by Simon to "replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist" (Simon 1955: 99). Briefly, we can state that while maximization entails the three-step sequence: (1) enumerate all the options on offer, (2) evaluate each, (3) choose the best option, a satisficing behaviour follows the sequence: (1) set an aspiration level such that any option which reaches or surpasses it is good enough, (2) begin to enumerate and evaluate the options on offer, (3) choose the first option

[8] Even though it was not introduced under this specific label.

which, given the aspiration level, is good enough (Pettit 1984: 166–167). Slote's conception of satisficing departs from Simon's, however, in an essential way. As he argues, "the sort of satisficing involved [in his own theory] is not (merely) the kind familiar in the economics literature where an individual seeks something other than optimum results, but a kind of satisficing that actually rejects the available better for the available good enough" (Slote 1984: 148). This proposal lies in stark contrast to the classical understanding of satisficing, where the individual appeals to it in order to "reduce the informational and computational requirements of rational choice" (Byron 1998: 71), but given two options which differ from the perspective of the utility produced, would always choose the better one.[9, 10] Phrased in this way, it is not immediately clear whether Slote's notion of satisficing can make sense as a rational strategy, since it would explicitly reject a better alternative in favour of a worse one. In order to yield some intuitive plausibility to the notion, Slote appeals to a number of examples.

In the first such example, you are asked to imagine that you are at work and have just finished eating lunch. You return to your desk and realize that there is a candy bar or a coke in the refrigerator which is placed right next to you. While you are no longer hungry or thirsty, you are not satiated to the point where consumption of the candy bar or coke would not give you additional pleasure. However, you still choose not to consume them (Slote 1984: 143–144). In the second example, we are asked to think of a fairy-tale hero who is given the opportunity to have a single wish granted and does not choose a big pot of gold or a million dollars, but just enough to enable him and his family to live a comfortable life (Slote 1984: 147). In the third example, we are asked to imagine a situation where a family's car breaks down in the middle of the night next to a hotel. The family is quite poor so they cannot afford to rent a room or purchase a meal. Given these conditions, the hotel manager arranges for them to be accommodated, free of charge, in one of the vacant rooms, although not the most expensive one, and receive a meal, also free of charge, from the evening's left-overs, although not the best meal available (Slote 1984: 149–150). The strand that ties together all these examples is the fact that the agent responsible for making a choice had a set of alternatives available before him and de-

[9] Slote himself admits that this is the position of Simon, and further states that the "idea of rational satisficing implies only that individuals or firms do not always seek to optimize and are satisfied with attaining a certain 'aspiration level' less than the best that might be envisaged. It does not imply that it could be rational actually to reject the better for the good enough in situations where both were available" (Slote 1984: 145).

[10] The reason why I do not take into account classical satisficing, but rather the Slotean version, is precisely the fact that Simon's individual would satisfice due to time or informational constraints and such issues do not usually bear much weight in normative theories. Simonian satisficing is thus unlikely to provide the groundwork for any interesting analysis of sufficientarianism.

liberately chose a sub-maximizing one. Slote claims that in each one of the cases discussed however, the strategy of choosing less than the best can be construed as rational in any common-sense interpretation. The primary reason which Slote offers is that individuals might sometimes exhibit a form of moderation which precludes them from taking more benefits rather than fewer.[11] In his own words, "the moderate individual [...] is someone content with (what he considers) a reasonable amount of enjoyment; he wants to be satisfied and up to a certain point he wants more satisfactions rather than fewer, to be better off rather than worse off; but there is a point beyond which he has no desire, and even refuses, to go" (Slote 1986: 60).

It is, of course, not clear whether the examples provided by Slote couldn't be otherwise grounded by various rational (in the classical sense) reasons, thereby making his claim about the non-maximizing character of his proposal collapse. This idea is suggested by Pettit (1984), who discusses the examples offered above and, while agreeing with Slote that they are instances of satisficing behaviour, he adds that they can be construed as rational precisely because of other considerations which the agent weighed in her decision-making process. Only if no such reasons are brought into the picture, the unmotivated sub-maximization which results is in Pettit's terms irrational. In deference to the possibility that individuals satisfice for the sake of a more sophisticated brand of maximization, Slote proposes a distinction between two types of satisficing, namely *instrumental* satisficing on the one hand and non-instrumental or *intrinsical* satisficing on the other (Slote 2004: 14). The instrumental view of satisficing holds that an individual might deliberately choose an inferior alternative only when this course of action would lead to an overall maximization of welfare. In various forms (see for instance Schmidtz 2004 or Narveson 2004), the plausibility of this general view encounters no major resistance amongst political and moral philosophers. The non-instrumental view of satisficing however, which Slote himself admits has been "decidedly the minority view on the rationality involved in satisficing" (Slote 2004: 14), claims that limiting consumption of goods before reaching the point where the marginal utility experienced is null has intrinsical value. The plausibility of this idea is much more controversial and no common ground is reached in this respect.[12] Since the cases which I build in the following section do not rest on a particular view of Slotean satisficing I will not provide a defence of the intrinsical conception, but rather interpret the idea of Slotean satisficing in accordance with the instrumental conception. If the intrinsical conception would turn out to

[11] See, however, Schmidtz (2004: 32) for a disentanglement of the ideas of moderation and satisficing, which Slote often uses interchangeably (Slote 1984: 147, Slote 1986: 65, Slote 2004: 16).

[12] For a wider view on the debate between critics and defenders of satisficing views in moral and political philosophy see Byron (2004).

be plausible, the arguments developed would analogously apply to that interpretation as well.

What is important to note, however, is that the notion of Slotean satisficing is not equivalent to other notions used to build subjective thresholds in some sufficientarian views, such as Frankfurt's account of contentment. As noted in section 2, the idea of contentment proposed by Frankfurt does not imply that the individual who reaches her subjectively-set threshold cannot gain any further welfare above this level. Instead, since in Frankfurt's view "the use of the notion of 'enough' pertains to *meeting a standard* rather than to *reaching a limit*" (Frankfurt 1987: 37, original emphasis), it is entirely plausible to claim that given two options, one of which is right at the threshold of contentment and the other one somewhat above it, the individual in question would choose the latter over the former, due to the higher output of utility, even though she would be in one sense satisfied with both. But the idea of satisficing, as used by Slote, has different implications. While additional resources would still yield an improvement in those aspects of welfare derived from the material consumption of goods, it would not lead to an *all things considered* increase in welfare due to the fact that it would cause counterbalancing disutility in other areas associated with welfare, such as individual attitudes towards moderation. Otherwise, the idea that someone could choose the *available good enough* over the *available better* would be conceptually inconsistent. This notion of satisficing, which is stronger than Frankfurt's idea of contentment, will be used in the subsequent sections.

4. *Resource sufficientarianism and violations of the principle of personal good*

Consider the following case:

> *Resource plenitude.* In a society composed of three individuals (*Alice*, *Brian* and *Charlie*), there are 60 resources available for distribution. Each unit of resource consumed yields exactly one unit of utility for every individual and none of them are satiated at any point. Alice is a satisficer, with her aspiration level set at 30 units of utility, Brian is a satisficer, with his aspiration level set at 10 units of utility and Charlie is a maximizer.

