Contents

ARTICLES

DISCUSSIONS

BOOK REVIEWS

REPORTS

Organon F 23 (4) 2016: 428-442

Against Laws of Nature as Truthmakers for Presentists*

PABLO RYCHTER

University of Valencia
Av. Blasco Ibáñez 30. 46010 Valencia. Spain
pablo.rychter@uv.es

RECEIVED: 01-03-2016 • ACCEPTED: 14-09-2016

ABSTRACT: This paper addresses the so-called 'truthmaker problem' or 'grounding problem' for presentism. In section 1, I set the stage by introducing presentism and the truthmaker problem. In section 2, I consider a proposed solution to it, which I call the 'laws of nature proposal' (LNP), recently defended by Markosian (2013). I argue that LNP fails as a solution to the truthmaker problem because it does not meet a constraint that is generally taken as constitutive of it: that the entities doing truthmaker work should be *categorical.* Then, in section 3, I discuss the prospects of abandoning this 'categoricity constraint'. The conclusion of this discussion is that the presentist should be allowed to such a move. This, however, is not completely good news for the friends of LNP, since the abandonment of the 'categoricity constraint' opens the door to simpler solutions, like what is often called 'Lucretianism'.

KEYWORDS: Laws of nature – presentism – truthmaking.

* I wish to thank the following people for their helpful comments and discussion on previous versions of this paper: Dan López de Sa, Ned Markosian, Sven Rosenkranz, Giuliano Torrengo, Ezequiel Zerbudis, the participants in the PERSP Metaphysics Seminar, and the audiences in Santiago de Compostela, Buenos Aires, Barcelona and Girona. The research leading to this work was partially funded by the research project FFI2015-66372-P (Spanish I+D Plan).

1. Presentism and the truthmaker problem

Presentism is the view that only the present is real. Or, in other words, that only the present exists: there are no merely future objects and no merely past objects. That is to say, the world does not contain past dinosaurs or future outposts in Mars—things that would exist only in the past, or only in the future. Presentism contrasts with a variety of non-presentist views according to which at least some of these entities do exist and are as real as present things.

Presentists face the so called 'grounding problem' or 'truthmaker problem'. This problem may be pinned down as the apparent conflict between presentism and two other, plausible views: the view that truth supervenes on being (that is, the view that if *p* is true, it would be necessary that the world be different from how it is in order for *p* not to be true), and the view that some statements seemingly about the past (and about the future, but let us focus on the past) are true. It is clear what this apparent conflict is. Let us take

(D) Dinosaurs once roamed the Earth

as an example of a true statement seemingly about the past. If presentism is true, it looks as if the truth of D fails to supervene on how the world is. The world of the presentist does not stretch beyond the present, and contains nothing but present things. And nothing in this world *necessitates* the truth of D. If it were false that there were dinosaurs, everything today could look very much the same.[1] In other words, if presentism is true, it seems that there is a possible world *w* that is indiscernible from the actual world except for the fact that D is not true at *w*. So it seems that the three views (presentism, supervenience of truth on being, and that D is true) are incompatible.

As it is common in this sort of predicaments, one can either take the apparent conflict at face value and avoid contradiction by dropping one of the views in conflict (a move that is in general accompanied by an explanation of why

[1] We would not have fossils of dinosaurs (given that by 'fossil' we mean something actually caused by a distant past being), but we could have qualitative replicas of fossils of dinosaurs. Such a world would still look like the actual world. D is a standard example in the contemporary discussion of presentism and truthmaking, but if this feature of the example is distractive, it could be replaced by a sentence about any extinct natural kind that did not leave any fossils.

the rejected view *seemed* true at the beginning), or explore the idea that appearances are deceiving and that the three views are not really in conflict. One can also try to combine these two approaches. This is the path followed by the 'laws of nature proposal' (LNP), which we will examine in Section 2. As we will see, LNP has a 'conciliatory' and a 'rejectionist' side. But before moving on to LNP, I would like make a few remarks about the truthmaker problem for presentism.

First, I would like to say something about the choice of formulating the problem in terms of the principle that truth supervenes on being. This principle is usually taken as the least controversial of a series of principles about truthmaking. On the opposite end, we find the principle that every true proposition has a truthmaker, an entity that makes the proposition true. This principle is stronger because, unlike the supervenience principle by itself, it requires the existence of a particular entity (usually thought to be a fact, a state of affairs, or a trope) that acts as a truthmaker of the relevant true proposition. However, following Sider's (2001, 35) influential discussion, I am using the supervenience principle for the formulation of the truthmaker problem for presentism.[2] My main reason for doing this is that, as emphasized by Sider, given certain assumptions about what is admissible as a supervenience base (which I discuss next), the supervenience principle is already sufficient to generate the truthmaker problem for presentism. Thus, given that there seems to be no need to appeal to the stronger truthmaker principle, it is a good policy to use the weaker one. That way the problem turns out to be of interest also for those who believe in *truthmaking without truthmakers* (that is to say, people who accept the supervenience but not the truthmaker principle).[3] As I said, this is my own reason for formulating the problem in terms of supervenience. But it should also be noticed that even those authors who initially frame the discussion in terms of the truthmaker principle (like Cameron 2010) quickly end up discussing the same issue as those who instead use the supervenience principle: namely, the issue of what properties are admissible in an appropriate supervenience base. I think the reason for this has to do with the relatively extended assumption that a state of affairs is nothing over and above a particular instantiating a property (see for instance Armstrong 1997). Thus, for instance, the state of affairs of the

---

2 I thereby follow what I take to be a fairly standard procedure in the discussion of the problem. See for instance Caplan & Sanson (2011).

3 For more references about *truthmaking without truthmakers*, see Rychter (2014).

world having once contained dinosaurs (and alleged truthmaker for D) is nothing over and above the world instantiating the property of having once contained dinosaurs. Given this assumption, even if we were interested in finding out what truthmakers (states of affairs) the world contains, the central issue turns out to be what properties it really instantiates. Once we are clear on what properties the world really has, the truthmakers come as a "free lunch".

Second, it is worth keeping in mind that both the truthmaker principle and the supervenience principle are initially motivated by an intuitive idea that we can gloss in any of the following formulations: that truth is *grounded* in reality, that truths are true *in virtue of* reality, that truths are true *because* of how the world is, etc. Both the truthmaker principle (as usually understood) and the principle that truth supervenes on being fall short of capturing this intuitive idea. This is because both principles are allegedly reducible to modal notions and, in general, modal notions are too coarse-grained to capture the notions expressed by the italicized expressions above. Thus, even if the supervenience principle is the one invoked in the formulation of the problem, it would be disappointing if a solution to the problem were not also an explanation of how propositions such as D can be *grounded* in reality in the sense of the intuitive idea just mentioned.

Third, I claimed above that the supervenience principle, although weaker than the intuitive idea that motivates it, is enough to generate the truthmaker problem for presentism. But this is so just in case we are somewhat selective as to what could constitute the supervenience base for the truth of D and the like. Not any feature of the world can be allowed into this supervenience base. In particular, it is generally assumed by proponents of the truthmaker problem that the presentist cannot attempt to solve the problem by appealing to the property (presently exemplified by so many things) of being such that dinosaurs once roamed the earth. Properties like this are ruled out at the outset because they are, as proponents of the problem say, 'past-directed', they 'point beyond their instances', and are 'hypothetical' rather than 'categorical'.[4,5] It is

[4] Notice that, for the same reasons, the property of containing fossils of dinosaurs is ruled out. That property is 'past-directed' in the sense under discussion. Of course, we could admit in our supervenience base the property of containing qualitative replicas of fossils of dinosaurs, but the truth of D would fail to supervene on such a base.

[5] Following Sider (2001), I use the locutions 'hypothetical property' and 'property that points beyond their instances' to the same purpose.

in my opinion far from clear what these complaints amount to. (I'll come back to this in section 3). But in any case, it is clear that the formulation of the truthmaker problem should be understood as including this kind of constraint on any putative supervenience base. Otherwise, the problem is too easily solved. Let me put it in other words: whoever takes the truthmaker problem seriously (be it a presentist who thinks it can be solved, or a non-presentist who thinks that the problem is fatal for presentism) is driven by the *aspiration to find a categorical supervenience base* for truths like D, a set of properties that do not 'point beyond their instances'. Whether or not this aspiration is a reasonable one for a presentist (and thus whether or not she must take the truthmaker problem seriously) is something I will discuss in Section 3.

Finally, I want to distinguish the truthmaker problem from another, related problem that is often discussed under the label of 'the problem of singular propositions' (see Markosian 2004). If presentism is true, there are no merely past individuals and so no merely past individuals are available as objects of reference or as members of domains of quantification. Thus, given that 'Socrates' purports to refer to a past individual, we fail to express a singular proposition by 'Socrates was a philosopher'. The problem of singular propositions is not that, allegedly, the presentist cannot explain what *grounds* the *truth* of the proposition expressed by this statement. It is rather that it cannot explain how the sentence is *meaningful* under the standard assumption that such a sentence expresses a singular proposition. Or in other words, the problem consists in explaining what proposition the sentence expresses and how it does it, given that it cannot express the singular proposition that it is generally thought to express. I think it is worth considering briefly how these two problems relate to each other. One may think that the problem of singular propositions concerns meaningfulness in addition to truth only because it is a more specific problem concerning one particular type of truths, but that the general worry underlying both problems is the same, and that the truthmaker problem is the more general and encompassing formulation of it. (This would explain why the truthmaker problem attracted so much more recent attention than the problem of singular propositions did). The problem, it is tempting to say, is just one: that the presentist's ontology is too sparse, and that it does not contain the materials necessary to ground the truth, and in some cases also the meaningfulness, of the the things we say seemingly about the past. But for reasons that I discuss in Section 3, I think this line of reasoning is misleading, and that we should keep the two problems apart. In order to do so, it will be useful, when

discussing the truthmaker problem, to focus on sentences like D, which do not express or presuppose any singular proposition seemingly about a past object. In fact, perhaps D is not the best example, since it may be thought to involve reference to an extinct natural kind. I'll keep using it nevertheless, but its job could also be done by something like 'There were round objects' (which does not seem to express or presuppose any singular proposition about the past) or 'Obama was a child' (which expresses a singular proposition but not about any past object).

2. The laws-of-nature proposal

Let us now turn to LNP, a proposal for solving the truthmaker objection that has recently been defended by Ned Markosian (see Markosian 2013). LNP's basic idea is roughly this: contrary to what we were assuming, many truths seemingly about the past, perhaps D itself, are necessitated by the presentist's temporally narrow world. This world is narrow, but it contains a system of laws of nature. This system of laws is either completely deterministic or somewhat indeterministic. If it is completely deterministic, then it determines, together with everything else in the (present) state of the world, how the world was and will be. (According to LNP, a system of completely deterministic laws of nature fix reality in both directions: just as there is only one possible future given how the present is, so is there only one possible past).[6]

---

[6] That is to say, according to LNP, laws of nature are *bi-directionally* deterministic (to the extent that they are deterministic at all). Markosian (1995, 100) makes this assumption explicit and offers references for further discussion of it. A nice statement of the assumption of bi-directionality is offered by Lewis (1979, 460), who also makes it when discussing determinism: 'A *deterministic* system of laws is one such that, whenever two possible worlds both obey the laws perfectly, then either they are exactly alike throughout all of time, or else they are not exactly alike throughout any stretch of time. They are alike always or never. They do not diverge, matching perfectly in their initial segments but not thereafter; neither do they converge'. That deterministic laws determine also the past (given the present state of the world) may sound surprising because in the philosophical discussion of determinism it is *future* determination that is most often invoked. Nevertheless, as noticed by Hoefer (2016, sec. 2.3), 'for a wide class of physical theories (i.e., proposed sets of laws of nature), if they can be viewed as deterministic at all, they can be viewed as bi-directionally deterministic.'

So contrary to what we assumed before, there is no possible world *w* that is indiscernible from the actual world except for the fact that D is not true in *w*. In envisaging such a world, we were forgetting about the completely deterministic system of laws of nature. (Notice that somehow or other that system is a feature of the actual world, and so any world that is indiscernible form the actual world is indiscernible with respect to it). If the laws of nature are somewhat indeterministic, on the other hand, less things about the past and about the future are determined by how the world is. But according to LNP, it is ok for the presentist to deny the truth of everything that is not so determined. In particular, it is ok to say that D is not true, if it really turns out that the (present) state world is compatible with the falsity of D—i.e. if it is nomologically possible, given the (present) state of the world, that D is not true. This is why I said before that LNP's approach to the truthmaker problem had two sides. LNP says: if the laws of nature are deterministic, then presentism, supervenience, and D are all true, contrary to appearances. If the laws are indeterministic, on the other hand, then D and other statements about the past that we take to be true are in fact not true.

I have several worries about LNP, the discussion of which will help me to present what I think presentists should make of the truthmaker problem. As we have seen, LNP crucially appeals to the laws of nature: truths about the past and about the future are true in virtue the world's (presently) being governed by certain laws of nature (together with any other states of affairs). My first worry is this: to the extent that laws of nature are suited to do this work, they will also fail to satisfy the categoricity constraint mentioned above. Nomic properties, the properties in virtue of which laws of nature apply, will not satisfy the aspiration (that animates proponents of the truthmaker problem) of finding a categorical supervenience base. Laws of nature are paradigmatically hypothetical: they tell us how the world *would* be if such and such conditions *obtained*.

You may want to say: 'laws of nature may well be categorical entities after all. If the Humean view about laws of nature is correct, laws of nature are nothing but (or are determined by) regularities among particular events, and these in turn may be reduced to the instantiation of categorical properties'. I agree, but I also think that this Humean view is not available for presentists. The view is available to non-presentists that embrace past and future events, but it looks as if the presentist's repertoire of events is too sparse to determine many laws of nature. In particular, I can't see how present events, i.e. all the events that

exist according to the presentist, could determine diachronic laws of the sort 'if you shake a bunch of Fs at t_1, you *will* get a G by t_2. (You may want to protest here and say: 'but that we *will* get a G by t_2 is a present event!' Fair enough, but laws of nature determined by this sort of future-directed events cannot be used in a solution to the truthmaker problem. Remember the constraint that only 'categorical' properties be allowed in the supervenience base.) That present events cannot determine diachronic laws is especially problematic, because it is precisely this kind of laws that may be thought to ground truths seemingly about the past and the future. So it seems that LNP is incompatible with the Humean conception of laws, and that it *requires* a different conception. This is already very significant because, it seems to me, the Humean picture about laws and the aspiration to solve the truthmaker problem (which, remember, includes finding a *categorical* supervenience base for truths seemingly about the past) are very close relatives, two projects arising from a single source: the picture of reality as 'vast mosaic of local matters of particular fact, just one little thing and then another', with no necessary connections between them, no pointing to each other. So it seems odd that in order to solve the truthmaker problem the friend of LNP should give away the Humean conception of laws and go shopping for a different conception. But this is in fact what Markosian (2013) suggest we do, mentioning the 'Armstrong-Dretske-Tooley' account of laws as a promising alternative. Let us see how things stand if we follow this path.

On the Armstrong-Dretske-Tooley account ('ADT account' for short), laws of nature are relations between universals and so, given this account, LNP amounts to the idea that D is grounded in a relation between universals (or in more than one). But what are these universals? Universals are sometimes thought of as abstract entities, but I think that the friend of LNP who adopts the ADT account should rather follow Armstrong in thinking of universals as constituents of the concrete, material world. It would be odd if at this point the friend of LNP said that it is the relations between a bunch of abstracta, rather than a feature of the concrete material world, what grounds or makes true that dinosaurs once roamed the Earth. (In fact, I think that it is in general the friend of ADT, and not only the friend of LNP who subscribes ADT, who should think of universals as constituents of the material world: I find very implausible the idea that lawlike connections in the concrete, material world hold *because* a bunch of abstracta bear a particular relation.)

So universals must be thought of, as Armstrong did, as concrete constituents of the material world. But Armstrong is a non-presentist and so his universals have merely past and merely future instances (at which they are multiply located). Friends of LNP cannot follow Armstrong this far. If they adopt ADT, they must think of universals as *presently* existing constituents of the material world. But this has some unappealing consequences. First, it seems possible that some universals are *alien* to the present: that is to say, that the present fails to contain some universals that nevertheless had or will have instances. Suppose, for instance, that in the past the world contained *G* particles, a kind of fundamental particle that is now extinct. Since there are no such particles, *being a G particle* is not a universal and there are no laws about *G* particles. The consequence of this is clear: friends of LNP cannot ground any alleged truth about *G* particles or about anything involving *G* particles (if dinosaurs were in part constituted by *G* particles, the alleged truth of D cannot be grounded, for instance). So friends of LNP must either deny the possibility of universals alien to the present, or bite the bullet and deny the truth of many of the propositions about the past that we initially thought to be true. Second, and most importantly, the problem of diachronic laws that LNP faced on the assumption of the Humean theory, re-appears here on the assumption of ADT. Suppose that we say that N(F, G) is a law of nature relating universals F and G. Suppose further that this law together with the fact that a bunch of Fs are shaken at t_1 jointly determine that we *will* have a G at t_2. On these assumptions, N is at least necessarily co-extensional with a relation N* that is *future directed*: the relation of *having instances that will exist right after/before instances of…* And although we do not yet have a full explanation of what it is for a property or relation to be future-directed, it seems reasonable to suppose that being future-directed is a feature that is shared between necessarily equivalent properties and relations. But if so, the appeal to N violates the categoricity constraint and thus does not constitute a satisfactory solution to the truthmaker problem. This confirms what we should have been suspecting from the start: an appeal to an unexplained necessary connection between distinct existences can hardly be welcome by someone who is moved by the aspiration to ground all truths in a *categoric* supervenience base.

Before moving on, let me mention other additional worries about LNP. First, even if LNP succeeded in making laws of nature part of an appropriate supervenience base for the truth of D, and thus succeed as a solution to the truthmaker problem as stated above, it would not thereby succeed in finding

an appropriate *ground* for D. It seems implausible to say that it is true that dinosaurs existed partly *because of* (or *in virtue of*) the fact that such and such laws of nature presently hold. Of course, that such and such laws of nature presently hold can be good evidence for *believing* that there were dinosaurs, but not what *makes it the case* that there were dinosaurs. (If anything, as Humeans have it, the other way around looks more plausible: such and such laws of nature presently hold partly because there were dinosaurs.) Second, the 'rejectionist' side of LNP leaves the view in need of some supplementation. In order to see this, suppose that the laws of nature are indeterministic enough so that D is not true. If we accept the principle of bivalence, we should also say that, in these circumstances, D is false. Now my question is: is D as bad as C?

(C) Centaurs once roamed the Earth.

If the laws of nature are sufficiently indeterministic (something that is the business of science to determine), the presentist who adopts LNP has to say that both D and C are false. Nevertheless, I think it is reasonable for the friend of LNP to distinguish between C and D, and to confer on D some kind of positive status that C lacks. Perhaps D is in some sense closer to truth than C. Or perhaps D somehow expresses a true proposition, when understood non-literally. In any case, I think it is important for the friend of LNP to have some story to tell about the positive status of D vis à vis C. Otherwise, I think the proponent of the truthmaker problem will be unsatisfied by LNP, since the proposed solution would amount to reject one of the three views that generated the problem (the view that D is true), without any explanation of why it seemed true in the first place.

3. Against the categoricity constraint

We have seen that the main obstacle that LNP stumbles upon is the categoricity constraint—a constraint that, as we have stressed, should be understood as constitutive of the truthmaker problem. In this section, I want to address whether this constraint is reasonable for a presentist and thus whether presentists should be worried about the truthmaker problem. My conclusion will be disappointingly negative: I think presentists have no good reason to accept the categoricity constraint, and no reason to be worried about the truthmaker problem. This, however, is not good news for friends of LNP, given that

LNP is offered as *solution* to the truthmaker problem, i.e. as a view that strives to meet the demands imposed by the proponents of the problem. Once we drop the categoricity constraint, there are other, simpler and more natural views for the presentist to hold.

Let us start by considering the view that has been called *Lucretianism.* According to this view, the world presently exemplifies the property of having once contained dinosaurs. If we take this property as part of how the world is (i.e., as part of the relevant supervenience base for the principle that truth supervenes on being), there is no world that is indiscernible from ours with respect to being, and such that D is not true in it. It is precisely against proposals like this that the categoricity constraint is put forward. As proponents of the truthmaker problem say, Lucretian properties like *having once contained dinosaurs* are inadmissible because they are 'past-directed', they 'point beyond their instances', they 'float free from the world', and thus constitute a case of 'cheating'. As we noted earlier, it is difficult to find a clear, definite statement of the worry behind these complaints.[7] Sider (2001) admits that the notion of the categoric is 'elusive', and Schaffer (2008) admits that current characterizations of the notion 'could use further work'. But despite the fact that we lack a clear formulation of the categorical constraint, we can get some grasp on it by considering particular examples of objectionable properties and the broader assumptions and theoretical context in which the truthmaker objection seems pressing.

---

[7] An exception is Cameron (2010) who analyzes categoricity in terms of present intrinsic nature. On this proposal, we can admit in the relevant supervenience base any property that contributes something to the present intrinsic nature of its exemplifier. The property of having contained dinosaurs is ruled out because it does not tell us anything about how the world presently *intrinsically* is. I do not intend to offer a full assessment of this proposal, but I would like to mention two initial worries. First, the notion of intrinsicness is not unproblematic and relying on it may not be the most fruitful strategy in this context. A popular way of understanding intrinsicness is in terms of perfect duplication. But how can we rule out the view that if something has the property of *having once been F,* so does any *perfect* duplicate of it. The problem, more generally, consists in securing a characterization of intrinsicness that rules out past-directed properties. Second, I doubt this analysis of categoricity captures what Sider and others were after in their complaints about past-directed properties. What seems to bother these people about properties like *having contained dinosaurs* is the fact that these properties *point to the past*, not the fact that they do *not* point to the present. (See Caplan & Sanson 2011 for a similar diagnosis).

I admit that there is something suspicious about Lucretianism. If you ask yourself '*what feature of the world makes it true that once there were dinosaurs?*', it indeed sounds as cheating to say that it is the fact that the world instantiates the Lucretian property of having once contained dinosaurs. But the presentist's story about truthmaking need not end at that point. Rather, I think the presentist should insist on the following 'core idea': what is true is true in virtue of how the world is, *or in virtue of how the world was and will be*. That there once were dinosaurs (D) is true in virtue of how the world *was.* If, when talking to the proponent of the truthmaker problem (who is armed with the supervenience principle), the presentist chooses to rephrase this by saying that D is true because the world instantiates the Lucretian property of having contained dinosaurs, that is fine. But it should be kept in mind that this is mere rephrasing. Saying that the world instantiates the Lucretian property is nothing but saying the world is such that there were dinosaurs – or simply that once there were dinosaurs.

Let us come back to what I think should be the presentist's core idea: truths are true in virtue of how the world is, *or in virtue of how the world was and will be*. This idea allows the presentist to happily join truthmaker theorists in their central motivation: *truth is grounded in reality*. Truth is grounded in reality, the presentist says, but not just in how reality *is.* It is also grounded in how reality *was* and *will* be.[8] In other words, there is a way things were, and those propositions that correctly describe that way things were are true. Those that do not correctly describe the way things were are not true. There is no present feature of the world that makes D true because D is not about how the world *is.* D is about how the world *was,* and so is true in virtue of a feature that

[8] You may think that it is illegitimate to move here from *truth is grounded in reality* to *truth is grounded in how reality is*. That is to say, you may think that these two slogans are best understood as representing two different views. The first slogan may be best understood as representing standard truthmaker theory, according to which each truth is true in virtue of the existence of a particular entity (so the first slogan would be equivalent to *truth is grounded in existence*). The second slogan may be best understood as representing the weaker (and to my mind much more plausible) view that truth supervenes on what things exist *and how they are*. Now, the "core idea" that I am here putting forward on the presentist's behalf depends on the assumption that it is the second doctrine, and not the first, the one that matters. I think this is correct, and that it would be bad news for the proponent of the truthmaker problem if the problem in the end presupposed the stronger doctrine.

the world *had* (and no longer does). Far from cheating, this all sounds to me like perfectly sane, honest common sense (I'll say more about honesty below).[9] What are the consequences of this 'core idea' for the truthmaker problem? This is one way of thinking about it: the presentist's core idea amounts to keeping the supervenience principle while rejecting the categoricity constraint and thus admitting Lucretian properties. But this way of thinking about the presentist position is most appropriate for proponents of the truthmaker problem, not for presentists themselves. That is to say, it is the way in which presentists are forced to describe the position *when talking to proponents of the truthmaker problem*. It is not the way in which presentist would naturally describe their own view. And it is only when the presentist makes this effort of communication when the presentist's discourse sounds as cheating.

This takes me to note something peculiar about the dialectics in this area. There are two prima facie different ways of arguing against a philosophical position: one may attempt to show that the position harbors internal tensions or inconsistencies, or one may attempt to show that the position is inconsistent with a view that is uncontroversially true and widely accepted. Clearly, proponents of the truthmaker problem are not following the first strategy. They attempt to show that presentism is incompatible with widely accepted views. But they do not succeed in this. For although it seems uncontroversial that truth is grounded in reality (and as we have seen, the presentist agrees with this), it is not equally uncontroversial that truth supervenes on a base of *categorical* properties. In fact, as we already mentioned, the precise content of this constraint is not even spelled out. It may be thought that even if we lack a precise formulation of the categoricity constraint, there are nevertheless clear reasons that should lead us to accept *something* in the vicinity. But I can't see there are such pressures. The motivation for the categoricity constraint is, perhaps, as Carroll (1994, 5) puts it, 'simply the gut feeling' that non-categorical properties are somehow not fundamental. Or perhaps it is the epistemological worry (close to Hume's original one) that we cannot *see* the difference between a world that exemplifies the property of having once contained dinosaurs and one that does not. None of these strike me as powerful reasons to adopt roughly defined view like the categorical constraint. Of course, a philosopher may still be moved by

[9] Here I am indebted to Rosekranz (2012), who makes similar remarks in defense of Ockhamism (the view that future contingents are true without there being anything past or present that makes them true) rather than presentism.

her gut feelings to adopt such a view as a working hypothesis and even as a central tenet of her metaphysics. But then she can't really use the principle as a premise of an argument against someone else. In conclusion: proponents of the truthmaker problem have succeeded in showing why *they* cannot be presentists, not in showing that *presentists* are wrong in holding their views.

A remaining worry: Sider (2001) emphasizes the fact that presentists who appeal to Lucretian properties would be in the same boat as other metaphysicians (those accepting brute dispositions, etc) that seem clear cheaters. I cannot engage in a comparison between presentism and these other views, or in a discussion of whether the alleged cheaters are really cheaters. But I want to finish by pointing to a dimension of *honesty* in presentism. I agree that it would be suspicious if the presentist, holding such an austere ontological position, could easily cook up a ground for the truth of *any* proposition that we ordinarily take to be true.[10] That would be cheating. It would be like refusing to pay the price of ontological austerity. But the presentist is not in the position of offering grounds for any alleged truth we like. She cannot ground the truth of 'Socrates was a philosopher', for instance, given that Socrates does not presently exist. Here is where the distinction we made above between the truthmaker problem and the problem of singular propositions becomes relevant. If the presentist follows the Lucretian path, she can ground the truth of all *general* propositions seemingly about the past. Given that, as we are assuming, *dinosaurhood* is a presently existing property (and reducible to presently instantiated properties), the presentist has in his ontology the materials to ground the truth of D, or any other past-tensed purely general proposition. She has the materials to 'construct' the Lucretian property of having once contained dinosaurs. But given that Socrates is not in the presentist's ontology, she does not have the materials to construct the property of having once contained Socrates, and so she cannot ground the truth of *Socrates was a philosopher* in the same way as she grounds the truth of D. Thus the presentist does not cheat: she does pay a price for her ontological austerity.

[10] I have in mind easy recipes of the following kind: 'it was the case that *p* because the world exemplified the property of being such that *p*', or 'because it contained the fact that *p*', etc.

4. Conclusion

The presentist should not follow the friend of LNP in trying to ground truths 'about' the past in the laws of nature. In Section 2, I argued that LNP is an unsuccessful attempt to meet the challenge set by proponents of the truthmaker problem. In Section 3, I have argued that presentist should not feel obliged to meet the challenge in the first place.

References

ARMSTRONG, D. M. (1997): *A World of States of Affairs*. Cambridge University Press.

CAMERON, R. (2010): Truthmaking for Presentists. *Oxford Studies in Metaphysics* 6, 55-100.

CAPLAN, B. & SANSON, D. (2011): Presentism and Truthmaking. *Philosophy Compass* 6, No. 3, 196-208.

CARROLL, J. (1994): *Laws of Nature*. Cambridge University Press.

HOEFER, C. (2016): Causal Determinism. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/archives/spr2016/entries/determinism-causal/.

LEWIS, D. (1979): Counterfactual Dependence and Time's Arrow. *Noûs* 13, No. 4, 455-476.

MARKOSIAN, N. (1995): The Open Past. *Philosophical Studies* 79, No. 1, 95-105.

MARKOSIAN, N. (2004): A Defence of Presentism. *Oxford Studies in Metaphysics* 1, 47-82.

MARKOSIAN, N. (2013): The Truth About the Past and the Future. In: Correia, F. & Iacona, A. (eds.): *Around the Tree: Semantic and Metaphysical Issues Concerning Branching Time and the Open Future*. Springer, 127-141.

ROSENKRANZ, S. (2012): In Defence of Ockhamism. *Philosophia* 40, No. 3, 617-631.

SIDER, T. (2001): *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford University Press.

RYCHTER, P. (2014): Truthmaker Theory without Truthmakers. *Ratio* 28, No. 3, 276-290.