Consider now that, irrespective of the procedure used, the resource sufficientarian,[13] who claims that what is important from the point of view of justice is that *enough* resources are distributed to each individual, has established that the sufficiency threshold is at 20 units of resources. Fortunately, from the resource sufficientarian's point of

[13] I take Orr's (2005) view to be the standard version of resource sufficientarianism. While Orr does not provide answers to a number of questions which a complete sufficientarian theory should standardly address, the endorsement of resources as a currency of sufficientarianism is enough for my present purposes.

view, there are just enough resources to be distributed so that everyone reaches the threshold proposed, thus 20 resources will be distributed to each individual. Call this distribution D1 [Alice – 20; Brian – 20; Charlie – 20]. This distribution of resources will in turn be converted into 20 units of utility for Alice, 20 units of utility for Charlie and 10 units of utility for Brian, considering that he does not gain any extra benefits from resources above the amount of 10.[14] But consider the alternative distributions D2 [Alice – 25, Brian – 10; Charlie – 25] and D3 [Alice – 30; Brian – 10; Charlie – 20], which would map into either 25 units of utility for Alice and Charlie and 10 for Brian (in D2) or 30 units of utility for Alice, 20 units of utility for Charlie and 10 for Brian (in D3). Both D2 and D3 yield more aggregate utility than D1 without making the situation worse-off for anyone. Still, the resource sufficientarian is bound to claim that D1 is, at least *in one way*, better than D2 and better than D3, since D1 is the only distribution where everyone reaches the threshold of sufficiency. Thus, resource sufficientarianism violates what Broome has called the *principle of personal good*, which states that "if we take two distributions that have the same population, and if one of them is better than the other for someone, and at least as good as the other for everyone, then it is better"[15] (Broome 2004: 58). If we take this principle seriously, as many political and moral philosophers do (e.g. Broome 1991, Broome 2004, Vallentyne 1993, Tungodden 2003), we have a strong reason to object to resource sufficientarianism. Furthermore, not only is this view clashing with the principle of personal good, but it is also committed to benefit destruction, since it prescribes wasting 10 resources, which in an alternative distribution could have otherwise benefited either Alice or Charlie. In addition, the two problems raised here are proportionally amplified when: (1) the difference between the resource threshold set by the theory and the aspiration levels which are below the threshold increases and (2) the number of individuals with aspiration levels below the threshold increases.

One possible objection to the idea that resource sufficientarianism might be committed to violations of the principle of personal good and to benefit destruction is that the example proposed is simply implausible, since the aspiration level of an individual would not be positioned below the resource threshold. In the absence of any particular specification of a resource threshold in the sufficientarian literature, it is difficult to reply to this objection in a very concrete manner. However, one general response is that for resource sufficientarianism to gain any moral plausibility, the threshold cannot be located at very low levels, since at such

[14] The alternative would be that Brian's utility actually decreases when further receiving resources. I do not take this stronger case into consideration here, since the weaker case suffices for making the intended point.

[15] This can also be interpreted as a strong form of the Pareto Principle. Tungodden remarks that while the two are structurally identical, the principle of personal good is "stated in the space of individual good or well-being and not in the space of individual preferences" (Tungodden 2003: 8).

levels the negative thesis would no longer appear attractive. Consider that such a low threshold would be the level where individuals would have only very basic access to food, water, clean air and so forth, so that they can survive on a day-to-day basis. It would be, I think, correct to claim that no aspiration level can be found lower than this threshold. If such a threshold was in place, however, it would also mean that justice should not be concerned with the difference in resources between someone who has enough to barely survive for another day and a billionaire like Bill Gates, a position which intuitively appears to be radically implausible. Defending a multi-level version of resource sufficientarianism might partially mitigate this problem, in that the lowest threshold might be placed at a level below which no aspiration level would reasonably be located. But introducing a higher threshold, which is required in order to retain the attractiveness of the negative thesis, opens up the real possibility that the aspiration level of some individuals falls under this threshold, for reasons which have to do with attitudes towards moderation, religious attitudes etc. If we take case-implication critiques (Sen 1979: 197) seriously, then this possibility grounds a plausible objection against resource sufficientarianism.

5. *Welfare sufficientarianism and violations of the principle of equal importance*

Now consider a second case:

> *Resource scarcity*. In a society composed of three individuals (*Alice*, *Brian* and *Charlie*), there are 40 resources available for distribution. Each unit of resource consumed yields exactly one unit of utility for every individual and none of them are satiated at any point. Alice is a satisficer, with her aspiration level set at 30 units of utility, Brian is a satisficer, with his aspiration level set at 10 units of utility and Charlie is a maximizer.

Let us first consider the response which a subjective-threshold welfare sufficientarian, such as Frankfurt, would give to this case. Since Alice and Brian have aspiration levels set at 30 and 10, respectively, consider these levels as their subjective thresholds.[16] Further, according to Frankfurt, what is important from the point of view of justice in cases of resource shortages is that the incidence of sufficiency is maximized. The two positions converge to yield a precise distribution in this case, which is: D4 [Alice – 30; Brian – 10; Charlie – 0]. This distribution is the only one which maximizes the incidence of sufficiency, understood in a subjective sense, since it is the only one in which two of the three individuals have reached the threshold. Since Charlie has no threshold of contentment, he will receive no resources. Furthermore, if a windfall should occur, yielding 20 more resources for distribution (thereby

[16] Noting that they are not only levels of contentment, in Frankfurt's sense, but the stronger types of aspiration levels implied by Slote's conception of satisficing.

transforming this case into *Resource plenitude*), a subjective-threshold welfare sufficientarian is bound to say that we should be indifferent between giving any amount of resources to Alice, Brian or Charlie, even though Charlie is in a position where he has no resources at all.[17]

Secondly, consider the response which a specific type of objective-threshold welfare sufficientarian, namely one who is committed to the headcount claim would provide to the case at hand. Since the example is one of resource scarcity, we will assume that not all individuals can be raised to the threshold with the resources available. Consider therefore that the objective welfare threshold is set at 20. The type of sufficientarianism in which we are interested here would then prescribe distribution D5 [Alice – 20; Brian – 0; Charlie – 20]. The reason why this is the case is that D5 is the only distribution in which two of the three individuals reach the objectively established threshold. Furthermore, if a windfall should occur, yielding 20 more resources for distribution (once again, transforming this case into *Resource plenitude*), the objective-threshold welfare sufficientarian committed to the headcount claim would state that we should be indifferent between providing any amount of resources for Alice, Brian or Charlie, since any amount of resources which we provide Brian with will not be enough for him to reach the welfare threshold, although he is in a position where he has no resources at all.[18]

[17] As one anonymous reviewer has pointed out, it might be objected that subjective-threshold welfare sufficientarianism would not necessarily entail a distribution of 0 resources for Charlie—due to the fact that he is a maximizer—but that we could instead impute an average satisficing level and set that as a distributive threshold for him. This objection is unsuccessful, however, in cases involving sufficientarian views of this type, precisely because the subjectivist manner of deriving distributive thresholds precludes attaching externally built features to it. As subjective thresholds appeal only to the preferences of the individual in question, imputing the average satisficing level (or any other form of externally produced level) amounts to a collapse into objective-threshold welfare sufficientarianism, a separate view from that which was scrutinized in this paragraph (and which will be subsequently examined).

[18] It may be worth questioning if the unappealing prescriptions offered by subjective-threshold sufficientarianism might not draw their force simply from the fact that Frankfurt's version (which I used as a standard operationalization of this type of sufficiency view) is itself committed to the headcount claim as well. If this is correct, than Frankfurt's own sufficientarian position might seem less plausible in light of the example, but other subjective-threshold sufficientarian views might be unaffected. My reply to this argument is that even if the headcount claim is dropped from subjective-threshold sufficientarianism, the view simply cannot accommodate individuals who do not have contentment levels regarding the distribution of resources (this is perhaps most vivid in the case which Frankfurt himself discusses, that of monetary resources). If a person does not have a contentment level (at least in the weaker sense proposed by Frankfurt), then prioritarian or other types of arrangements for distributions under the threshold simply cannot count her in the distribution, at least while there are still other individuals that might reach their thresholds. The subjective-threshold view is therefore committed at a much deeper level than any other sufficientarian view examined here to make homogeneous

What do these two responses, derived from different normative principles, have in common? In both cases, one person appears to be significantly disadvantaged by the distribution, since she is up to a point entitled to no resources whatsoever and only after a certain point (where all others have reached the sufficiency thresholds) she has claims which are on par with the other subjects of the distribution, but not more pressing, even when she is at a miserable level of welfare and the others are at a blissful level. This result appears to be deeply at odds with some basic moral claims. To illustrate this, consider for instance Dworkin's *principle of equal importance*, according to which government should "adopt laws and policies that insure that its citizens fate are [...] insensitive to who they otherwise are—their economic backgrounds, gender, race, or particular sets of skills and handicaps" (Dworkin 2000: 6). I take this ethical principle to be relatively uncontroversial, since it expresses a more generic impartiality condition which has been a staple of the literature on distributive justice within the past decades[19]. If we take the principle of equal importance seriously, then the distributions prescribed by both welfarist subjective-threshold sufficientarians and objective-threshold sufficientarians committed to the headcount claim appear to be problematic as they assign *unequal* importance to individuals based on an internal characteristic, namely the type of rationality that they hold, which is morally arbitrary.[20] The unequal treatment of maximizing individuals in subjective-threshold welfare sufficientarianism and the unequal treatment of satisficing individuals who cannot reach the aspiration levels set in objective-threshold welfare sufficientarianism committed to the headcount claim, therefore count as serious objections against them.

assumptions regarding the rationality of individuals, since the existence of maximizers not only renders this view morally implausible but it raises a challenge to the coherence of the view as a whole.