SCHAFFER, J. (2008): Causation and Laws of Nature: Reductionism. In: Sider, T., Hawthorne, J. & Zimmerman, D. W. (eds.): *Contemporary Debates in Metaphysics*. Oxford: Blackwell.

Organon F 23 (4) 2016: 443-459

Defective Equilibrium

DANNY FREDERICK

Slate House, Hunstan Lane
Old Leake. Boston. PE22 9RG. United Kingdom
dannyfrederick77@gmail.com

RECEIVED: 19-06-2016 • ACCEPTED: 22-09-2016

ABSTRACT: I argue that the conception of reflective equilibrium that is generally accepted in contemporary philosophy is defective and should be replaced with a conception of fruitful reflective disequilibrium which prohibits ad hoc manoeuvres, encourages new approaches, and eschews all justification in favour of continuous improvement. I suggest how the conception of fruitful disequilibrium can be applied more effectively to moral enquiry, to encourage genuine progress in moral knowledge, if we make moral theory empirically testable by adopting a meta-ethical postulate which is independently plausible.

KEYWORDS: Ad hoc – defective equilibrium - empirical testability – growth of knowledge – justification – reflective equilibrium.

1. Reflective equilibrium as state and method

Reflective equilibrium has been recommended with respect to empirical science by Otto Neurath (see Neurath 1983, 94-95), with respect to logic by Nelson Goodman (see Goodman 1983, 61-64), with respect to ethics by John Rawls (see Rawls 1999, 40-46), and across the whole field of enquiry by other philosophers, including W. V. Quine (see Quine 1951, 39-43), though the *term* was coined by Rawls (1999, 18). A standard exposition of reflective equilibrium might run as follows.

We are in reflective equilibrium when the set of propositions we accept satisfies two conditions: its elements are mutually consistent; and some of its elements provide the best available explanation for some of the others (the latter are then said to 'support' the former). As Norman Daniels puts it:

> [W]e achieve reflective equilibrium when we arrive at an acceptable coherence … An acceptable coherence requires that our beliefs not only be consistent with each other (a weak requirement), but that some of these beliefs provide support or provide a best explanation for others. (Daniels 2011, sec. 1).

Mutual consistency is, in principle, an all-or-nothing affair: either two propositions are consistent or they are not. In practice things are not so simple because of vagueness or indeterminacy. Explanatory coherence, on the other hand, is a matter of degree. For example, suppose that we are in a reflective equilibrium in which a proposition, *P*, which is in our accepted set of propositions, explains a subset, *S*, of those propositions. Suppose, further, that a proposition, *P′*, which is currently outside our set of accepted propositions, explains *P*, and thus also *S*, and also explains some propositions in our set which *P* does not explain. Then, by adding *P′* to our set of accepted propositions, we will increase the coherence of our set of accepted propositions, thereby moving from one state of reflective equilibrium to a more coherent, and thus better, state of reflective equilibrium.

Philosophers usually recommend reflective equilibrium as a desirable *end-state*, so they recommend that we seek to remove any inconsistencies in the set of propositions we currently accept and that we choose the best of any available rival explanations. The process through which equilibrium is achieved is sometimes called the *method* of reflective equilibrium. It involves comparing our accepted propositions, noting any inconsistencies, and then revising either particular or general propositions to achieve greater coherence by eliminating the inconsistencies and perhaps also improving the explanations:

> The method of reflective equilibrium consists in working back and forth among our considered judgments (some say our "intuitions") about particular instances or cases, the principles or rules that we believe govern them, and the theoretical considerations that we believe bear on accepting these considered judgments, principles, or rules, revising any of these elements

> wherever necessary in order to achieve an acceptable coherence among them. (Daniels 2011, sec. 1).

One might wonder why our enquiries should be governed by the norms of reflective equilibrium theory. Since enquiry is a purposeful activity, one would expect reflective equilibrium to be commended as helpful in achieving the purpose of that activity. However, on the reasonable assumption that the purpose of enquiry is to extend our knowledge, the practice of reflective equilibrium would seem to be counter-productive. Specifically, reflective equilibrium theory has the following shortcomings:

(i) it permits inconsistencies to be removed by ad hoc manoeuvres;
(ii) it fails to acknowledge explicitly the essential contribution of increases in reflective *disequilibrium* to the growth of knowledge;
(iii) it takes static equilibrium, rather than ongoing improvement, as the ideal.

I explain these shortcomings of reflective equilibrium theory in turn, making use of illustrations from the history of science, and I propose a dynamic conception of fruitful reflective disequilibrium to replace it. I outline some apparent difficulties in applying the conception of fruitful reflective disequilibrium to promote the growth of moral knowledge; then I suggest that the difficulties might be overcome if we adopted a reasonable meta-ethical postulate that renders moral theories empirically testable.

2. Ad hoc manoeuvres

If we discover an inconsistency within the set of propositions we accept, there are better and worse ways of eliminating it. An ad hoc manoeuvre is one that eliminates the inconsistency without teaching us anything new (cf. Popper 1959, sections 19-20). If our purpose in enquiry is to extend our knowledge, to learn something new, then we should demand that a move in the direction of reflective equilibrium that removes an inconsistency should also explain something new. The point can be illustrated with two examples from the history of science.

First, in the mid-nineteenth century, the observed motions of Uranus conflicted with the predictions of Isaac Newton's theory. There were thus incon-

sistencies between accepted observation statements and a previously successful explanatory theory. The inconsistencies could have been removed by simply rejecting the observation statements as hallucinatory, or rejecting Newton's theory, or amending Newton's theory so that it made an exception of Uranus, or positing the existence of a special force acting on Uranus which had no other effects. However, each of those manoeuvres would have been ad hoc: they would have eliminated the inconsistencies without teaching us anything new.

A better way of eliminating the inconsistencies was proposed by Urbain Leverrier. Realising that the refuted predictions followed from the conjunction of Newton's theory with accepted background knowledge, he proposed to replace some of that background knowledge. Thus, he accepted both Newton's theory and the observation statements about Uranus, but he denied that the known planets are all the planets there are: he hypothesised a new fact, namely, the existence of another planet with just the properties necessary to account for Uranus' anomalous motions in terms of Newton's theory. What saved this from being ad hoc was that the new hypothesis could be tested independently, for it implied that the new planet would be seen in a specific portion of the sky at a particular time. The hypothesis passed that test: Neptune was discovered (cf. Kuhn 1957, 261-262). Thus, Leverrier's manoeuvre did not simply remove the inconsistencies in our accepted propositions; it also explained (indeed, successfully predicted) something new, namely, the observation statements concerning the positions of Neptune.

Second, in the late seventeenth century a number of observation statements accepted by the Astronomer Royal were inconsistent with the predictions of Newton's theory. The observation statements could have been rejected as hallucinatory or the result of incompetence, or Newton's theory could have been rejected; but such moves would have been ad hoc. Instead, Newton amended background knowledge concerning the way that the Earth's atmosphere refracted light, which makes heavenly bodies appear to be some distance from where they actually are. The amended theory of atmospheric refraction explained why the previously accepted observation statements were false and also why they had seemed to be true; and it survived independent tests and thus explained something new (see Lakatos 1978, 215-216).

Rawls recognises that when judgements about particular cases are rejected because they conflict with a general principle, it would be an advantage to have an explanation for why the particular judgements seemed acceptable, but he

does not require that such an explanation be provided and he does not require that the explanation also explain something new:

> When a person is presented with an intuitively appealing account of his sense of justice (one, say, which embodies various reasonable and natural presumptions), he may well revise his judgments to conform to its principles even though the theory does not fit his existing judgments exactly. He is especially likely to do this if he can find an explanation for the deviations which undermines his confidence in his original judgments and if the conception presented yields a judgment which he finds he can now accept (Rawls 1999, 42-43).

It might be rejoined that the standard accounts of reflective equilibrium recognise implicitly the superiority of adjustments to particular or general propositions which are not ad hoc, in that eliminating an inconsistency in a way which explains something new increases the coherence of the set of propositions we accept and thereby achieves a better reflective equilibrium. However, if the stricture against ad hoc manoeuvres is not stated explicitly as a *requirement*, such manoeuvres will be deemed acceptable whenever there happens to be no better, more coherent, equilibrium currently available. That removes the imperative to *increase* our knowledge; it thereby condones stagnation.

It may be complained that the prohibition on ad hoc manoeuvres is too challenging. For instance, it might be said that, if we think that all ravens are black, then discover a white raven and thus give up the theory that all ravens are black, we are doing nothing amiss; we are rather making reasonable adjustments to achieve a new reflective equilibrium. However, if the purpose of enquiry is to extend our knowledge, we should not be content with such an ad hoc manoeuvre. We should instead try ways of removing the inconsistency that promise to teach us something new. For example, we could impugn the assumption that white is the natural colour of the anomalous raven. A new hypothesis that the raven had been painted white would explain why the observation statement that the raven is white appeared to be true and (in conjunction with background knowledge) it would have testable implications concerning how the white colour could be removed. If the implications survive the tests, the hypothesis has predicted, and thus explained, something new. Alternatively, we might add a qualifying condition to the generalisation that all ravens are black that not merely allows some non-black ravens but also implies that

we will find other non-black ravens under specific circumstances that we can either discover or construct. If the modified generalisation survives testing, it will teach us something new. It is true that the proscription of ad hoc manoeuvres is a challenging demand that it will often be difficult to meet, and theorists may sometimes have to make numerous attempts to improve upon their resolution of an inconsistency in order to meet it. But wherever an inconsistency is removed without meeting the demand, that should be highlighted as a defect requiring eventual amelioration.

It might be objected that the notion of teaching us something new, on which the identification of ad hoc manoeuvres depends, is vague. After all, any ad hoc adjustment of a theory will teach us how the unadjusted theory can be made consistent with the proposition(s) with which is inconsistent; so it will teach us *something* new. However, such a novelty is not novel enough: an adjustment that is not ad hoc successfully predicts or explains something, or solves a problem, that is independent of the problem it was introduced to solve. Ad hocness is a matter of degree, so we can expect some borderline cases; and whether something is explanatory or novel or a solution to a problem often involves qualitative considerations and thus judgement, so we can expect some disputed cases. There are, however, many clear-cut cases. It seems clear that the successful Newtonian adjustments outlined above are not ad hoc, as they entailed unexpected empirical predictions which survived testing; and the same goes for the hypothesis that the raven was painted white, if it survives testing. It seems clear, too, that amending Newton's theory to make an exception for Uranus would have been ad hoc. The fact that there can be no general algorithm for ad hocness does not detract from the notion's usefulness.

3. Creative disequilibrium

One thing we know from the history of science is that the growth of knowledge is brought about by people who create reflective disequilibrium by discovering or generating an inconsistency. That sets in train attempts to eliminate the inconsistency by making modifications to our accepted body of theory; and those equilibrating modifications will in turn make further contributions to the growth of knowledge, so long as they are not ad hoc. That may be done in three connected ways.

First, it is characteristic of science that accepted theories are subjected to experimental tests. But an experimental test of a theory is a serious attempt to refute it, that is, an attempt to produce an inconsistency between the theory and an experimental result. Experimental results inconsistent with a previously successful theory can be acceptably explained away if the explanations also explain something new, as in the cases of Leverrier and Newton mentioned in section 2.

Second, scientific knowledge is often augmented by means of thought-experiments which disclose inconsistencies within existing theories and thereby lead to conceptual change, as with Galileo Galilei's criticism of Aristotelian dynamics. Conceptual changes that dissipate paradoxes should also explain something additional, as Galileo's distinction between average and instantaneous speed not only resolved the paradoxes revealed in his thought-experiment but also enabled the solution of problems involving accelerated motion (cf. Kuhn 1977).

Third, we increase our scientific knowledge by proposing new theories that contradict previously successful theories, as Johannes Kepler's astronomy contradicted Nicolaus Copernicus's system, as Newton's mechanics contradicted both Kepler's and Galileo's theories, and as Albert Einstein's relativity theories contradicted Newton's (cf. Kuhn 1957; Popper 1983, 75, 131-149). Where a new theory contradicts a currently successful theory, removing the inconsistency will require that at least one of them is rejected (at least in its current form). Rejecting the new theory on grounds of tradition, or conservatism, would be ad hoc, as would rejecting the old theory because it is old. But if the new theory can explain not only the success of the old theory but also something else, then accepting the new in preference to the old is not ad hoc. For example, Newton's theory not only contradicted Kepler's and Galileo's theories, it also explained why they were successful (their predictions about celestial and terrestrial motions, respectively, were approximately accurate) and it explained other things besides, such as the motions of the tides (see Popper 1983, 139-145, 190-191).

We should add a couple of qualifications. First, explaining the success of a prior theory does not necessarily mean explaining everything it explained. Some of the questions answered by the prior theory may have been artefacts of the assumptions of the theory, so that when a new theory jettisons those assumptions it does not answer those questions but rather explains them away as pseudo-problems (as relativity treats 'what is the absolute velocity of the

earth?'). Second, some problems solved by the prior theory which are not thus explained away might remain unsolved by the new theory for some time; but until they are solved, perhaps by further novel developments of theory, the new theory cannot be regarded as having fully superseded the prior theory.

Thus, the intellectual pioneer creates an inconsistency which he then attempts to eliminate in a way that increases overall explanatory coherence. If he succeeds, the reflective disequilibrium he introduced was temporary and was a means to improved knowledge. The growth of knowledge requires an interplay of equilibrating and disequilibrating changes.

Rawls notes that the "kind of reflective equilibrium that one is concerned with in moral philosophy," involving the idea of being "presented with all possible descriptions to which one might plausibly conform one's judgments together with all relevant philosophical arguments for them," may lead to a "radical shift" in one's views; but he goes on to say that "[t]he most we can do" is to study traditional theories "and any further ones that occur to us" (Rawls 1999, 43). In contrast, a commitment to the growth of knowledge would urge not that we leisurely wait for new possibilities to occur to us, but rather that we *actively contrive* new theories to create a disequilibrium that may lead to a radical shift in our views.

Other recent philosophers seem to have even less appetite for revolutionary change. W. V. Quine and J. S. Ullian explicitly counsel conservatism: "In order to explain the happenings that we are inventing it to explain, the [new] hypothesis may have to conflict with some of our previous beliefs; but the fewer the better" (Quine & Ullian 1978, 66); "one is not apt to be tempted by a hypothesis that upsets prior beliefs when there is no need to resort to one" (Quine & Ullian 1978, 67). Quine and Ullian do allow that the maxim of conservatism can be set aside where a revolutionary new theory offers dramatic gains in simplicity or generality (cf. Quine & Ullian 1978, 75-76); but if theorists took the maxim of conservatism seriously, they would never take the time and trouble to work out a revolutionary new theory, so they would rarely, if ever, encounter a situation in which the maxim of conservatism could be set aside. It seems that Quine and Ullian see the purpose of enquiry as being to settle upon a coherent set of propositions with which we feel comfortable. But if our purpose in enquiry is to extend our knowledge, we should welcome, and encourage, the discomforts created by theoretic innovators.

4. Progress

Advocates of reflective equilibrium pay little attention to the pioneer and they usually incorporate no explicit prohibitions on ad hoc manoeuvres. Further, they portray the state of static equilibrium as an ideal. For example, Daniels (2011, sec. 1) says:

> We arrive at an optimal equilibrium when the component judgments, principles, and theories are ones we are un-inclined to revise any further because together they have the highest degree of acceptability or credibility for us.

David Lewis writes:

> Our "intuitions" are simply opinions; our philosophical theories are the same … and a reasonable goal for a philosopher is to bring them into equilibrium. Our common task is to find out what equilibria there are that can withstand examination, but it remains for each of us to come to rest at one or another of them. (Lewis 1983, x)

Rawls allows that it is doubtful whether one can ever reach the state of reflective equilibrium, but he still regards the state as a "philosophical ideal" (cf. Rawls 1999, 43-44). Geoffrey Sayre-McCord agrees (cf. Sayre-McCord 1996, 142).

In addition, accounts of reflective equilibrium are often combined with coherence theories of truth or of justification. Thus, Goodman opines:

> [D]eductive inferences are justified by their conformity to valid general rules, and … general rules are justified by their conformity to valid inferences. But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.* The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either. (Goodman 1983, 64)

Similarly, Rawls says that, in reflective equilibrium, "we have done what we can to render coherent and to justify our convictions… A conception of justice

cannot be deduced from self-evident premises or conditions on principles; instead, its justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view" (Rawls 1999, 18-19). Sayre-McCord agrees: "as one approaches a (wide) reflective equilibrium one thereby increases the extent to which the beliefs one holds are epistemically justified" (Sayre-McCord 1996, 143). Such accounts could be welcomed by a complacent dogmatist who is more concerned to 'get his story straight' than to better understand the world (for an effective critique of such accounts of justification see Stich 1998).

If, in contrast, our aim is the growth of knowledge, a rest-state of reflective equilibrium, far from being an ideal, is not even desirable. We want development, not stasis. If at some time we happened to achieve consistency and coherence in our accepted propositions, our next theoretical task should be to upset that equilibrium by seeking novel facts or paradoxical implications to refute some currently successful theory or by developing a novel theory to replace an existing one. Ideally, a reflective equilibrium would never be attained: progress toward a reflective equilibrium would always be upset by a new disequilibrating intervention followed by equilibrating efforts which are in turn challenged by further disequilibrating novelty; and so on indefinitely. Our ideal is a fruitful reflective disequilibrium that generates unending improvement in our knowledge by means of:

(i) active search for inconsistencies within our currently accepted knowledge;
(ii) pursuit of new facts and development of new hypotheses which are inconsistent with our current knowledge and that offer the prospect of radical change;
(iii) achievement of greater explanatory coherence by removing inconsistencies in ways which are not ad hoc and by supplanting an old theory by a new one when, but only when, the new one provides better explanations.

This dynamic conception of fruitful reflective disequilibrium is not fitted to provide an account of truth or of justification. If our aim is the growth of knowledge, and we recognise that new knowledge often contradicts previously accepted theories and observation statements, we should never maintain that our currently accepted propositions, however good a set they may make, are

either true or justified. We should view them always as more or less ephemeral steps in the progress of our knowledge. We should not seek to justify our theories: we should seek to replace them with better ones.

It might seem that advocates of reflective equilibrium could accommodate these criticisms. Few of them, if any, would claim to have reached the ideal of reflective equilibrium; they would regard any equilibrium currently attained as being justified only defeasibly and thus open to revision in the light of new discoveries. Further, given a choice between an ad hoc resolution of an inconsistency and an alternative resolution that teaches us something new, they would be expected to prefer the latter, other things being equal. That, of course, is true. The problem with reflective equilibrium theory is that it does not *encourage* the growth of knowledge: it does not *require*, or even *commend*, the active search for counterexamples, paradoxes, novel theories, and equilibrating adjustments that avoid ad hoc manoeuvres. As we have seen, its advocates discourage change, especially revolutionary change, and when they concede that a change is necessary, they are happy to accept an ad hoc change if it leads them back to a state of rest. Reflective equilibrium is for the shiftless, who are more interested in attempted justification than in improvement through criticism. Fruitful reflective disequilibrium is for those who are restless for the growth of knowledge:

> [S]cientific progress is revolutionary. Indeed, its motto could be that of Karl Marx: 'Revolution in permanence'. (Popper 1994, p. 12)

5. Moral knowledge

The dynamic conception of fruitful reflective disequilibrium was illustrated above with examples from empirical science, but it should also apply to moral enquiry. If the aim of such enquiry is to improve our understanding of moral matters, to extend our moral knowledge, then we should not want simply to achieve consistency and coherence in our particular moral judgements and general moral principles. We should rather strive for progress in moral enlightenment by insisting that:

(i) consistency is achieved in ways that are not ad hoc;
(ii) complacency (the search for justification) is abjured;

(iii) there is vigorous encouragement of attempts to increase disequilibrium by attempting to generate
- paradoxes within existing theory,
- inconsistencies between accepted theory and judgements about particular cases,
- novel theories which contradict accepted theories and which promise to have greater explanatory merit.

We noted above that reflective equilibrium theory includes none of those demands and is thus unfit to guide intellectual enquiry. Our illustrations showed that it is out of line with good practice in empirical science. How does it compare with the practice of moral enquiry?

Contemporary moral philosophers generally accept reflective equilibrium theory, but while their practice conforms to the theory in ignoring demands (i) and (ii), it diverges from it in partially meeting demand (iii), in two ways. First, moral philosophers often contrive imaginative thought-experiments, describing unusual possibilities, through which our moral theories can be re-shaped, or which yield intuitively acceptable moral judgements about particular cases against which moral principles can be tested. Rawls' original contractual position is such a thought-experiment (cf. Rawls 1999, 11-19, 102-168). Second, some moral philosophers are prepared to propose or consider revisionary moral theories, such as Stoicism, act-utilitarianism, egoism, feminist ethics, vegetarianism and so on, that contradict previously accepted moral principles and which generate moral judgements about particular cases that contradict previously accepted moral judgements about those cases. Thus, what moral philosophers actually do is better than what they say they should do.

Moral theorists would make a stride forward if, adhering to (ii), they insisted on progress in moral knowledge by refusing to accept that a moral theory or judgement can be justified and resolving instead to try to improve upon even the best moral theory that they have so far achieved. They would then reject Donald Davidson's claim that "we should expect people who are enlightened *and fully understand one another* to agree on their basic values" (Davidson 2004, 49). If, as I have argued, enlightenment involves an ongoing process of discovery, rather than being a state that can be (or is already) achieved, then, however enlightened people are, they should seek to become more enlightened, so they should encourage each other to dispute currently accepted moral propositions.

Moral theorists would make a further stride forward if, following (i), they insisted that amendments to theory which remove inconsistencies should not be ad hoc. Unfortunately, it is not easy to see how moral enquiry could generally meet the demand that acceptable amendments to theory should explain something new in addition to solving the problem for which they are proposed. In the empirical sciences, a new amendment to theory may entail novel factual propositions which can then be checked by observation, as in the examples given in section 2. However, that option for avoiding ad hoc manoeuvres is not available to moral theorists given the a priori character of moral enquiry. They would instead need to ensure that an amendment to theory explains a moral judgement or a moral principle that was not previously explained, in addition to resolving the difficulty for which it is proposed. That may seem to make the stricture against ad hoc manoeuvres very demanding, perhaps impossibly so.

There is another worry. Suppose that moral theorists could meet that demand. It might be that all the rival moral theories could, given time, be modified in non-ad-hoc ways to overcome internal inconsistencies. In that case, we face the prospect of numerous alternative moral theories, each in reflective equilibrium, each contradicting each of the others, and none having greater explanatory merit than any other, since each is satisfactorily explanatory in its own terms. Lacking an empirical test, we seem to have no way of rating the rival moral theories as epistemically better or worse. Yet, from a commonsensical point of view, it may seem clear that some moral theories are epistemically better than others. For example, suppose that a Buddhist moral theory and an Islamic Fundamentalist moral theory are each self-consistent and that each is modified in non-ad-hoc ways in response to judgements about unusual actual cases or imagined possible cases. On what has been said so far, we have no argument available for preferring one over the other. Yet, if we compare how people flourish or suffer in communities which adhere largely to the one moral theory vis-à-vis those which adhere largely to the other, we are presented with a stark contrast. That contrast seems highly relevant morally, which suggests that the point or function of morality is to facilitate the fulfilment of persons. Indeed, that idea is explicit in theistic accounts of morality that invoke God's plan for His creation. As Robert Young put it: "For many, including Judaeo-Christians, promoting the well-being of humans (and perhaps all sentient creatures) *is* the whole point, or a large part of the point, of having moral principles at all" (Young 1981, 162). The idea is not peculiar to theists: rule-consequentialist accounts of morality make the point explicitly, and act-consequentialist

accounts make it crudely; contractarian and contractualist accounts attribute morality to agreement, either for selfish benefit or for mutual benefit or for the benefit of all; and some evolutionary accounts explain morality as an adaptation that benefits the species. That suggests that we may be able, with some ingenuity, to render moral theories empirically testable, if we accept something like the following meta-ethical postulate:

(m) The true moral theory is the one such that the best prospects for the fulfilment of persons would be realised if that theory were to be adhered to universally by people as they actually are.

Different moral theories assign different rights and duties to people, and different assignments of rights and duties imply different social structures. For example, unequal rights between races or sexes imply institutions of slavery or the subordination of women (or men), and duties to avoid some specific types of sexual activity imply the social exclusion of homosexuals. An understanding of the effects on human fulfilment of people universally doing their duty as defined by a particular moral theory therefore requires a social-scientific investigation of the consequences of types of action in types of social circumstances. So, ascertaining the consequences for human fulfilment if a particular moral theory were universally acted upon is a matter for the social sciences.

The elaboration of that idea for rendering moral theories empirically testable, and thus better able to meet the challenge of non-ad-hoc development, would require a separate (book-length) discussion. Here, though, we can consider the objection that (m) is an ethical theory, rather than a meta-ethical postulate, because it is tantamount to the claim that the fulfilment of persons is the ultimate moral value. The objection is misplaced for a number of reasons. First, (m) does not imply that there is an ultimate moral value. It may be, for example, that the best prospects for the fulfilment of persons would be realised if people adhered to a moral theory which incorporated a plurality of moral values none of which is overriding or ultimate. In particular, the postulate (m) does not reduce to 'maximise human fulfilment,' any more than rule-consequentialism reduces to act-consequentialism (on the latter, see Frederick 2016, 25-26). Second, while it is true that (m) is contestable and may be rejected by some, it should elicit broad assent because it appears to be at least tacitly accepted by so many different moral theorists. An aspiration for universal assent

would plainly be unrealistic. Third, it may be instructive to compare (m) with the following 'meta-descriptive' postulate:

(d) Rival descriptions of reality can be evaluated epistemically by testing them against our sensory experiences.

Accepting (d) does not commit us to the claim that sensory experiences are the ultimate reality; and it does not reduce to the claim that descriptions of reality must use only observational terms. It does involve the descriptive claim that observations generally bring us into contact with reality; but that claim is assumed by the great majority of inquirers, so (d) should elicit broad assent, even though it is denied by some mystics and even by some rationalist philosophers, such as Parmenides.

6. Conclusion

It is a standard view in contemporary philosophy that our intellectual enquiries are, or should be, an endeavour to achieve a state of reflective equilibrium in which the propositions to which we subscribe are rendered consistent and explanatorily coherent. I have argued that that view is unacceptable if the aim of our enquiries is to extend our knowledge, because:

(i) it does not take sufficiently serious account of the fact that equilibrating moves may be ad hoc and thus fail to extend our knowledge;
(ii) it does not encourage the disequilibrating moves that are necessary for the growth of knowledge;
(iii) it regards a settled equilibrium, particularly one which is supposed to confer 'justification' on the propositions we accept, or on our acceptance of them, as an ideal, rather than as a deplorable state of stagnation.

Consequently, I have proposed that the method of reflective equilibrium should be replaced by a method of fruitful reflective disequilibrium which:

(i) demands moves toward reflective equilibrium which eschew ad hoc manoeuvres;

(ii) encourages moves toward reflective disequilibrium which generate inconsistencies either within accepted theories, or between accepted theories and accepted observation statements, or between accepted theories and potential new rivals;

(iii) abhors the idea of a static equilibrium and embraces the ideal of unending improvement.

The method of fruitful reflective disequilibrium is implicit in the progress of scientific knowledge, as I have indicated with some illustrations. That is not to say that all scientists practise it, let alone advocate it. It is to say only that, insofar as science has made genuine progress, which it does seem to have done, the method of fruitful reflective disequilibrium can be seen at work. In areas where science has stagnated, that might be explained by adherence to the method of reflective equilibrium.

Application of the method of fruitful reflective disequilibrium to moral theory should stimulate the pursuit of moral enlightenment; but there are some doubts as to how effectively that can be done insofar as moral enquiry is pursued a priori. I have suggested that better progress might be made if we rendered moral theories empirically testable by adopting a plausible meta-ethical postulate linking the truth of a moral theory with the consequences for human fulfilment that would follow upon universal adherence to the theory. We might then hope to see advances in moral knowledge akin to the advances in the empirical sciences.

The champions of reflective equilibrium could deflect my critique by saying that the aim of their enquiries is not to extend our knowledge. One would then be left to wonder what the value of their enquiries might be.[1]

References

DANIELS, N. (2011): Reflective Equilibrium. In: Zalta, E. (ed.): *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), available at: http://plato.stanford.edu/archives/spr2011/entries/reflective-equilibrium/

DAVIDSON, D. (2004): The Objectivity of Values. Reprinted in: Davidson, D.: *Problems of Rationality*. Oxford: Clarendon, 39-57.

[1] I thank an anonymous referee for this journal for objections and queries that helped me to clarify the argument of this paper.

FREDERICK, D. (2016): The Good Bishop and the Explanation of Political Authority. *De Ethica* 3, No. 2, 23-35.

GOODMAN, N. (1983): *Fact, Fiction, and Forecast*. 4th ed. Cambridge: Harvard University Press.

KUHN, T. (1957): *The Copernican Revolution*. Cambridge: Harvard University Press.

KUHN, T. (1977): A Function for Thought Experiments. Reprinted in: Kuhn, T.: *The Essential Tension*. Chicago: University of Chicago Press, 240-265.

LAKATOS, I. (1978): Newton's Effect on Scientific Standards. In: *Philosophical Papers*, *Vol. 1*. Worrall, J. & Currie, G. (eds.), Cambridge: Cambridge University Press, 193-222.

LEWIS, D. (1983): *Philosophical Papers*, *Vol. 1*. Oxford: Oxford University Press.

NEURATH, O. (1983): Protocol Statements. Reprinted in: Neurath, O.: *Philosophical Papers 1913–1946.* Cohen, R. S. & Neurath, M. (eds.), Dordrecht: Reidel, 91-99.

POPPER, K. (1959): *The Logic of Scientific Discovery*. London: Hutchinson.