[19] See however the critical position adopted by Steinhoff (2014) against the ideas of equal concern and respect, which implicitly encompasses a criticism of the Dworkinian principle of equal importance.

[20] I do not mean to suggest that some forms of satisficing would not perhaps be desirable in some cases. But I maintain that there are at least two arguments in defence of the claim that satisficient individuals warrant no special priority in the distribution of resources. The first one concerns the possibility that satisficers are in fact not always moderate, since moderation is not necessarily connected to satisficing (as Schmidtz 2004 shows). If the aspiration level of an individual would be so high that reaching it would require a drainage of resources which could otherwise be distributed to maximizers in order that they reach a decent level of welfare, than it seems clear that we should not give any sort of priority to the satisficing individual. Further, if we consider satisficing and maximization as actual behavioral features (and not simply useful assumptions for theory-building), it is questionable to what extent they are traceable to individual choices and it would seem more likely that they are not. Therefore, it would be highly controversial to punish or reward individuals for being endowed with a trait for the formation of which they can claim no responsibility.

6. *A resilient competitor: Crisp's sufficientarian view*

Let us now examine how a distinct version of sufficientarianism, i.e. one which proposes an objective welfare threshold but is at the same time not committed to the headcount claim, would respond to cases such as *Resource plenitude* and *Resource scarcity*. Since Crisp's (2003) formulation of the sufficiency view meets both demands,[21] I will take his version as the standard-bearer for this type of sufficientarianism. What would such a view entail in the case of *Resource plenitude*? Assume, again, that the objective threshold is set at 20. First, since the view endorses prioritarianism below the threshold, all other things being equal, it would sequentially raise each individual with one unit of welfare until all of them reach the level 10. Up to this point, 30 resources have been distributed, so 30 units remain. Since Brian no longer gains any further utility after having 10 resources, the next 20 units would be distributed sequentially to Alice and Charlie, until both of them reach the sufficiency threshold. Finally, the remaining 10 resources are distributed between Alice and Charlie, since no further distribution towards Brian would manage to raise him over the threshold. If we follow Crisp's (2003: 758) suggestion that utilitarianism would be a plausible pattern for distribution over the threshold, all possible distributions of the last 10 resources to Alice and Charlie are equally preferable. The distribution prescribed by Crisp's sufficientarian view would therefore be either D2 [Alice – 25, Brian – 10; Charlie – 25], D3 [Alice – 30; Brian – 10; Charlie – 20] or some other version which distributes between 20 and 30 resources for Alice and Charlie and 10 resources for Brian. Thus, Crisp's view avoids violating the principle of personal good, since it considers that both D2 and D3 are preferable to D1, and at the same avoids destroying benefits, since it does not give more resources to Brian than he can convert into welfare.

Let us now see how Crisp's sufficientarianism fares in the *Resource scarcity* case. It once again begins by sequentially distributing one resource to each of the three individuals until all of them reach a level of 10 welfare. Since Brian no longer derives any utility from receiving more resources, the final 10 resources to be distributed are then equally allocated to Alice and Charlie, resulting in D6 [Alice – 15; Brian – 10; Charlie – 15]. This is because the threshold is set too high for all individuals to reach it and below the threshold, inequalities are to be arranged in a prioritarian manner. Thus, since we attach more weight to the distributive claims of individuals the lower their welfare levels are, we cannot proceed with distributing one more unit to an individual who is better-off, while there is still one individual who is worse-off and could be made better-off. This grounds both our reasons to distribute an equal amount of resources to all individuals until they reach level 10

[21] As it prescribes an objective threshold at the level where an impartial spectator would no longer feel compassion for the individual in question and it prescribes a prioritarian distribution below the threshold.

and our reasons not to distribute any more resources to Brian after this level, since he can no longer be made better-off. A theory which claims that D6 should be enacted instead of either D4 [Alice – 30; Brian – 10; Charlie – 0] or D5 [Alice – 20; Brian – 0; Charlie – 20] in the case of *Resource scarcity* has great intuitive appeal since it avoids violating the principle of equal importance. It does not punish or otherwise mistreat either Charlie (who in D4 would have received nothing) for being a maximizer, or Brian (who in D5 would have received nothing) for being a satisficer. Taking this into consideration, an objective-threshold welfare sufficientarian theory which is committed to prioritarianism below the threshold (of which Crisp's view would be a classical example), is in a position to provide a better reply to cases such as *Resource plenitude* and *Resource scarcity* than resource sufficientarianism, subjective-threshold welfare sufficientarianism or objective-threshold welfare sufficientarianism committed to the headcount claim are able to do.

7. *Conclusions*

The sufficiency view has drawn a considerable amount of attention in the literature on distributive justice in the past two decades, albeit much less than firmly established rivals such as the egalitarian and, more recently, prioritarian views. In this paper, I sought to open a new line of criticism as well as comparison between sufficiency views, which has been until this point unexplored, namely what sort of responses will sufficientarian theories offer to cases where individuals act on the basis of different conceptions of rationality. In order to construct a plausible view of the way in which individuals might be differentially rational, I appealed to the classical notion of a maximizing behavior on the one hand and the notion of Slotean satisficing on the other. I then assessed the responses provided by four different types of sufficiency views in cases based on these different accounts of rationality. The conclusions drawn in this article support objective-threshold welfare sufficientarianism committed to a prioritarian distribution under the threshold, the classical version of such a theory being that of Crisp, which I claim responds correctly to both cases presented. By contrast, I argue that resource sufficientarianism offers the wrong response to cases such as *Resource plenitude*, since it violates the principle of personal good, while allowing for benefits to be wasted rather than distributed, and both subjective-threshold welfare sufficientarianism and objective-threshold welfare sufficientarianism committed to the headcount claim offer the wrong response to cases such as *Resource scarcity*, since they violate, in opposite fashions, the principle of equal importance. It is, of course, possible to either object to these conclusions, by claiming that the principle of personal good or the principle of equal importance are simply not morally salient, or that there may be other implications of objective-threshold welfare sufficientarianism committed to prioritarianism below the threshold that might prove, on balance, more

problematic.[22] It is also possible to reduce the force of my objections by accommodating them within the framework of the criticized views through an appeal to value pluralism, in order to avoid violations of the above mentioned principles. Regardless, the article still provides a strong reason[23] in favour of Crisp's (and similarly constructed) version of sufficientarianism against other types of sufficiency views, e.g. those of Frankfurt and Orr, as they presently stand.

References

Anderson, E. 1999. "What is the Point of Equality?" *Ethics* 109 (2): 287–337.

Arneson, R. 1989. "Equality and Equal Opportunity for Welfare." *Philosophical Studies* 56 (1): 77–93.

Arneson, R. 2000. "Welfare Should be the Currency of Justice." *Canadian Journal of Philosophy* 30 (4): 497–524.

Arneson, R. 2006. "Distributive Justice and Basic Capability Equality: 'Good Enough' Is Not Good Enough." In Kaufman, A. (ed.). *Capabilities Equality: Basic Issues and Problems*. New York: Routledge.