POPPER, K. (1994): The Rationality of Scientific Revolutions. Reprinted in: Popper, K.: *The Myth of the Framework*. London: Routledge, 1-32.

POPPER, K. (1983): *Realism and the Aim of Science*. London: Routledge.

QUINE, W. V. (1951): Two Dogmas of Empiricism. *Philosophical Review* 60, No. 1, 20-43.

QUINE, W. V. & ULLIAN, J. S. (1978): *The Web of Belief*. 2nd ed. New York: McGraw-Hill.

RAWLS, J. (1999): *A Theory of Justice*. Revised ed. Cambridge: Belknap Press.

SAYRE-MCCORD, G. (1996): Coherentist Epistemology and Moral Theory. In: Sinnott-Armstrong, W. & Timmons, M. (eds.): *Moral Knowledge?* Oxford: Oxford University Press, 137-189.

STICH, S. (1998): Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity. In: DePaul, M. & Ramsey, W. (eds.): *Rethinking Intuition*. Lanham: Rowman and Littlefield, 95-112.

YOUNG, R. (1981): Theism and Morality. Reprinted in: Helm, P. (ed.): *Divine Commands and Morality*. Oxford: Oxford University Press, 154-164.

Organon F 23 (4) 2016: 460-487

Logic and Rational Requirements

DIEGO TAJER

University of Buenos Aires – Conicet
Puan 480. 1406 Buenos Aires. Argentina
diegotajer@gmail.com

RECEIVED: 31-04-2016 • ACCEPTED: 30-06-2016

ABSTRACT: In this paper, I discuss the relation between logic and rationality. I develop (formally and conceptually) a rational requirement which can respond to the classic objections by Harman (1986). On the one hand, the requirement pays attention to the *relevance* of the premises and the conclusion, which is formally expressed by the notion of weak relative closure. The requirement also takes care of the *complexity* of the inferences. This notion of complexity is formally represented by a partially ordered scale of the difficulty of inferences, which is weaker than the notion of complexity as number of steps.

KEYWORDS: Logic and rationality – normativity of logic – rational requirements.

1. Introduction

In this paper, I discuss the relation between logic and rationality. The notion of rationality is too complex, so I will just focus on some aspects which are relevant to the discussion. I am interested, following Brome (1999) and MacFarlane (2004), on developing a specific *rational requirement* for logic. Rational requirements are statements which express what rationality asks from us with respect to a certain epistemic or practical issue.

Many authors, such as Broome (1999), Kolodny (2005), Way (2010), and Shpall (2013), discussed which was the best way of expressing rational re-

quirements. One of the main issues that this discussion introduced is the difference between a wide scope and a narrow scope for the rationality operator ("rationality requires that …").[1] Taking R as a rationality operator, the narrow scope principles have the form $A \rightarrow R\mathrm{B}$, while the wide scope principles have the form $R(A \rightarrow B)$. For example, many authors have discussed the following rational requirements (where "WS" means Wide Scope and "NS" means Narrow Scope):

(NS Evidence) If you believe that there is conclusive evidence that p, then rationality requires you to believe p.

(WS Evidence) Rationality requires that (you do not believe that there is conclusive evidence that p, or you believe p)—cf. Kolodny (2005, 521).

(NS Enkrasia) If you believe that you ought to do F, and you believe that you can do F, then rationality requires you to intend to do F.

(WS Enkrasia) Rationality requires that (either you don't believe that you ought to do F, or you don't believe that you can do F, or you intend to do F)—cf. Broome (2014, 171).

Here I will not focus on the precise formulation of each non-logical rational requirement. I will take for granted that some of these pragmatic or epistemic requirements are indeed true. In this paper, I will focus on the following two possible requirements of logical rationality. The first one has narrow scope, while the second one has wide scope:

(NS Validity) If $\Gamma \vDash \mathrm{A}$, then if you believe Γ, rationality requires you to believe A.

(WS Validity) If $\Gamma \vDash \mathrm{A}$, then rationality requires that (you do not believe some sentence of Γ, or you believe A).[2]

---

[1] Strictly speaking, most authors in this discussion use deontic operators such as "should …" or "has reasons to …". Following Broome (1999; 2014), I prefer to use rationality operator, and then discuss whether the requirements can be read as duties or reasons. On the other hand, a rationality *predicate* could also be used instead of an operator, but it does not give more clarity to the discussion (for example, we would need to add sentence names, etc.).

[2] As two anonymous referees observed, some authors such as MacFarlane have argued for this negative requirement: if Γ implies A, then rationality forbids you to believe

The aim of this paper is to develop a new requirement for logical rationality, which will be based on WS Validity. But before going into this, it is convenient to say some words about another discussion, which is sometimes taken as more fundamental: should we be rational; or in other words, is rationality normative?

There are different arguments in favor and against the idea that rationality is necessarily normative. In general, the arguments in favor can be Kantian or utilitarian. Kantians consider that rationality is a fundamental aspect of the human being, and as such, it is certainly normative (cf. Southwood 2008). Utilitarians claim that following rational requirements leads us to taking better decisions or believing true propositions. For example, Joyce (1998) appeals to accuracy arguments to justify the rational requirement of epistemic coherence.[3]

Arguments against the normativity of rationality usually point out two things. First, that rationality can lead us to taking wrong decisions or believing false propositions (see Kolodny 2005). Second, that in cases in which rationality takes us "closer" to the right action or the true belief, it is superfluous, since it can be subsumed under other requirements. For example, the rational requirement of epistemic coherence can be subsumed under the evidential norm of believing what the evidence suggests (cf. Kolodny 2008).

Γ and $\neg A$. Under some plausible assumptions (such as Explosion), WS Validity gives a similar result: it forbids you to believe $\Gamma \cup \{\neg A\}$ and disbelieve something else (i.e. there is no rational way of believing $\Gamma \cup \{\neg A\}$ without going fully trivial). If we add a requirement of non-triviality, then WS Validity implies MacFarlane's negative requirement.

Now, why is WS Validity better than MacFarlane's negative requirement? Because the negative requirement is just equivalent to a consistency requirement. But there is something more to tell about logical rationality: if someone believes that Canada is a country, and that every country has a capital city, but does not yet believe that Canada has a capital city, there is *something wrong* with the belief set of this person. Ignoring the obvious consequences of your beliefs is something to be criticized. WS Validity (unlike MacFarlane's negative requirement) can point out this kind of mistake.

[3] Joyce (1998) shows that an incoherent probability distribution (i.e. one which does not correspond to the probability calculus) is necessarily "dominated" by a coherent distribution. This means that the coherent one will be closer to the truth in every possible world.

The word "rationality", and similarly the word "normativity", have been used to name different things. This is why the discussion on the normativity of rationality is often confusing. To be clear, I can specify what I mean when I say "normativity of rationality":

(Normativity of rationality)
If rationality requires you to do *F*, then you *ought* to do *F*.[4]

In this paper, I will not take a stance on the normativity of rationality. I will be interested, mainly, in formulating a requirement of logical rationality. Moreover, given my suspension of judgment on the normativity of rationality, I will offer logical rational requirements that are *compatible* with the possible normativity of rationality. In other words, I will provide some requirements such that, if rationality requires you to do *F*, assuming that you *ought* to do *F* does not lead to inconsistency.[5]

2. Scope and normativity

In this section I will explore the problem of the scope of logical rational requirements. In particular, I will mention the Bootstrapping objection (which affects the narrow scope requirements) and discuss which kind of normative force corresponds to a logical rational requirement.

---

[4] A referee observed that this definition does not clarify the concept of normativity. Admittedly, it is not a proper analysis of normativity, but rather a semantic clarification. Normativity is understood in many different ways across the literature; for example, it can be applied to meaning or content (cf. Boghossian 2003). Even the normativity of rationality can be understood in more inflationary ways (see Southwood 2008). The identification between "normativity" and "ought" is usual but not trivial; thus, the semantic clarification could be useful for some readers.

[5] This methodology was also adopted by Broome (2014, chap. 11). A referee observed that this definition makes the requirements incompatible with the *non-normativity* of rationality. This is true in one respect: the requirements cannot logically imply the non-normativity of rationality. If this were the case, they would be *a priori* incompatible with the normativity of rationality. According to the methodological principle I adopted, rationality could be normative or non-normative, but this should not be a logical consequence of rational requirements.

2.1. Bootstrapping

The Bootstrapping objection has frequently been raised against the normativity of rationality, although it affects mainly the narrow scope formulations. This problem can be expressed in this way, schematically:

> (Bootstrapping)
> Suppose that the requirement *r* of rationality has narrow scope, i.e. it has the form "If you have the attitude X, then rationality requires you to have the attitude Y". Suppose that you ought not to have the attitude Y. Now, in case you have the attitude X, rationality requires you to have the attitude Y anyway. If rationality is normative, then in this case you *ought* to have the attitude Y, which by hypothesis we assumed you ought *not* to have.

Until now, the setup was rather abstract. But we can illustrate the problem with some clear examples:

> (Bootstrapping for NS Enkrasia)
> For unjustified reasons, you believe that you ought to kill your son (and you believe you can do it). Therefore, if rationality is normative and NS Enkrasia holds, you ought to intend to kill your son. But obviously you ought not to do it.

> (Bootstrapping for NS Evidence)
> For unjustified reasons, you believe that there is conclusive evidence that the world is squared. Therefore, if NS Evidence holds and rationality is normative, you ought to believe that the world is squared. But obviously you ought not to believe it.

> (Bootstrapping for NS Validity)
> For unjustified reasons, you believe *p*. Given that logic is reflexive (i.e. *p* implies *p*), if NS Validity is normative, you ought to believe *p*. The same can be reproduced for any of your beliefs: you ought to believe everything you believe. But this is absurd.

Most authors in this discussion consider that Boostrapping rules out the normativity of narrow scope requirements.[6] Fortunately, the *wide scope* version

[6] For reasons of space, I will not discuss the positions which defend the narrow scope requirements against the Bootstrapping objection. See Schroeder (2009).

of those requirements is immune to the Bootstrapping objection. In those cases, rationality gives the option of having or lacking some attitudes, but it does not require adopting a specific attitude. For example, WS Enkrasia requires you not to believe that you ought to do *F*, or not to believe that you can do *F*, or to intend to do *F*.

The same holds for WS Validity: it just requires that, if Γ implies A, you do not believe some sentences of Γ, or you believe A. Moreover, WS Validity can respond to one objection by Harman (1986, 11). Harman observed that, even though A and A → B imply B, sometimes we believe A and A → B but we ought not to believe B (for example, when B is false). WS Validity is not affected by this problem, for it does not require believing B in this case, but it gives the option of revising the belief in the premises.

Therefore, it is promising to adopt the Wide Scope rational requirements, which may provide *duties*. In other words, there might be duties of complying with disjunctive requirements such as WS Validity, WS Enkrasia, and WS Evidence, among others. In what follows, we will focus on these wide scoped rational requirements, for they are compatible with the normativity of rationality.

2.2. Strict normativity?

In the last paragraph, we argued that rationality could be normative. But can this normativity be *strict*? I will claim that no specific requirement (including the logical requirements) can be strictly normative. Reisner (2011) developed some mental experiments in order to prove this point. This is the clearest one:

> (Reisner case)
> Suppose that a millionaire makes the following bet with you: he gives you billions of dollars in case you believe *p* and you don't believe (*p* or *q*). With that money, you could and would feed all the hungry people in the world.

According to Reisner (and I share his intuitions), in this case you ought *all things considered* to adopt a belief set which includes *p* but not (*p* or *q*). As Reisner (2011, 41) claims, "it would be quite hard to explain how it is that saving all the starving people in the world does not have deontic or normative priority over violating a principle of rationality". Therefore, you ought to violate WS Validity.

This shows that the normativity of WS Validity cannot be strict, but weak or defeasible. The duty of being logically rational (if there is such a duty) can be defeated by a different duty. MacFarlane (2004) was the first one to hold this idea. He exemplified this problem with the Preface Paradox (cf. Makinson 1965), in which a person can have an inconsistent but rational belief set. According to MacFarlane, this is a case in which a logical rational requirement conflicts with a more global epistemic requirement, and the last one dominates. Here I will adopt the same approach with respect to these conflicting cases: these cases do not show the absence of normative force in logical rational requirements, but their defeasible nature. It is worth remarking that, even admitting the defeasible normativity of logical rationality, it is still better to adopt a Wide Scope requirement than a Narrow Scope one. Given the bootstrapping problem, a Narrow Scope requirement would be defeated in every case in which I have a false belief; on the contrary, a Wide Scope requirement only fails in very specific cases such as the Preface Paradox or the imaginary Reisner cases.

Finally, even though I reject the possibility of a rational requirement with strict normativity, I admit the possibility that rationality, taken as a global property, could be strictly normative. If rationality is taken as a property which emerges from the fulfillment of different requirements (epistemic or practical), the strict normativity of this "global" rationality cannot be so easily ruled out.

3. Relevance

Until now, I have argued for a wide scope requirement, and I claimed it could possess a defeasible normativity. But the wide scope requirement I advocated for, WS Validity, is still affected by many problems.

The first one was described by Harman (1986, 12). Harman observes that we may intuitively ignore some *irrelevant* consequences of our beliefs. According to WS Validity, it is irrational to believe "it rains" and not to believe "it rains or $2 + 2 = 6$", "it rains or it is Tuesday", "it rains or $2 + 3 = 4$", and many other completely irrelevant sentences. But, according to Harman, this attitude is rational, since believing *all* the consequences of your beliefs would make you lose time, energy and mental space in many strange, trivial or irrelevant beliefs.

My way of solving the problem of irrelevant consequences is to add a clause to WS Validity, which specifies that the premises and the conclusion must be contextually relevant. As we will see, my specification is similar (but different) to the proposals by Broome (2014, 157) and Steinberger (2015, 25).

Broome specifies that the rational requirement holds whenever the agent "cares about the conclusion" (Broome 2014, 157). A paraphrase of his position is the following:[7]

> (WS Validity – Broome)
> If Γ implies A, and you care whether A, then rationality requires you not to believe some sentence of Γ, or to believe A.

This principle holds in most cases, but it is affected by some problems. One of them is the exaggerated subjectivity of the notion of *care*. Suppose that I hear the fire alarms, and I know that if the fire alarms sound, then the house is burning. But still, my belief set does not include "the house is burning", since I don't care about this in this particular moment (suppose I am writing a difficult article on logic, and all my attention is focused on that). Intuitively, my attitude is irrational, but Broome's notion takes it as rational, for in that case I don't *care* about the conclusion.

Steinberger (2015, 25) solves this point, for he changes the specification and he introduces the idea of "having reasons to consider the conclusion". He suggests the following requirement:[8]

> (WS Validity – Steinberger)
> If Γ ⊨ A, and you *consider* or *have reasons to consider* A, then rationality requires you not to believe some sentence of Γ, or to believe A.

Here, he makes room for a disjunctive notion between a subjective aspect (to actually consider) and an objective aspect (to have reasons to consider). In the

---

[7] Broome applies this idea to WS Modus Ponens, not to WS Validity. However, his considerations about relevance do not depend on that.

[8] This is not exactly Steinberger's formulation. His requirement also includes, as we will see, the fact that the agent believes that the inference is valid. However, for the purpose of this section, I ignore that aspect of the requirement (I leave it for the next section).

previous example, the agent does not consider the conclusion, but clearly has reasons to consider it.

Anyway, Steinberger's proposal is still affected by a problem, which is the emphasis on the *conclusion*. This is unreasonable. Suppose that my belief set includes a remote and complicated inconsistent set. For example, I believe in the axioms of naïve set theory, which I learned at primary school. But in the context, the discussion is focused on something completely unrelated, say, the size of the countries. In that context, I consider the proposition "Spain is larger than France", though I reject it. To put it simpler, my belief set is *Naïve set theory* ∪ {Spain is *not* larger than France}. According to Steinberger's notion, logical rationality does not permit me to be in that state, given that my remote inconsistent beliefs also imply "Spain is larger than France" (by Explosion).[9] In other words, given that I believe the axioms of Naïve set theory (which is inconsistent), and I consider "Spain is larger than France", I must also believe that sentence. This is not completely unjustified (after all, logical rationality does not permit me to have inconsistent beliefs), but it is clearly inadequate if the *relevance* of logical requirements is taken into account. For the inconsistent set I remotely believe is absolutely irrelevant in the context.

My position makes a modification to solve this problem, where both premises and conclusion must be relevant in the context. In other words, I will adopt Steinberger's notion of relevance, but also extended for the premises:

> (WS Validity + Relevance)
> If $\Gamma \vDash A$, and Γ and A are relevant in the context, then rationality requires you not to believe some sentence of Γ, or to believe A.

Following Steinberger, I will define "relevance in the context" in the following disjunctive way:

---

[9] As a referee observed, it is possible to avoid this problem by rejecting Explosion and adopting a paraconsistent logic. However, moving towards a paraconsistent logic such as *LP* or *FDE* has very strong consequences: it means rejecting the rational force of very intuitive rules such as disjunctive syllogism or modus ponens. I also think that my description of the situation is more accurate; the problem is not the background logic but the excessive demands of the ideal rational requirements.

(Relevance)
In the context *c*, the sentence *p* is relevant for agent *i* iff *i* considers or has reasons to consider *p*.[10]

In the fire alarms example, my beliefs are {the fire alarms are sounding; if the fire alarms sound, the house is burning}, and I have reasons to consider the belief {the house is burning}. Then, according to my criterion, the set of relevant propositions in the context is the union of these sets, say: {the fire alarms are sounding; if the fire alarms sound, the house is burning; the house is burning}. In this case I am violating WS Validity+Relevance: my belief set is not closed relatively to the set of relevant propositions.

Instead, in the case of naïve set theory as irrelevant belief, my belief set is *Naïve set theory* ∪ {France is not larger than Spain}, and it complies with WS Validity+Relevance. For, even though "France is larger than Spain" is relevant and can be deduced from my belief set, it cannot be deduced from my set of *relevant* beliefs. It is worth mentioning that the axioms of naïve set theory are irrelevant in this context for, even though I believe them, I am not considering them and I don't have any reason to do it.

3.1. Relevance: a formal approach

In what follows, I will formally develop the notion of relevance that I introduced in the last paragraph. In order to do it, I will use the concept of a consequence operator, which is widely used in non-classical logics and belief revision theories.[11]

In the literature on belief revision, it is usual to presuppose that the belief set is closed under logical consequence. Formally, there is an operator *Cn* which takes a set of sentences and gives as output the set of its logical consequences. In other words, $Cn(X) = \{A \mid X \vDash A\}$. A consequence operator is *Tarskian* iff it satisfies these three conditions:

(Inclusion) If $a \in X$, then $a \in Cn(X)$

[10] The reader can notice that I introduced the notion of *relevance* in the requirement, and then I defined it. Strictly speaking, I could have introduced the defined notion from the beginning. I presented the requirement in this way for simplicity.

[11] See Hansson (1999) for a complete introduction to belief revision theories, and Wójcicki (1988) for a classic monograph on consequence operators.

(Idempotence) $Cn(X) = Cn(Cn(X))$
(Monotony) If $X \subseteq Y$, then $Cn(X) \subseteq Cn(Y)$

The most popular logics (classical, intuitionistic, relevant, etc.) can be represented with a Tarskian operator, for they are structural (i.e. they satisfy monotony, reflexivity and cut). Belief revision theories usually take the belief set X to be closed under consequence, i.e. $Cn(X) = X$.

However, we have seen that it is exaggerated to ask a real individual to have a closed belief set. A non-closed belief set can be adequate, when the consequences of the beliefs which are not included in the set, or the sentences which work as premises, are irrelevant. Now I will try to give a formal characterization of these conceptual aspects of relevance.

Relative closure

In order to formally characterize the notion of relevance, I have to start from the definition of a *context*. As I said before, the evaluation of a belief set takes place in a context. The set Δ of relevant propositions is the set of propositions which the agent considers or has reasons to consider in a particular context.

By now, the only restriction on Δ is the following:

(Closure under negation)
If $A \in \Delta$, then $\neg A \in \Delta$

This cannot be so problematic. If a sentence is relevant in a context, its negation must also be relevant.[12] In general we will make a simplification to avoid Δ being necessarily infinite: we will allow the cancellation of double negations. So, if A and $\neg A$ belong to Δ, then $\neg\neg A$ may not be in Δ.[13] We will use the symbol $\pm$ to simplify, where $\pm\Gamma = \Gamma \cup \{\neg\gamma \mid \gamma \in \Gamma\}$.

---

[12] This condition is similar to the closure under negation in judgement aggregation. See List (2012) for an introduction to this area of research.

[13] This property is adequate in cases where A is equivalent to $\neg\neg A$, such as classical logic, K3, LP, FDE, etc. It may be not entirely adequate for intuitionistic logicians, for they do not regard A and $\neg\neg A$ as equivalent. However, this simplification may be dropped for philosophical reasons and all the essential features of the proposal would remain the same.

Another restriction over the contexts that could be adopted is taking every set Δ to be closed under subformulas:

(Closure under subformulas)
If $A \in \Delta$, and B is a subformula of A, then $B \in \Delta$.

This is fairly intuitive too. If "it rains and it is Wednesday" is relevant, then "it rains" and "it is Wednesday" are relevant. The same should apply to the other connectives.

I will use a notion from belief revision theory (see Hansson 1999, 32), which is the concept of relative closure. A set Γ is closed relative to Δ iff Γ contains all the consequences of Γ that also belong to Δ. Formally:

(Relative closure)
A set X is closed relative to a set Δ iff $Cn(X) \cap \Delta \subseteq X$.

For example, the set $\{p\}$ is not logically closed relative to $\{p, q, q \rightarrow p\}$, for $\{p\}$ does not include the sentence $q \rightarrow p$, which can be inferred from $\{p\}$. Instead, the set $\{p\}$ is closed relative to $\{p, q, q \vee r\}$, for $\{p\}$ includes all the consequences of $\{p\}$ that are included in $\{p, q, q \vee r\}$ (i.e it includes itself).

Let's see how this concept can be applied to more concrete cases. Suppose that the context in question is a football match, and the agent has the following belief set:

$\Gamma =$ {Messi will score a goal;
If Messi scores a goal, Barcelona will win}

In the context *c* of a football match (and of course, with a considerable amount of simplification), suppose that the set of relevant propositions is the following:

$\Delta =$ ±{Messi will score a goal;
If Messi scores a goal, Barcelona will win;
Barcelona will win;
Neymar is playing with number 5}

Here, the belief set Γ has a clear shortcoming: it includes a set of beliefs which are relevant in the context, but it does not include one relevant consequence of these beliefs ("Barcelona will win"). Formally we can say that the set Γ of the example is not closed relative to Δ.

Let's compare Γ with the following set:

Γ* = {Messi will score a goal;
If Messi scores a goal, Barcelona will win;
Barcelona will win}

This new set, unlike Γ, is closed relative to Δ. The same would happen to the following set:

Γ** = {Messi will score a goal}

Even though Γ** includes fewer elements than Γ, it is closed relative to Δ. This shows that the way of reaching a closed belief set is not necessarily to accumulate beliefs, but also to abandon beliefs when it is necessary.[14]

Weak relative closure

The notion of relative closure is much more realistic than the ideal notion of closure. However, it still has a shortcoming (that we mentioned in the previous part, from a conceptual point of view). Suppose that our belief set is inconsistent with respect to a completely irrelevant topic. Just to follow with the previous example, take the set of relevant propositions as Δ, but now the belief set is:

Γ′ = *Naïve set Theory* ∪ {Messi will score a goal}

Intuitively, this set should be taken as contextually adequate. Even though it includes an inconsistent belief, the inconsistency is not relevant in the context (since we are not considering it, and we have no reasons to do it in the context). With respect to the relevant propositions, Γ′ is actually closed.

However, following the previous definition of a relatively closed set, the set Γ′ is not closed relative to Δ, for "Neymar plays with the number 5" (and any other sentence in Δ) can be inferred from Γ′ by Explosion; but Γ′ does not contain that sentence. A way of solving this problem is to adopt a weaker notion of relative closure:[15]

---

[14] The concept of relative closure, like the concept of closure, is synchronic. In other words, it does not guide the processes, but it evaluates states.

[15] It is worth mentioning that there are many ways of solving this problem. Restall & Slaney (1995) use a background paraconsistent logic, so that an inconsistent belief does not make the set trivial. However, this proposal pays high costs: the paraconsistent logic

(Weak relative closure)
A set X is weakly closed relative to a set Δ iff $Cn(X \cap \Delta) \cap \Delta \subseteq X$.

Indeed, the set Γ′ is not closed relative to Δ, but it is weakly closed relative to Δ. For even though Γ′ does not include all the relevant consequences of its members, it does contain all the relevant consequences of its *relevant* members. This notion of closure is stronger than the previous one, and allows us to formally define a rational requirement which pays attention to the contextual relevance. Indeed, we can translate WS Validity+Relevance as the following requirement:

(WS Validity+Relevance – Formal)
Rationality requires your belief set to be contextually adequate; i.e. when the contextually relevant propositions are Δ, your belief set X must be such that $Cn(X \cap \Delta) \cap \Delta \subseteq X$.

This requirement asks the set to include the relevant consequences of those propositions that were initially relevant. This avoids the intermission of irrelevant beliefs that might imply other relevant propositions. In a nutshell, I argued that introducing the notion of weak relative closure in the logical rational requirement can respond adequately to Harman's objection.

4. Excessive demands

4.1. The objection and the first answers

The second important objection against WS Validity was also expressed by Harman (1986, 17):

(Excessive demands)
It is rational to ignore the least obvious consequences of our beliefs. For example, one can believe the Peano axioms and not believe some of its consequences, without being irrational.

This objection has received many answers.

---

they use (*FDE*) does not admit Modus Ponens, so the demands over the agents are considerably low.

The first answer, and probably the least interesting one, just denies the problem. According to this view, if we assign a belief set to an agent, and we represent the set as a set of possible worlds, logical closure will follow. This answer was formulated by Stalnaker (1987), though it is not particularly strong. In theories of rationality, idealizations are frequent. However, this does not mean that we do idealize at this point when we attribute beliefs. If that were the case, we would not understand how can someone ignore the consequences of her beliefs.

The second answer to Excessive Demands admits that sometimes we do not comply with logical closure. But it claims that logical rationality is an *ideal* condition, and as such, there is always some level of irrationality if you believe the Peano axioms but you ignore some of their consequences. This is, for example, Broome's first position (1999), and one of the proposals of MacFarlane (2004).

It is convenient at this point to introduce the important distinction between *ideal rationality* and *applied rationality*.[16] *Ideal* rationality is a set of requirements that can be used as a point of reference, or regulative ideal, for dealing with our beliefs or evaluating the beliefs of the others. According to this kind of rationality, the objection of excessive demands does not apply, since even when no agent can comply perfectly with the closure requirement, one can evaluate how close is her belief-set to the ideal. Every agent should, in any case, take logical closure as a point of reference.

On the other hand, *applied* rationality is a set of requirements that we use ordinarily to evaluate real agents and classify them as rational or irrational. Undoubtedly, logical closure is too demanding in this respect, for we do not classify an agent as irrational when she ignores the last consequences of her beliefs. It shall be clear that, in this paper, I am looking for a requirement of *applied* rationality. Therefore, this second answer to the problem of excessive demands is not useful for my purpose.

4.2. Epistemic variations

A common response to the problem of excessive demands, which was anticipated by Harman (1986, 17) and suggested by Field (2009, 253) and Steinberger (2015, 25), holds that logical rational requirements apply just in cases

[16] See Smithies (2015) for a defense of the distinction between ideal and applied rationality, or between "ordinary standards" and "ideal standards" for rationality.

in which the agent *recognizes*[17] that the premises logically imply the conclusion. In other words, their proposal is to replace WS Validity by the following requirement (the considerations about relevance that were introduced in the previous section will be ignored by now):

> (Recognized WS Validity)
> If you *recognize* that $\Gamma \vDash A$, then rationality requires you not to believe some sentences of Γ or to believe A.

This epistemic variation of WS Validity has, nevertheless, a clear shortcoming. There are some *obvious* cases of validity, which must have rational force even when one does not recognize them. In other words, it seems that, even though not closing your belief set under *recognized* consequences is wrong, it is also wrong not to believe some simple consequences of your beliefs. In what follows, I will present some variations of the requirement that are immune to this objection.

4.3. Objective and inferential scales

Adopting epistemic variations is not the only way of restricting the range of application of rational requirements to a subset of valid inferences. It is possible to develop more objective restrictions, based on the level of difficulty. If this strategy is adopted, the subset of inferences with normative force will not be the *recognized* inferences, but the *simple* inferences, according to a certain scale. The idea is that the agent must have her belief set closed under some simple inferences, but not necessarily under more complex inferences.

A usual strategy for restricting the requirements to "simple" cases of validity is to consider that the complexity of an inference can be measured by the *number of steps* that you need to prove its validity (i.e. the *length* of its shortest proof). Some authors, such as Field (2009), D'Agostino & Floridi (2009) and Jago (2009) have proposed ideas of this kind.[18] In this case, rationality could

---

[17] It is hard to mention all the subtleties of each position. Strictly speaking, the word "recognize" comes from Harman. Field uses "realizes" and Steinberger "believes".

[18] Actually, Field is the only one who appeals to inferential measures for logical rational requirements. D'Agostino & Floridi, and also Jago, just try to establish a reasonable measure of the complexity of an inference.

require that the agents believe what can be derived from their beliefs in a certain (small) number of applications of rules:

> (WS Validity – Proof-length criterion)
> If $\Gamma \vDash A$ can be proved in at most k applications of rules, then rationality requires you not to believe some sentence of Γ, or to believe A.