Axelsen, D. and Nielsen, L. 2015. "Sufficiency as Freedom from Duress." *Journal of Political Philosophy* 23 (4): 406–426.

Axelsen, D., Nielsen, L. 2016. "Capabilitarian Sufficiency: Capabilities and Social Justice." *Journal of Human Development and Capabilities* published online DOI:10.1080/19452829.2016.1145632.

Benbaji, Y. 2005. "The Doctrine of Sufficiency: A Defence." *Utilitas* 17 (3): 310–332.

[22] For instance, an anonymous reviewer has argued that Frankfurt's (1987: 30) original case, which motivated his dissatisfaction with egalitarianism, constitutes a serious challenge to Crisp's version of sufficientarianism. The case can be summed up as follows: A population consisting of 10 individuals lives in a situation of extreme poverty. There are 40 units of a certain resource (e.g. food or medicine) available for distribution amongst the population. Any individual that does not get at least 5 units of the respective resource will die. At first sight it might appear that a commitment of prioritarianism below the threshold would require that we distribute 4 resources to each individual, which would ultimately result in everyone dying. This implication, might be claimed, is more damaging for Crisp's view than the counterexamples offered here against alternative views. But a closer examination of what exactly a commitment to prioritarianism under the threshold would entail precludes such a conclusion. Since in any of its standard versions prioritarianism is welfarist in respect to the currency of justice, it would not endorse a distribution whereby at best, no utility is to be gained by anyone (if we consider that the utility output for death is 0) or at worst generates disutility for everyone (if we consider that the utility output for death is negative). Thus, much like Casal's (2007: 307–308) rejection of Frankfurt's claim that egalitarianism would imply an equal distribution of resources in this case, we can also reject the claim that prioritarianism below the threshold would imply an equal distribution of resources, since the state of affairs reached would be worse for at least some individuals without benefitting anyone (a violation of the principle of personal good which prioritarianism *cannot* conceptually make). Instead, prioritarianism under the threshold requires that in this case six individuals (presumably selected after a fair procedure) be brought to the threshold of survival, identically to Frankfurt's sufficientarian view.

[23] Even if not necessarily decisive.

Benbaji, Y. 2006. "Sufficiency or Priority?" *European Journal of Philosophy* 14 (3): 327–348.

Broome, J. 1991. *Weighing goods*. Oxford: Blackwell.

Broome, J. 2004. *Weighing lives*. Oxford: Oxford University Press.

Byron, M. 1998. "Satisficing and Optimality." *Ethics* 109 (1): 67–93.

Byron, M. (ed.) 2004. *Satisficing and Maximizing: Moral Theorists on Practical Reason*. Cambridge: Cambridge University Press.

Casal, P. 2007. "Why Sufficiency Is Not Enough." *Ethics* 117 (2): 296–326.

Cohen, G.A. 1989. "On the Currency of Egalitarian Justice." *Ethics* 99 (4): 906–944.

Crisp, R. 2003. "Equality, Priority, and Compassion." *Ethics* 113 (4): 745–763.

Dorsey, D. 2008. "Toward a theory of the basic minimum." *Politics, Philosophy & Economics* 7 (4): 423–445.

Dworkin, R. 2000. *Sovereign Virtue: The Theory and Practice of Equality*, Cambridge: Harvard University Press.

Frankfurt, H. 1987. "Equality as a Moral Ideal." *Ethics* 98 (1): 21–43.

Gheaus, A. (2016), "Hikers in flip-flops. Luck egalitarianism, democratic equality and the *distribuenda* of justice." *Journal of Applied Philosophy* EarlyView DOI: 10.1111/japp.12198.

Goodin, R. 1987. "Egalitarianism, Fetishistic and Otherwise." *Ethics* 98 (1): 44–49.

Huseby, R. 2010. "Sufficiency: Restated and Defended." *Journal of Political Philosophy* 18 (2): 178–197.

McDermott, D. 2008. "Analytical Political Philosophy." In Leopold, D. and Stears, M. (eds.). *Political Theory: Methods and Approaches*. Oxford: Oxford University Press.

Narveson, J. 2004. "Maxificing: Life on a Budget; or, If You Would Maximize, Then Satisfice!" In Byron, M. (ed.). *Satisficing and Maximizing: Moral Theorists on Practical Reason*. Cambridge: Cambridge University Press.

Nathanson, S. 2005. "Equality, Sufficiency, Decency: Three Criteria of Economic Justice." In Adams, F. (ed.). *Ethical Issues for the Twenty-First Century*. Charlottesville: Philosophy Documentation Center.

Orr, S. 2005. "Sufficiency of Resources and Political Morality." Presented at the *Priority in Practice seminars* held on 22–23.09.2005, University College London.

Parfit, D. 1997. "Equality and Priority." *Ratio (New Series)* 10 (3): 202–221.

Pettit, P. 1984. "Satisficing Consequentialism." *Proceedings of the Aristotelian Society, Supplementary Volumes* 58: 165–176.

Satz, D. and Ferejohn, J. 1994. "Rational Choice and Social Theory." *Journal of Philosophy* 91 (2): 71–87.

Schmidtz, D. 2004. "Satisficing as a Humanly Rational Strategy." In Byron, M. (ed.). *Satisficing and Maximizing: Moral Theorists on Practical Reason*. Cambridge: Cambridge University Press.

Sen, A. 1979. "Equality of What?" *The Tanner Lectures on Human Values*. Stanford University.

Shields, L. 2012. "The Prospects for Sufficientarianism." *Utilitas* 24 (1): 101–117.

Simon, H. 1947 [1997]. *Administrative Behavior*. New York: The Free Press.

Simon, H. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (1): 99–118.

Slote, M. 1984. "Satisficing Consequentialism." *Proceedings of the Aristotelian Society, Supplementary Volumes* 58: 139–163.

Slote, M. 1986. "Moderation, Rationality and Virtue." In McMurrin, S. (ed.). *The Tanner Lectures on Human Values*. Salt Lake City: University of Utah Press.

Slote, M. 2004. "Two views of satisficing." In Byron, M. (ed.). *Satisficing and Maximizing: Moral Theorists on Practical Reason*. Cambridge: Cambridge University Press.

Steinhoff, U. 2014. "Against Equal Respect and Concern, Equal Rights, and Egalitarian Impartiality." In Steinhoff, U. (ed.). *Do All Persons Have Equal Moral Worth? On 'Basic Equality' and Equal Respect and Concern*. Oxford: Oxford University Press.

Temkin, L. 2003. "Egalitarianism Defended." *Ethics* 113 (4): 764–782.

Tungodden, B. 2003. "The value of equality." *Economics & Philosophy* 19 (1): 1–44.

Vallentyne, P. 1993 "The connection between prudential and moral goodness." *Journal of Social Philosophy* 24 (2): 105–128.

Croatian Journal of Philosophy
Vol. XVII, No. 49, 2017

Self-deception and Selectivity: Reply to Jurjako

JOSÉ LUIS BERMÚDEZ
Texas A&M University, College Station, Texas, USA

Marko Jurjako's article "Self-deception and the selectivity problem" (Jurjako 2013) offers a very interesting discussion of intentionalist approaches to self-deception and in particular the selectivity objection to anti-intentionalism raised in Bermúdez 1997 and 2000. This note responds to Jurjako's claim that intentionalist models of self-deception face their own version of the selectivity problem, offering an account of how intentions are formed that can explain the selectivity of self-deception, even in the "common or garden" cases that Jurjako emphasizes.

Keywords: Action explanation, dispositions, epistemic virtue, self-deception, the selectivity problem.

I originally proposed the selectivity problem in Bermúdez 1997, 1999, and 2000 as an argument for intentionalist, as opposed to anti-intentionalist or deflationary, approaches to self-deception. Intentionalists claim that intrapersonal self-deception effectively mirrors interpersonal deception. In both cases the (self-) deceiver intentionally brings it about that the (self-)deceived person acquires a belief, or other propositional attitude. Just as the interpersonal deceiver intends to bring it about that his victim acquires a particular belief, so to does the intrapersonal self-deceiver intend to bring it about that he himself acquire a particular belief.