The proof-length criterion is initially plausible (in fact, I will apply a similar one). However, it faces several objections. The first one was observed by Field (2009, 260): there is no way of complying with this requirement without having a logically closed belief set. For, imagine a set which is closed under *one* application of rules. Could it be non-closed under *two* applications? Certainly not; if the first thing happens, every number of steps can be reached one by one.

Anyway, we may focus on what does the requirement ask in each case. And effectively, with respect to certain initial non-closed set, this requirement can point out which beliefs are we to blame for not adopting. For example, if you believe just p and $p \rightarrow q$, you can be blamed for not believing q,[19] but you cannot be blamed for not believing $\neg\neg\neg(\neg q \vee \neg(r \vee q))$.

However, even though this variation is promising, there is another important and not so commonly observed problem: the number of applications of rules is not a correct measure of the complexity of an inference. In fact, suppose that an agent has the beliefs $p_1, \ldots, p_{120}$. Intuitively, it is easy for the agent to infer $p_1 \wedge \ldots \wedge p_{120}$. However, this involves 120 applications of rules. Now, suppose that the agent believes in the two axioms of Naïve set theory. It is possible to prove a contradiction from them in a few steps, but it is not an easy proof (the intelligence of Russell was needed to find the proof). According to the inferential approach, finding a contradiction in Naïve set theory is much easier than introducing 100 conjunctions. In this sense, the inferential criterion

---

[19] I am using an *explicit* notion of belief: a type of belief that, when you have it, you know that you have it. Using an implicit notion of belief may be useful for other discussions, but it would obscure this particular discussion. For, according to the usual concept of "implicit belief", we implicitly believe the obvious consequences of our beliefs; in this way, a considerable part of the problem of deductive closure would be automatically (and artificially) solved. I find much more illuminating to *explain* (rather than to rule out) the failure of deductive closure; that's why my notion of belief is explicit.

assigns difficulty to simple inferences, and takes some difficult inferences as easy. Therefore, an inferential criterion cannot establish a good measure of complexity.

It is tempting to adopt a more skeptical position in this debate, and to claim that the task of finding a scale of inferential difficulty is impossible. This reaction is somewhat justified. Harman (1986, 3) draws a distinction between inference and reasoning. Logical *inferences* are cases in which an agent arrives to a conclusion from certain premises, by using a set of rules. Instead, in a process of *reasoning* an agent arrives to a conclusion from certain premises by different informal methods, such as mental maps, rational intuition, suppositions, implicit "logical rules", etc. Logical inferences, given their precision, are measurable, and therefore can be ordered by complexity. But pieces of reasoning are not so precise.

Anyway, there is no strong reason to embrace skepticism at this point. Even though reasoning does not psychologically work as a logical apparatus, there are certain similarities. There are clear cases of simple or complex beliefs that are classified as such by both perspectives (logical and psychological). So, even if there are many functional differences, the level of "intuitive" difficulty of an inference hopefully may be formally captured, as well.

My proposal in the next paragraph will take some elements from the proof-length approach. I will develop a formal theory which can be used as a measure of complexity. Unlike the inferential approach, which establishes a total order, my proposal will establish a partial order, where some inferences are necessarily more difficult than others.

4.4. From recognized to recognizable

Conceptually, the restriction I will adopt has an element of subjectivity, but not as strong as in Recognized Validity. It is clear that, for each agent, some cases of logical consequence are *recognizable* and some are not. What is recognizable for each agent depends on her inferential capacity.[20] This allows to restrict the requirement in the following way:

---

[20] Anyway, the two axioms of recognizability I introduce later are compatible with an objective interpretation of the criterion, where every agent could recognize the same set of inferences.

(Recognizable WS Validity)
If you *could* recognize that $\Gamma \vDash A$, then rationality requires you not to believe some sentence of Γ or to believe A.

It is important to remark that my proposal is potential. This means that the agent is not limited by the inferences she actually recognizes, but by the inferences she *could* recognize. The modal element depends on the inferential capacity of the agent. In this sense, logic has a rational force over the agent no matter the actual logical knowledge she has; the only thing that matters is the knowledge she *could* have. Admittedly, potential notions such as "could know" are not completely clear. However, they describe the rational requirements in a much more accurate way. In ordinary talk, we usually appeal to *abilities*: for example, someone is responsible for not avoiding the death of another person whenever she had the *ability* or the *possibility* to save the other person. The same applies to logical rational requirements: if you are able to realize that A implies B, you can be criticized for believing A and not believing B; but you cannot be criticized for believing the Peano Axioms and ignoring whether the Goldbach conjecture is true or false. What does "ability" precisely mean is still an ongoing debate in philosophy, and is far from the scope of this paper. In what follows, I will provide some axioms which, at least for the cases of logical recognizability, help to make the notion more precise.

Now I will introduce some precisions on the notion of recognizability. In the last section I rejected the proof-length approach, for it doesn't give a correct analysis of the difficulty of inferences. My proposal is based on two conditions that every set of recognizable inferences should satisfy:

(Set of recognizable inferences)
Let R_i be the set of recognizable inferences for an agent *i*. R_i should satisfy the following two properties:

(Reflexivity) If $A \in \Gamma$, then $(\Gamma \vDash A) \in R_i$.

(Order) If every proof of $\Gamma \vDash A$ includes a proof of $\Gamma \vDash B$, then: $(\Gamma \vDash A) \in R_i$ only if $(\Gamma \vDash B) \in R_i$.

This definition establishes that if you can recognize that $p, q, r \vDash (p \wedge q) \wedge r$, then you can also recognize that $p, q, r \vDash (p \wedge q)$. For, in order to prove the first case of validity, you must be able to prove the second one. But it does not establish any previous relation between "incomparable" inferences: for example, between the inference from $p_1, p_2, \ldots, p_{100}$ to $p_1 \wedge p_2, \ldots \wedge p_{100}$ and the

inference from Naïve Set theory to absurdity, even though the latter inference is shorter than the former. In this way, it avoids the objection I presented against the proof-length approach.

An anonymous referee observed that this criterion is still arbitrary, since there is no principled way of establishing the set of recognizable inferences for an agent. Admittedly, my criterion has some degree of arbitrariness; but it is still better than the proof-length perspective, for it does not suppose that every agent has a numerical limit *k* (a very unrealistic assumption). The set of recognizable inferences depends on each agent; the two axioms I provided give some restrictions on the structure of this set. Unlike the proof-length approach, my two axioms are intuitive and realistic. They are still too weak to determine *a priori* what an agent can recognize. But this is not necessarily a problem. The recognizability set can be thought as analogous to a possible world: there is no logical way of determining *a priori* what is true in a possible world, but there are some structural restrictions that every possible world satisfies.

In order to make Order more precise, it is necessary to specify some proof method. As I claimed above, no proof-theoretical apparatus "corresponds" perfectly to natural language reasoning. The discussion about which proof system is more similar to natural language reasoning is too complex to be covered here, so my proposal will take a simple proof method, with fairly intuitive rules. I will use a simple *tableaux* system,[21] although it is important to remark that the schematic definition of Proof-inclusion is suited for other calculi as well, such as *sequent calculus*. A *tableaux* is a tree which represents an argument *ad absurdum*. In the first step we enumerate the premises and deny the conclusion, i.e. we start the tree in the following way:

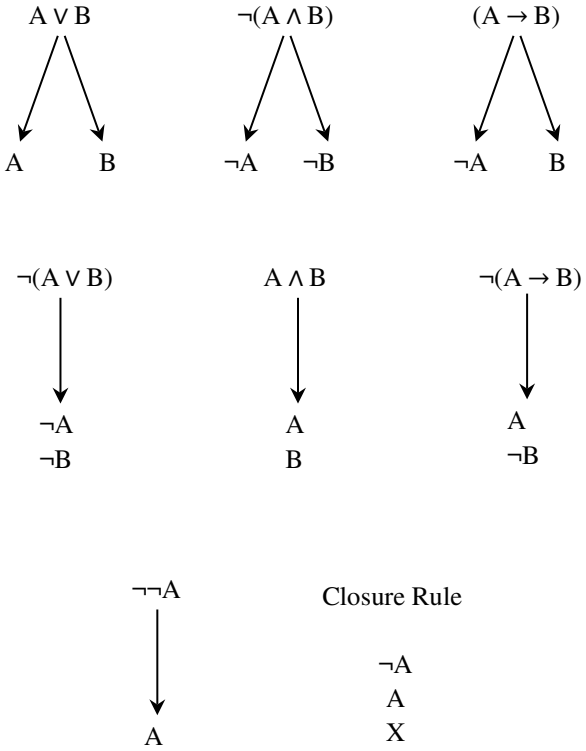
Premise 1
…
Premise *n*
¬Conclusion

Then, by valid transformation rules, we try to reach a contradiction. An application of a rule may extend a branch or open two new branches (for example, if A ∨ B is in the three, there will be a branch with A, and another branch with B). When a branch includes A and ¬A for one formula A, we say that the branch is closed, and we put the symbol "X". The formulae in

[21] See D'Agostino et al. (1998) for a detailed presentation of *tableaux* systems.

each branch are *nodes*. The *length* of a proof is the number of nodes in it, without counting the nodes that express the premises and the initial hypothesis *ad absurdum*.[22]

The tableaux system for classical logic has the following rules. Each rule "decomposes" a formula, and reduces it to formulae with fewer symbols:



In propositional logic, one must decompose each formula which appears in a branch at most once. Put otherwise, once a rule *R* is applied for *A*, you

[22] This measure of proofs is developed *ad hoc* for treating the problem we are discussing.

should not apply it again to *A*. The proof ends when we reach a contradiction in every branch; i.e. when we can close every branch. In this case, the inference is valid. Otherwise, the inference is invalid, and there are open branches in the *tableaux*. This system is complete and correct for classical logic (cf. Priest 2008, 16-17).

Let's see some examples of how this method is used. For example, a proof of $\neg p, p \vee q \vDash q$ can be performed this way:

It is now possible to define more precisely what is for a proof to include another one, in order to determine more clearly how the axiom of Order is applied:

> (Proof inclusion)
> A proof of $\Gamma \vDash A$ *includes* a proof of $\Gamma \vDash B$ iff, by erasing nodes in the proof of $\Gamma \vDash A$, you can obtain a proof of $\Gamma \vDash B$.[23]

For example, we can think of the inference $\neg p, p \vee q \vDash q \vee r$, which might be obtained with the following *tableaux*:

[23] This notion of inclusion could be extended to a variety of proof systems. For example, in natural deduction, one proof includes another one iff one can obtain the latter by erasing steps of the former. However, introducing other notions of inclusion would also involve introducing other proof systems, and this task is far from the scope of this paper.

It is easy to observe that, if the grey nodes are erased, we get a proof of $\neg p$, $p \vee q \vDash q$. This implies that the second proof includes the first one.[24]

Now we can apply the notion of Order to the tableaux system:

(Order-Tableaux)[25]
If every *tableaux* proof of $\Gamma \vDash A$ includes a proof of $\Gamma \vDash B$, then: $(\Gamma \vDash A) \in R_i$ only if $(\Gamma \vDash B) \in R_i$.

The general idea is that, if you need to prove B in order to prove A, then if A is recognizable for an agent, B is recognizable for the agent too. The axioms

[24] As I claimed before, the same considerations could be represented with a sequent calculus. As an example, see the following derivation:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{p \Rightarrow p}{p \Rightarrow p,q} \quad \dfrac{q \Rightarrow q}{q \Rightarrow p,q}}{p \vee q \Rightarrow p,q}}{\neg p, p \vee q \Rightarrow q}}{\neg p, p \vee q \Rightarrow q \vee r}$$

[25] Notice the difference between Proof Inclusion and Order. Proof inclusion establishes a binary relation on proofs. Instead, Order asks for *every* proof of a certain valid argument to include a proof of another valid argument. In this sense, Order has a higher level of generality.

of Order and Reflexivity can be proved to be consistent. For example, the inferential criterion of the previous section, where R_i is the set of inference that are feasible in k steps,[26] satisfies both axioms:

Theorem
An inferential measure of an agent's recognizability (i.e. where R_i includes the valid arguments that can be proven in k or less steps) satisfies Reflexivity and Order.

Proof
Let us establish that $f(\Gamma \vDash \mathrm{A})$, a measure of the length of inference, is equal to the number of nodes in the shortest proof of $\Gamma \vDash A$, without counting the premise nodes and the initial hypothesis *ad absurdum*. The set R_i of recognizable inferences for an agent i is the set of valid arguments that can be proved in k or fewer steps.
Reflexivity holds because $f(\Gamma \vDash \mathrm{A}) = 0$ when $A \in \Gamma$, given that the *tableaux* is closed just after the enumeration of the premises and the hypothesis *ad absurdum* $\neg$A. Then, given that $k \geq 0$, $(\Gamma \vDash \mathrm{A}) \in R_i$.
Order also holds. If every tableaux proof of $\Gamma \vDash A$ includes a proof of $\Gamma \vDash \mathrm{B}$, then the shortest proofs of $\Gamma \vDash A$ include a proof of $\Gamma \vDash \mathrm{B}$. The length of these proofs of $\Gamma \vDash \mathrm{B}$ will be as much $m \leq n$, where n is the length of the shortest proofs of $\Gamma \vDash A$. So the shortest proofs of $\Gamma \vDash \mathrm{B}$ have length $j \leq m \leq n$. Therefore, necessarily $f(\Gamma \vDash \mathrm{B}) \leq f(\Gamma \vDash \mathrm{A})$, i.e. the measure of difficulty of $\Gamma \vDash \mathrm{B}$ is less or equal than the measure of $\Gamma \vDash \mathrm{A}$. So, for an arbitrary k, if $(\Gamma \vDash \mathrm{A}) \in R_i$, then $(\Gamma \vDash \mathrm{B}) \in R_i$. □

It is worth remarking that, even though the inferential measure satisfies Order and Reflexivity, many other measures may satisfy it (including measures where a short proof is not necessarily easier than a large one).[27]

Following the concept of recognizability, it is possible to develop a consequence operator relativized to the recognizable inferences. In other words, if

---

[26] Above we made an objective interpretation of the inferential criterion, according to which the limit k is identical for every agent. We could also make a subjective interpretation, where the limit k is different for each agent.

[27] The axiom of Order introduces a partial order between inferences that *share premises*. With respect to other valid arguments, there are no restrictions and they could be incomparable.

$\vDash_i$ means "recognizable for agent i": $Cn_i(X) = \{A \mid X \vDash_i A\}$. This consequence operator satisfies some structural properties.

First, Cn_i is reflexive. For an arbitrary i, the axiom of Reflexivity guarantees that $X \subseteq Cn_i(X)$. Now, Cn_i does not necessarily satisfy Monotony (if $X \subseteq Y$, then $Cn_i(X) \subseteq Cn_i(Y)$). In our definition of Order, the inferences have the same set of premises; therefore, inferences with different premises (even though one set of premises includes the other one) could be just incomparable. For the recognizability measure to satisfy Monotony, it is necessary to relax the axiom of Order for inferences with different sets of premises, and to admit *tableaux* where not every premise appears at the beginning. In this way, naturally every proof of $\Gamma \vDash A$ will include a proof of $\Gamma \cup \Delta \vDash A$. Otherwise, one could just add a third axiom to the concept of Recognizability:

(Monotony) If $(\Gamma \vDash A) \in R_i$, then $(\Gamma \cup \Delta \vDash A) \in R_i$

Monotony is also consistent with a purely inferential measure (since the premise nodes do not extend the length of the proof). But Monotony is still a controversial axiom. One might argue that adding premises to an inference makes the inference *less* recognizable, for now it is necessary to find which premises need to be used.