In opposition to intentionalism, anti-intentionalists such as Al Mele argue that self-deceiving belief acquisition can be explained solely in terms of motivational bias and similar mechanisms, without assuming any intention to acquire a belief (Mele 1997, 2001, 2012). In his 1997 account, for example, Mele proposes the following four jointly sufficient conditions for S to acquire a belief through self-deception.

1) The belief that *p* acquires is false
2) S treats data seemingly relevant to the truth of *p* in a motivationally biased way.

3) This biased treatment non-deviantly causes S to come to believe that *p*
4) The evidence that S possesses provides greater warrant for ~*p* than for *p*.

The selectivity problem is directed in particular at components (2) and (3) of this account. My objection is that the anti-intentionalist does not have the resources to explain why motivational bias should be brought to bear in some cases and not in others:

> Self-deception is paradigmatically selective. Any explanation of a given instance of self-deception will need to explain why motivational bias occurred in *that* particular situation. But the desire that *p* should be the case is insufficient to motivate cognitive bias in favor of the belief that *p*. There are all sorts of situations in which, however strongly we desire it to be the case that *p*, we are not in any way biased in favor of the belief that *p*. How are we to distinguish those from situations in which our desires result in motivational bias? I will call this the selectivity problem (Bermúdez 2000: 317)

Only intentionalist models of self-deception can solve the selectivity problem, I claim. In order for a self-deceiver to come to believe that *p* there must be not simply a desire that *p* be the case, coupled with various biased mechanisms of belief formation, but also an intention to believe that *p*.

Jurjako raises the very interesting objection that intentionalist models face their own version of the selectivity problem. He starts with the plausible assumption that intentions are formed for reasons, typically beliefs and desires.

> So, in order to explain why in this particular instance self-deception occurred, we need to invoke a desire and a belief. But now we can ask why in this particular situation a desire that *p* be the case caused an intention to believe that *p* is the case? As Bermúdez noted, we have all kinds of desires that, nevertheless how strong, do not cause us to believe that *p* is the case; similarly we can say that we have different strong desires to believe that *p* be the case (or that we believe that *p* is the case), that nevertheless do not cause an intention to believe that *p*. So in this way we can raise the selectivity problem against the intentionalist account. Namely, we can raise the question why in this *particular* situation the desire that *p* be the case (or to believe that *p*) caused an intention to believe that *p* is the case since, according to Bermúdez, in all kinds of situations, no matter how strongly we desire that *p* be the case it does not cause us to believe that *p* is the case. (Jurjako 2013: 155)

Jurjako proposes two options that an intentionalist can take to resolve this new version of the selectivity problem. The first option is to assume that self-deceptive intentions emerge "by sheer chance" from the reasons that precede the intention.[1] The second option is to suppose

[1] Actually, Jurjako refers to "intentions to self-deceive", but intentionalists about self-deception are certainly not committed to holding that a self-deceptive intention is always an intention to deceive oneself. I can (self-deceptively) intend to bring it about that I believe that *p* without intending to deceive myself. For further analysis of how to understand self-deceptive intentions see Bermúdez 2000.

that self-deceptive intentions result from a conscious decision. According to Jurjako, intentionalists are caught on the horns of a dilemma here. The first option is highly implausible and in any case does not provide a satisfying answer to the selectivity problem. The second option, on the other hand, does resolve the selectivity problem, but over-intellectualizes what is going on in self-deception in a way that makes it inapplicable to common or garden varieties of self-deception.

I completely agree with Jurjako that the first option is a non-starter and will say no more about it. I also agree with him that intentions are not determined by standing beliefs and desires. Intentionalist models of self-deception could not possibly work unless forming an intention is in some sense an autonomous mental act. In that respect intentionalist models are committed to something like the commonsense view of the progress from thought to action sketched out by David Wiggins at the beginning of his paper "Weakness of will, commensurability, and the objects of desire" (Wiggins 1978). According to this commonsense view, "we need autonomous and mutually irreducible notions of believing, desiring, deciding *that*, deciding *to*, intending" (Wiggins 1978/9: 244). A similar view of the autonomy of intention is defended in Holton 2009.

Both Holton and Wiggins primarily analyze intentions that result from choice, where choice typically results from a process of deliberative practical reasoning. Again, I agree with Jurjako that it is not helpful to see typical examples of common or garden self-deception as the result of deliberative practical reasoning. But we can escape from the dilemma that he poses for intentionalist approaches to self-deception by recognizing other ways of thinking about how intentions are formed. Deliberative and reflective choice is one end of a spectrum, rather than the only game in town.

As standardly understood, intentions lead straight to action (modulo weakness of will), which is why they bridge the gap between beliefs, desires, and other propositional attitudes, on the one hand, and action on the other. But of course this immediately raises the question of how the gap is bridged between propositional attitudes and intentions. The canonical model, going at least as far back as the Aristotelian practical syllogism, sees intentions as resulting from means-end reasoning about how best to satisfy desires (taking "desire" broadly enough to include what Aristotle would have called the apparent good). But there are some important passages where Aristotle appears to recognize that even as an idealization the deliberative model often fails to apply. Looking at those passages points towards an alternative that helps make better sense of self-deception.

In an illuminating passage in Book VI of the *Nichomachean Ethics* Aristotle discusses the distinctive character of practical wisdom (*phronesis*) and what distinguishes it from intelligence (*nous*). He writes:

> That practical wisdom is not knowledge is evident; for it is, as has been said, concerned with the ultimate particular fact, since the thing to be done is of

> this nature. It is opposed, then, to comprehension; for comprehension is of the definitions, for which no reason can be given, while practical wisdom is concerned with the ultimate particular, which is the object not of knowledge but of perception—not the perception of qualities peculiar to one sense but a perception akin to that by which we perceive that the particular figure before us is a triangle.[2]

The key idea here is that practical wisdom involves perception. How one acts is, in the last analysis, a function of how one *sees* things.

One way of understanding what is going on here emerges when we recall the basic form of the Aristotelian practical syllogism, which contains both a major premise and a minor premise. The major premise is typically portrayed in a way that aligns it with belief. In *De Motu Animalium* Aristotle gives the example: All men ought to walk. The minor premise, though, typically comes across differently. The minor premise is how a general belief is seen to be applicable to *this* particular situation. Here is another important passage from *De Anima*. Aristotle is considering the question (rather strange to modern ears) of whether the faculty of knowing moves or is at rest.

> The faculty of knowing is never moved but remains at rest. Since the one premise or judgment is universal and the other deals with the particular (for the first tells us that such and such a kind of man should do such and such a kind of act, and the second that this is an act of the kind meant, and I a person of the type intended), it is the latter opinion that really originates movement, not the universal; or rather it is both, but the one does so while it remains in a state more like rest, while the other partakes in movement.[3]

The minor premise (dealing with the particular) is, to use the earlier phrase, what bridges the gap between beliefs, desires, and action. It is what allows me to see that the situation I am in is one to which *this general belief* or *this desire* is applicable.

Without getting into the question of how to reconcile Aristotle's various comments about action, choice, and deliberation,[4] it seems to me that there is an important insight in these two passages, pointing towards an alternative way of thinking about how intentions emerge. An intention to act in a certain way can come about because of how I interpret or understand the situation in which I find myself. To use a very non-Aristotelian term, intentions can result from *framing* a situation in a certain way. There are many different types of frame. Some are highly intellectualized. But many are not. Framing a situation can be as simple a matter as identifying which other situation it is most similar to, highlighting one feature over another, or finding an affective

[2] Aristotle, *Nichomachean Ethics* Bk. VI 1142a23–1142a28, translated by W. D. Ross, revised by J. O. Urmson (in J. Barnes, *The Complete Works of Aristotle*, Vol. 1)

[3] Aristotle, *De Anima* Bk. III 433b17–433b21, translated by J. A. Smith (in J. Barnes, *The Complete Works of Aristotle*, Vol. 1).

[4] For helpful discussion and further references see Price 2008.

valence. Framing (and re-framing) a situation in a different way can often open up new possibilities for action.[5]

This way of thinking about intention and choice offers a way out of the dilemma Jurjako poses. The intentions that drive common or garden self-deception do not have to be viewed as emerging either randomly or from conscious acts of deliberative choice. Instead we can see them as emerging from how the self-deceiver frames the situation in which they find themselves. Of course, what is being framed in self-deception is not, as it were, the object of the self-deceiving belief. The spouse determinedly convinced of his spouse's fidelity despite all the evidence to the contrary may well be framing his spouse's behavior in all sorts of ways, but that is not what generates the self-deception (more likely, it is explained by the self-deception). What matters for self-deception is how the self-deceiver frames the situation in which he believes that his spouse is faithful. He might, for example, frame this as an act of trust and loyalty. Having a certain belief is part of the person that he wants to be, and it is because he sees things that way that he intentionally comes to form the self-deceptive belief. Here it seems correct to say both that the intention to form a certain belief is what ultimately explains his self-deception, and that the intention does not emerge from an over-intellectualized process of conscious choice.

References

Bermúdez, J. L. 1997. "Defending intentionalist accounts of self-deception." *Behavioral and Brain Sciences* 20: 107–08.

Bermúdez, J. L. 1999. "Autoinganno, intenzione, e credenze contraddittorie." *Sistemi Intelligenti* 11: 521–32.

Bermúdez, J. L. 2000. "Self-deception, intentions, and contradictory beliefs." *Analysis* 60: 309–19.

Bermúdez, J. L. Forthcoming. *The Power of Frames: New Tools for Rational Thought*. Cambridge: Cambridge University Press.

Holton, R. 2009. *Willing, Wanting, Waiting*. New York: Oxford University Press.

Jurjako, M. 2013. "Self-deception and the selectivity problem." *Balkan Journal of Philosophy* 5: 151–62.

Mele, A. 1997. "Real Self-Deception." *Behavioral and Brain Sciences* 20: 91–102.

Mele, A. 2001. *Self-Deception Unmasked*. Princeton University Press.

Mele, A. 2012. "When Are We Self-Deceived?" *Humana Mente Journal of Philosophical Studies* 20: 1–15.

Price, A. W. 2008. "The practical syllogism in Aristotle: a new interpretation." *Logical Analysis and History of Philosophy* 11: 151–62.

Wiggins, D. 1978. "Weakness of will, commensurability, and the objects of deliberation and desire." *Proceedings of the Aristotelian Society* 79: 251–77.

[5] For a very suggestive view of reasoning and choice from which I have learnt a lot see Schick 1991 and 1997. Schick talks about understandings rather than frames. The role of framing in reasoning is developed in more detail in Bermúdez (Forthcoming).

Croatian Journal of Philosophy
Vol. XVII, No. 49, 2017

Book Discussion

Possible Uses of Tennant's Methodology in Secondary Education

RUDI KOTNIK
Faculty of Arts, University of Maribor, Maribor, Slovenia

The paper addresses the issue whether Tennant's textbook Introducing Philosophy*, a demanding textbook based on the methodology of Analytical philosophy, can be useful for high school teachers not trained in Analytical methodology. The pedagogical background is presented through a conceptual framework of problematization, conceptualisation and argumentation, and I follow Tennant's methodology through these three principles. The issue which I discuss is how Tennant's methodology can help teachers to foster the three analytical abilities in students. I will show how his presentation of topics as content demonstrate his methodology and how particular examples can be used by teachers in secondary education, as well as in introductory university courses in philosophy. If teachers pay attention to this methodology within the content, they can apply it to other topics.*

Keywords: Teaching, content, methodology, argumentation, conceptualisation, problematization, thought-experiment.

1. *Introduction*

In this paper I would like to discuss a specific issue related to a very specific domain of teaching philosophy in secondary education, which could be relevant for general introductory courses at the graduate level as well. Usually philosophy teachers in secondary education (at least in the continental Europe), especially those who were educated in the previous decades, did not have an education in analytical philosophy. Therefore, it is a special challenge to examine and find out how a textbook based on analytical methodology could be helpful and used by these teachers. Tennant's book *Introducing Philosophy* offers this kind

of challenge, because a teacher without knowledge and experience in analytical philosophy can very soon get lost in reading and studying the book. What is the novelty, thus, of this author's methodology, and can it be accessible to teachers and consequently for students?

2. *Philosophy education: the background*

Within theoretical approaches to philosophy education (didactics of philosophy) there are several approaches concerning how to teach philosophy. Despite their varieties, their common ground can be reduced to three basic principles: problematization, conceptualization and argumentation. They can be included in the aims and objectives of teaching of philosophy, in its methodology of teaching, and in assessment criteria. The French author M. Tozzi (2008) talks about three processes in which philosophy happens: problematization, conceptualization and argumentation, which to a certain extent represent a methodology. These activities develop appropriate abilities, or we could say that these abilities form the basis for the activities: i.e., for doing philosophy (Kotnik 2014: 152). These processes, however, are not separated but interwoven and interrelated: "Conceptualisation is an attempt to philosophically clarify the concept, problematization undermines it, and argumentation corroborates the thesis. All three are the aspects of reflection" (Šimenc 2007: 29). This is the conceptual ground of our philosophy education. We are going to follow Tennant's book through these three methodological principles and processes trying to find out to what extent and in what ways this textbook can provide teachers some of the benefits of this book.

3. *Outline of the book through the three principles*

The main division of Tennant's book is between philosophical content and methodology. Although the subtitle *God, Mind, World, and Logic* partly refers to the content as traditional topics, these topics serve as a demonstration of methodology.[1] His emphasis is on providing methodology, as he puts it "groundwork, orientation, and wherewithal: concepts; distinctions; characterization of important '-isms'; and philosophical methodologies such as analysis, explication and thought-experiment" (p. XXI). He says that "it provides a more methodical survey of the basic tools for thinking that the beginning philosopher must acquire" (p. XV). This is elaborated systematically, carefully, and thoughtfully, occupying the first half of the book, which consists of 433 pages. The introductory chapter (Part I) The Nature of Philosophy is followed by the chapter (Part II) Philosophy and Method and continued in two

[1] The book review by Reeve (2015) surprisingly presents Tennant's book as dealing with content.

more chapters (Part III: The Existence of God and Mind and Part IV: Body and External World) presenting two topics which can be read as a demonstration of methodology. The method of philosophy is elaborated through eight subchapters: What is Logic?, Inductive Reasoning, The Method of Conceptual Explication, The Method of Thought-Experiment, Intellectual Creativity and Rigor, Deduction in Mathematics and Science, and The Methodological Issue of Reductionism. Following these sections step by step, we can also recognize principles of argumentation, conceptualization and problematization. While the first two seem more explicit, the last one can be noticed in each section as well. At the end of each section the author invites the reader to think about the questions he raises. In the section Intellectual Creativity and Rigor he explains this and we can understand this as problematization:

> A great philosopher, likewise, is one who can identify concepts and fundamental beliefs of great importance; offer interesting, illuminating analyses of those concepts, or necessary and sufficient conditions for the truth of those beliefs; and construct imaginative counterexamples to defective rival analyses (p. 162).

Since he is addressing teachers as professional philosophers and those who want to become teachers, we ask the question: can teachers in secondary education or in introductory university courses help students who will not be professional philosophers to learn philosophy by means of Tennant's textbook? His highly demanding methodology seems inaccessible for average high school students and even for some of their teachers. Is it, therefore, an impossible task for teachers to use Tennant's textbook in the philosophy class? In the following section I'll try to show the scopes and limits of using this textbook for this purpose. My guidelines will be the above mentioned three principles.

4. *The nature of philosophy through the three principles – emphasis on conceptualization*

Before proceeding to methodology, Tennant's extensive introductory chapter "The nature of Philosophy" explains important concepts and distinctions as well as opposing -isms. This can be understood as a necessary clarification of terms. For my purpose these clarifications have a wider significance. Following Tennant's approach, we can notice that he is already raising problems and that problematization is there from the very beginning, together with argumentation and conceptualisation. By presenting and discussing the major conceptual distinctions (appearance/reality, mind/body, objective/subjective, abstract/concrete, descriptive/normative, empirical/rational, necessary/true/false/impossible, theory/evidence, and in a special section Kant's distinctions a priori/a posteriori, analytic/synthetic), he shows the necessity to introduce new concepts and distinctions by italicizing them and reminding the reader about their importance in philosophical inquiry. These ital-

ics appear throughout whole book and have a significant educational role inviting the reader's mindfulness. It is, therefore, worthwhile to follow his approach carefully to see how he makes these distinctions throughout this section. This way of doing it can be a learning experience in itself. Of course, this refers to the whole textbook.

Let me illustrate this section with the distinction between subjective/objective, which is useful for high school students. The term *subjective* is often used without further explanation or justification and students are happy with that. Tennant draws the readers' attention "to make clear the exact sense in which one is intending the notions of 'subjective' and the 'objective' to be understood, in the context at hand" (p. 44). For this purpose, he offers five contrasts, between secondary and primary qualities, perspective limitation and group consensus, probability and objective chances, projections onto the world and properties of agents (in ethics), first person perspective and shared experience (p. 44–45).

The pedagogical significance of this approach is not only in offering further distinctions to clarify particular concept and/or distinction but also in learning *a new philosophical habit*, attitude not taken concepts for granted and being mindful for them, which is one of the beginner's way to practice conceptualisation as well as problematization. As in all other sections or chapters he ends the section with dilemmas and questions which, again, is an example of problematization.

The section *Important Opposing -isms* is of equal significance as previous one. Opposing -isms are not just that but also author's mindful reminder of the nature of philosophy which approaches to a problem because of its controversy. They show to a beginner that philosophical approach as -ism is a view from a certain position regarding what draws philosopher's attention. Often we follow a philosophical discussion by ending with classification of opposing views or ending by identifying certain position as one of the -isms. Tennant reminds the reader that this is not enough and offers to the beginner clarifications of these -isms indicating problems which some of them deals with in detail in Part III.

What would be the use (usefulness) of Tennant's isms? The most important aspect is to be reminded that -isms, which are used so easily and sometimes without care, can be questioned about their precise meaning. For both, teachers and students, can be useful: they are reminded to challenge obviousness of -isms with scrutiny. They can clarify their knowledge about them more precisely. The teacher can help students by equipping them with a framework to map their already obtained knowledge and therefore to put particular pieces of knowledge to the map of -isms and consequently to have systematic insight into the whole. Moreover, Tennant's explanations could be useful for the students to overcome common sense understanding of particular concepts and relations among them.

5. *Tenant's methodology through the three principles*

The extensive chapter on methodology starts with logic and symbolization. How important logic is for Tennant, can be seen from his words: "A philosopher who shies away from formal analysis is like a surgeon who ignores the need for basic hygiene" (p. XVIII). In comparison with other introductory textbooks, he consciously "makes uninhibited use of logical analysis, schematization, and regimentation in order to clarify important views or methods as they are laid out" (p. XVII). High school teachers can go with students to the limit where students can follow. They can learn the basics of logic but they can also learn its significance, which Tennant explains and illustrates in quite an impressive way. The chapter includes the basics of inductive reasoning, methods of conceptual analysis, the method of conceptual explication, and the method of thought-experiment. The section Intellectual Creativity and Rigor introduces problematization and continues with issues of Deduction in Mathematics and Science, ending the chapter with The Methodological Issue of Reductionism.

The section on conceptual analysis provides an important pedagogical aspect for our purpose. Tennant's intention is to inform the beginner in philosophy that "a great deal of contemporary philosophical discussion in the journals is concerned with providing counterexamples to proposed conceptual analyses" (p. 125). Although his step by step detailed presentation of conceptual analysis as a technique illustrated with examples (such as "Gettier cases") aims for a "professional" analytical philosophy, high school teachers can still gain something valuable for doing philosophy with students. Students can learn not to take concepts for granted and to question them as described by Tennant: "stating individually necessary and jointly sufficient conditions for the application of the concept in question" (p.126). In the section as a whole, we can notice a method as a unity of problematization (questioning concepts), conceptualization (conceptual analysis) and argumentation, which can be applied to introductory courses of philosophy. A part of the above analysis is philosophically "sharpened intuitions" which "lead to the construction of thought-experimental counterexamples to faulty conceptual analyses on offer" (p. 125). This section then introduces the necessary and important method of Thought-Experiment, which needs attention in a special section.

The Method of Thought-Experiment

This section can be useful for high school teachers. Thought-Experiments (TEs) can be very creative and this creativity could be productive in philosophy class, since "one tests to the limit the application of concepts of philosophical importance. One imagines wildly different 'possible worlds' or bizarre situations which serve to bring out distinctions among concepts that might otherwise be taken to be the 'same',

by virtue of applying to the same objects under normal circumstances" (p. 153). Although students are usually not as interested in testing application of concepts as professional philosophers, they could be interested in "'possible worlds' or bizarre situations." Many students are familiar with Descartes' thought-experiment of the evil demon. Teachers report that usually they show interest discussing the well-known movie *Matrix*, and they could be inspired to go further to other 'bizarre situations'. Tennant challenges his students to engage themselves in TEs, putting "aside their beliefs concerning the probability or likelihood or feasibility of the imagined scenarios" for "acquiring this intellectual skill" (p. 155). He offers several TEs. However, for this purpose, I was (despite understanding his purpose) disappointed, that he does not offer more than a short summary of any particular TE. Maybe he could think about expanding this section in the next edition.

6. *Content as a demonstration of methodology in action*

Content is presented as "an explanation of … certain main philosophical Problems. They are the ones that the author finds both engaging and tractable by the intellectual methods that he has available, as someone coming from a background of logic and foundations within Analytical Philosophy" (p. XVII). For my purpose I will take four examples.

Anselm's Ontological Argument

After a methodological introduction explaining the nature of this a priori argument in comparison with mathematical theorems and scientific hypotheses, the problem "Does God Exist?" is presented in a systematic, extensive and detailed way: The original text in Latin, the English translation, a reconstruction in "logician's English," and exegesis of the argument in its formal shape, and various criticisms which are examined extensively and in detail. The first objection is that "Anselm tacitly uses a mistaken principle about linguistic understanding" (p. 217). The second is that "Anselm mistakenly treats existence as a property of things." The third is that the "Ontological Argument keeps bad company" and that "There are other arguments, of the same form, for patently unacceptable conclusions" (p. 219). The fourth (raised by the anti-realist) is that the "Ontological Argument uses a strictly classical form of *reductio ad absurdum* to which the anti-realist would object (p. 220). These objections are followed by a "completely rigorous regimentation of the argument" (p. 221) and by "Translating Anselmian chunks into logical notation" (p.225) and by offering "Further reading on the Ontological Argument" (p. 227). The four objections are enough for the teacher and students to follow and understand the reasons Russell had in mind when he said that "it is much easier to be persuaded that ontological arguments are no good than it is to say exactly what is wrong with them. This helps to explain why ontological arguments have fas-

cinated philosophers for almost a thousand years" (Oppy, 2016). This common journey with students has its limits, at which it makes sense to stop. Nevertheless, what follows is the advanced level. If students are motivated and equipped with the tools of analytical philosophy, they can proceed with the rigorous regimentation of the argument and its logical notation. Tennant's detailed, exhaustive, thorough and systematic analysis is welcome because it offers what the many textbooks lack. He also shows how particular issues in the critique are not definite and are still open and subject to different approaches (e. g. realism vs antirealism). Students can, again, learn that in philosophy there is no single solution to a problem and much depends on the perspective from which it is approached.

The Liar Paradox

An example of the philosophical content in Tennant's book describes how to approach some of the famous paradoxes. Among the reasons why paradoxes are worth studying, he mentions that "they are deeply puzzling, and often inspire young thinkers to pursue Philosophy more seriously" (p. 369). Let us illustrate the approach with two of them.

Tennant approaches the liar paradox in the following way:

1 The Liar is meaningless.
2 The Liar is meaningful, but the question of its truth or falsity cannot arise, since it does not 'engage with' any language-independent subject matter in a suitably 'grounded' way.
3 The Liar is meaningful, but is neither true nor false.
4 The Liar is meaningful, but is both true and false.
5 We should not use a language in which the Liar can be expressed; for such a language is incoherent.
6 We can and should use a language in which the Liar can be expressed; the alleged incoherence arising from the paradox is neither here nor there, and cannot threaten any serious scientific purposes. (p. 376)

For teachers, it can be useful to clarify which concepts and distinctions can be used and the extent that students can learn how to employ them (meaning, truth, language, coherence) and at the same time to realize that there is no one single solution to a problem. Tennant's approach can be used to explain to students how the issue is controversial, i.e. how controversy is in the nature of philosophy. In this case the concepts of meaning, truth, language, coherence as perspectives reveal the controversy.

Zeno's Paradox

Another well-known example is Zeno's paradox, which is presented as a mathematical paradox. "Zeno (mistakenly) thought that this temporal

sum would have to be infinite. So, he concluded, the arrow would never reach its target. We can see today exactly how Zeno's reasoning was mistaken. It is possible for an infinite series of finite numbers (such as 1/2, 1/4, 1/8, ...) to have a finite sum. Zeno did not realize that. Paradox dissolved" (p. 378). This can be learned from high school mathematics. However, according to Tennant, if we want to discuss and solve the problem with students, we just need to look at Zeno's *assumption* and his *belief* about it: we need to introduce the concept of infinite series of finite numbers and their sum, which Zeno mistakenly believed was infinite. This mathematical *concept*, so obvious to mathematicians and analytical philosophers, needs to be recognized as an *assumption*, and this is the task of the teacher to help students, if they are not able to do so. By doing this, we train students to look for assumptions. Although the example itself has a simple solution, it invites students to deal with other examples, and to develop the habit of looking for assumptions and of articulating, expressing assumptions into appropriate form. This is something that is obvious to professional philosophers, but is an ability that still needs to be developed with students.

It is worth emphasising that the content presented serves as a demonstration of the author's methodology, which is the focus of my attention: how it can be used by teachers in their work with students.

Mind/body as an example of a content demonstrating methodology as the unity of the three principles

The mind/body topic is one of the traditional topics in high school or university introductory courses. Although teachers have many resources for designing their work with students, Tennant's textbook can still provide them new possibilities and clarifications. One of them would be the presentation of the contrast between Descartes' contribution in mathematics, "the system of Cartesian coordinatization" and his solution of the "phenomenon of mind" which leads Tennant to use a "different order of exposition" of Descartes' Meditations. It is worth following this interesting pedagogical approach to Descartes' dualistic solution of mind/body problem. However, there is another value to this approach in the continuation of the topic. Tennant carefully expose the *problems* of this solution and offers a very clear presentation of Ryle's critique and his indication of categorical mistake. In his *argumentation,* he clearly explains and illustrates the *concept* of categorical mistake, which is again useful pedagogical contribution. Moreover, he shows the difficulties, *problems* of Ryle's approach which leads him to present attempts to solve these difficulties (Materialism and Supervenience) and new *problems* attempted by Functionalism etc. The same method, therefore, continues through the elaboration of all the approaches presented in the chapter—and throughout the whole book. If teachers carefully follow the development of this chapter and the author's methodology as a unity of the three principles, they can find the value and relevance

for high school teaching. If they pay attention to this methodology, they can apply it to other topics.

8. *Conclusion*

Tennant's textbook as a possible source for high school teachers, especially those who prepare students for final exams like A-level or International Baccalaureate, provides a very demanding and unique way of looking at the methodology of philosophy as a unity of problematization, conceptualization and argumentation. Teachers can make a use of these principles, if they carefully examine *how* Tennant employs them and if they apply them in an appropriate way.

It is of special importance that Tennant, as "one of the most notable figures" in the field of contemporary philosophy, is devoted not only to research but also to pedagogical issues of philosophy. Tennant's textbook is praiseworthy because of its pedagogical contribution. The scrutiny of demanding philosophical research is transferred to the (theory of) philosophy education. The implications are far reaching: the book can remind departments of philosophy to think about not only how to design the study of philosophy but also how to develop teaching methodology and perform particular courses.[2] Since my particular interest is philosophy education within secondary education, it is worthwhile to emphasise the challenge to what extent the scrutiny of philosophy can be implemented in the teaching of philosophy in secondary education in general and in the domains of problematization, conceptualisation and argumentation in particular.

Least but not last, Tennant's textbook is an example of developing a pedagogical approach to philosophy, an approach which by emphasising the importance of teaching methodology, demonstrates the necessity of a distinction between philosophical content, its form, and the process in which doing philosophy takes place.

Although the teacher as a reader must keep in mind the author's "liberty of presenting certain matters from its author's point of view" (p. XV), this does not diminish the pedagogical value of the book.

References

Kotnik, R. 2014. "Philosophy Textbooks: A Gap between Philosophical Content and Doing Philosophy." *Croatian Journal of Philosophy* 14 (40): 151–158.

Oppy, G. 2016. "Ontological Arguments." *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2016/entries/ontological-arguments/>.

Reeve, N. 2015. "Book Review. Introducing Philosophy: God, Mind, World and Logic." *Practice: Social Work in Action*: 1–2.

[2] Additional help for teachers with resources is available in the companion to the book on the website https://godmindworldlogic.wordpress.com

Šimenc, M. 2007. *Didaktika filozofije*. Ljubljana: Filozofska fakulteta.
Tozzi, M. 2008. "Compétences et discussions à visée philosophique." Université Montpellier 3. sept. 2007. (Retrieved 25.7.2016) http://www.philotozzi.com/2008/01/de-la-question-des-comptences-en-philosophie/

Croatian Journal of Philosophy is published three times a year. It publishes original scientific papers in the field of philosophy.

Croatian Journal of Philosophy is indexed in
The Philosopher's Index, Philpapers, Scopus
and in *Arts & Humanities Citation Index (Web of Science).*

Payment may be made by bank transfer
SWIFT PBZGHR2X
IBAN HR4723400091100096268

Instructions for Contributors

All submissions should be sent to the e-mail: cjp@ifzg.hr. Submissions must be in English and formatted to be double-spaced with suitably wide margins, an A4 page size, and automatic page numbering.

Articles are normally no longer than 8,000 words of main text including bibliography. The Journal will consider longer papers, but, once these limits are exceeded, authors should bear in mind the editorial policy that the acceptance bar raises with increasing length.

Manuscripts should be compiled in the following order: cover page; title; abstract (not exceeding 200 words); keywords (3 to 6); main text; appendices (as appropriate); references.

All the authors of a paper should include their full names, affiliations, postal addresses, telephone and fax numbers and email addresses on the cover page of the manuscript. If a paper is co-written, one author should be identified as the Corresponding Author. The cover page must be submitted as a separate document. All submitted manuscripts must be prepared for blind review, with revealing acknowledgements and self-identifying references removed.

Sources are cited in the text by the author's last name, the publication date of the work cited, and a page number if needed, e.g. (Barber 2007: 324). Full details appear in the reference list in which the year of publication appears immediately after the author's name:

Barber, A. 2007. "Linguistic Structure and the Brain." *Croatian Journal of Philosophy* 21 (7): 317–341.

Williamson, T. 2013. *Identity and Discrimination*. Oxford: Wiley-Blackwell.

ISSN 1333-1108
9 771333 110001