Finally, the operator Cn_i does not satisfy Transitivity (if $A \vDash_i B$ and $B \vDash_i C$, then $A \vDash_i C$). For example, you could recognize that $p \vDash p \vee q$ and $p \vee q \vDash r \rightarrow (p \vee q)$, without recognizing that $p \vDash r \rightarrow (p \vee q)$. The only thing that cannot happen, according to the axiom of Order, is that $p \vDash r \rightarrow (p \vee q)$ is recognizable and $p \vDash p \vee q$ is not, given that every proof of the first argument includes a proof of the second one.[28]

5. Relevance and difficulty

In this section, I will put together the considerations of the two previous sections, i.e. the concepts of relevance and complexity. The combination of WS Validity+Relevance and WS Validity+Recognizable is straightforward:

[28] Indeed, the failure of transitivity makes this notion functional. A transitive notion of recognizability would arguably become trivial.

(Logical rationality)
If $\Gamma \vDash A$, and both Γ and A are relevant; and you could recognize that Γ implies A; then rationality requires you not to believe some sentence of Γ or to believe A.

This requirement asks you to believe the *relevant* consequences of your *relevant* beliefs, whenever you *could recognize* that those beliefs imply those consequences. I can now develop a formal version of this requirement, using the two formal concepts of the previous sections: the set of relevant propositions Δ and the epistemically constrained operator Cn_i. The final requirement is the following:

(Logical Rationality – Formal)
Let i be your inferential capacity, and Δ the set of relevant propositions. Rationality requires your belief set X to be such that $Cn_i(X \cap \Delta) \cap \Delta \subseteq X$.

For example, let us assume that my belief set is $\Gamma = \{PA, \neg T\}$, where PA are the Peano axioms, and T is a formula that can be derived from PA but in a thousand of difficult steps. Suppose that I am seriously discussing T with a colleague; i.e. the set Δ is {PA, T, ¬T}. My belief set is inconsistent, but it is not irrational. Suppose that my inferential capacity is R_i, where $(PA \vDash T) \notin R_i$. In this case, Γ is intuitively a rational belief set, in the applied sense of "rational". Even though Γ is not closed under logical consequence (it is indeed inconsistent), the proof from PA to T is too complex, so that the inference stays out of the set R_i. Therefore, I am not supposed to follow that inference, and my belief set Γ is logically rational.

6. Conclusion

In this paper, I proposed a specific rational requirement for characterizing logical rationality. First, I argued in favor of WS Validity, for it can avoid the Bootstrapping problem. Then I added a restriction to WS Validity in order to solve the problem of irrelevant consequences. The new version of the requirement, WS Validity+Relevance, restricts the requirement to the cases in which premises and conclusion are relevant (i.e. the agent considers or has reasons to consider it). This conceptual notion can be formally represented with the concept of weak relative closure.

After this, I introduced a second modification to address Excessive Demands. According to the new criterion, a valid argument has rational force over an individual whenever the individual *could* recognize its validity. Then WS Validity+Recognizable is obtained. The concept of recognizability can be formally characterized by two axioms. Reflexivity establishes that if $A \in \Gamma$, then $\Gamma \vDash A$ is recognizable; while Order establishes that, if in order to prove $\Gamma \vDash B$ you have to prove $\Gamma \vDash A$, then if you can recognize the first inference you can also recognize the second one.

Finally, I combined these two elements in the final requirement Logical Rationality: if you could recognize that Γ implies A, and both Γ and A are relevant in the context, rationality requires you not to believe some sentences of Γ or to believe A. This is, I think, a complete and formally precise logical rational requirement.

References

BOGHOSSIAN, P. (2003): The Normativity of Content. *Philosophical Issues* 13, No. 1, 31-45.

BROOME, J. (1999): Normative Requirements. *Ratio* 12, No. 4, 398-419.

BROOME, J. (2007): Wide or Narrow Scope? *Mind* 116, No. 462, 359-370.

BROOME, J. (2014): *Rationality Through Reasoning*. New Jersey: Wiley-Blackwell.

CARIANI, F. (2013): Ought and Resolution Semantics. *Noûs* 47, No. 3, 534-558.

D'AGOSTINO, M. & FLORIDI, L. (2009): The Enduring Scandal of Deduction. *Synthese* 167, No. 2, 271-315.

D'AGOSTINO, M., GABBAY, D., HANLE, R. & POSEGA, J. (eds.) (1998): *Handbook of Tableaux Methods*. Dordrecht: Springer.

FIELD, H. (2009): The Normative Role of Logic. *Proceedings of the Aristotelian Society* 82, 251-268.

HANSSON, S. (1999): *A Textbook on Belief Dynamics*. Dordrecht: Kluwer.

HARMAN, G. (1986): *Change in View: Principles of Reasoning*. Cambridge MA: MIT Press.

HINTIKKA, J. (1962): *Knowledge and Belief*. Ithaca: Cornell University Press.

HORTY, J. (2012): *Reasons as Defaults*. Oxford: Oxford University Press.

JAGO, M. (2009): Logical Information and Epistemic Space. *Synthese* 167, No. 2, 327-341.

JOYCE, J. (1998): A Non-Pragmatic Vindication of Probabilism. *Philosophy of Science* 65, No. 4, 575-603.

KOLODNY, N. (2005): Why Be Rational. *Mind* 114, No. 455, 509-562.

KOLODNY, N. (2008): Why Be Disposed to Be Coherent. *Ethics* 118, 437-463.

KORSGAARD, C. (1996): *The Sources of Normativity*. Cambridge: Cambridge University Press.

LIST, C. (2012): Judgment Aggregation: An Introductory Review. *Synthese* 187, No. 1, 179-207.

MACFARLANE, J. (2002): Kant, Frege and the Logic in Logicism. *The Philosophical Review* 111, No. 1, 25-65.

MACFARLANE, J. (2004): In What Sense (if Any) Is Logic Normative for Thought. Talk at APA, April 21, 2004; available at the author's website: http://johnmacfarlane.net/normativity_of_logic.pdf.

MAKINSON, D. (1965): The Paradox of the Preface. *Analysis* 25, No. 6, 205-207.

PRIEST, G. (2008): *An Introduction to Non-Classical Logics*. 2nd ed. Oxford: Oxford University Press.

RAZ, J. (2005): The Myth of Instrumental Rationality. *Journal of Ethics and Social Philosophy* 1, No. 1, 2-28.

REISNER, A. (2011): Is there Reason to Be Theoretically Rational? In: Reisner, A. & Steglich-Petersen, A. (eds.): *Reasons for Belief*. Cambridge: Cambridge University Press, 34-53.

RESTALL, G. & SLANEY, J. (1995): *Realistic Belief Revision*. Technical Report, Automated Reasoning Project, Australian National University.

SCHROEDER, M. (2004): The Scope of Instrumental Reason. *Philosophical Perspectives* 18, 337-364.

SCHROEDER, M. (2009): Means-End Coherence, Stringency and Subjective Reasons. *Philosophical Studies* 143, 223-248.

SHPALL, S. (2013): Wide and Narrow Scope. *Philosophical Studies* 163, No. 3, 717-736.

SMITHIES, D. (2015): Ideal Rationality and Logical Omniscience. *Synthese* 192, No. 9, 2769-2793.

SOUTHWOOD, N. (2008): Vindicating the Normativity of Rationality. *Ethics* 119, 9-30.

STEINBERGER, F. (2015): Three Ways in Which Logic May Be Normative. Manuscript; available at the author's Academia.edu site.

STEINBERGER, F. (2016): Explosion and the Normativity of Logic. *Mind*, forthcoming.

WAY, J. (2010): Defending the Wide-Scope Approach to Instrumental Reason. *Philosophical Studies* 147, No. 2, 213-233.

WÓJCICKI, R. (1988): *Theory of Logical Calculi. Basic Theory of Consequence Operations*. Dordrecht: Reidel.

Organon F 23 (4) 2016: 488-502

The Environment and Its Ontological Status

SALVATORE ITALIA

Department of Education, Psychology and Philosophy. University of Cagliari
Via Is Mirrionis 1. 09123 – Cagliari. Sardinia. Italy
salvatore.italian29k@gmail.com

RECEIVED: 16-12-2015 • ACCEPTED: 13-04-2016

ABSTRACT: Is reality a 'Ready-Made World' or an entity constructed by individuals and social activity? The concept of the environment seems to be the boundary that clearly shows how we can simultaneously adhere to our apparently contradictory intuitions—that is, those about the external and autonomous features of reality independent of human intervention, and those about its undeniably constructed character. The environment, then, seems to be a concept that shows how non-epistemic and epistemic notions of reality (i.e. respectively seeing reality as independent from and dependent on us) can be understood cohesively.

KEYWORDS: Affordances – environment – niche construction – natural selection – realism.

1. Realism, 'dependence', and 'independence'

During our daily activities, when we walk, play football, or talk to a friend about various facts and events—about their *truth*—we refer to a reality that we assume to be independent. The shift from the concept of truth to that of reality occurs naturally because we talk about what is real by discussing what is true, and we trust our truths because they are supposed to refer to autonomous states of affairs. However, our intuitions about the descriptive and 'correspondentist' aspects of truth (that the objectivity of truth is linked to the idea of an autonomous and independent reality yet to be described) must be

balanced with another important and natural intuition: the need to conceive of truth as potentially revisable (the view known as 'fallibilism').[1] As finite beings, we cannot exclude the possibility that an assertion or a belief, even if justified now, could turn out to be false at some point in the future, since we have many examples in the history of science that support the idea that 'truth' can be revised.

These two characteristics of the concept of truth shed light on the features of our intuitive understanding of reality. We conceive of our truths as *constituting* reality (through what is real *for us*) and, at the same time, our truths play valuable roles in our daily lives (in both practical and theoretical activities) given that we also conceive of them as *corresponding to* reality. Thus, truth has an adaptive role (cf. Wuketitis 1999). We consider our truths to be the most reliable instrument for orienting ourselves (practically and theoretically) in a world that we cannot completely keep under control and that sometimes surprises us. We need to protect ourselves from this element of surprise, as predictability is necessary for projecting and creating some order in our lives. Truths are the basis of this process, as they are essentially linked to predictability. If truths are both (C) *constitutive* and (D) *descriptive* of reality[2]—as respectively endorsing an epistemic and a non-epistemic conception of realism—we obtain a concept of reality that is both constructivist and realist: reality is created by our truths and is simultaneously independent of them.[3]

To be clear, realism is the belief in the independent existence of certain entities. Different kinds of realism depend on the entities that are referred to (here I will refer to that particular entity called the 'environment') and on the degree of independence that is ascribed to them. Moreover, independence is a matter of degree, and we can adhere to either an epistemic realism according to which the limits of reality are the limits of our knowledge (current or possible), or a non-epistemic realism according to which the limits of reality are unknowable (currently or absolutely).[4]

---

[1] Fallibilism is in general defined as the view that it is not inevitable that our beliefs be certain; as such it implies that their truth is not unquestionable.

[2] (C): What turns out to be true shapes our idea of what reality consists of; (D): Something is true because it is supposed to correspond to how things really are.

[3] This also entails the problem of relativism: why our truths and not "theirs" are a better description of reality?

[4] To be sure: *non-epistemicity* too, as well as *epistemicity*, is a matter of degree.

Going back to the relationship between truth and reality, it is important to show that by putting too much emphasis on (C), we could arrive at an uneasy problem, as described by Boghossian:

> The world did not begin with us humans; many facts about it obtained before we did. How then could we have constructed them? For example, according to our best theory of the world, there were mountains on earth well before there were humans. How, then, could we be said to have constructed the *fact* that there are mountains on earth? (Boghossian 2006, 26)

This explains that constructivism could and should be balanced with realism: we can construct our concepts, but not the properties they are supposed to refer to; we can construct descriptions but not 'things' or facts. Is that true? There are different degrees of constructivism: according to the most extreme version all kind of entities are formed by individual and social activity; according to the softer type, only a subclass of entities is constructed (like language or social institutions). Referring to environment, I hold the thesis that 'our' environment is constructed, but not entirely, being it also subjected to pressures and constraints coming from other 'external' realities. The field of ontological positions is broad (metaphysical realism, scientific realism, common sense realism, structural realism, internal realism, external realism). Here I will hold a particular version of external realism where "externality" (and thus independence) comes in degrees. We move pragmatically in a world conceived as:

(1) independent from us, a feature that is the basis of our truths' *objectivit*y, which, in turn, is the basis for our trusting these truths; and
(2) dependent on our possible intervention, a feature that is the basis of our desire for further knowledge, to the extent that we want to know how things are in order to modify established 'truths' as required.

Before turning to the environment-notion as an exemplary concept that clearly shows how these two aspects of reality can be held together, there is another point worth mentioning. The environment is an entity that is determined to exist based on a particular science (ecology), so the degree of certainty science can provide us is now of utmost importance. Avoiding both scientism and skepticism—the former of which is a way to trust science as

providing us with absolute certainties,[5] and the latter as endorsing an attitude to never trust science because of its unavoidable contextual character—we can pragmatically say that science, even if dependent on contingent and contextual interests and values (as all humans activities are), can give us objective results (or 'truths') to the extent that such truths are of productive value when directly or indirectly tested in reality.[6]

Scientific results cannot be regarded as *absolutely* true because for both their genesis and their acceptance, they depend on assumptions that are assumed within the 'Lifeworld' where they originate (borrowing this concept from Habermas 1999, 32).[7] Thus, science cannot be an absolute explanation of what is and is not, nor can it explain evolution; instead, it can only be understood as a *contextual* absolute explanation. In other words, it is an explanation that has an absolute value only within a given context. This means that we cannot draw the boundaries between what exists and what does not independently of us in any absolute terms, and therefore we cannot define the various parts of a non-epistemic reality in a way that does not feature the limits of our own particular perspective on the reality that we experience.[8] However, we cannot avoid referring to our truths in absolute terms, at least until there are any compelling reasons to give them up. This is simply a pragmatic reaction to scepticism where we trust science *under condition*—that is, under the awareness of both its contextual dependence and its fallibility.

---

5 This is an extreme version of scientism, being it the general view that the scientific methods, categories and results achieved in natural sciences are the only elements of reliable inquiries (whether in social sciences or humanities or in philosophy).

6 The extreme version of scientism I have been dealing with here can be also defined as an attitude of absolute trust in the power of science and potentially results in having to give up on fallibilism—that is, the awareness that what we hold to be true can turn out to be false someday.

7 In Habermas's philosophy, the notion of the 'Lifeworld' refers to an already-interpreted world in which we move as speakers and actors, made up of a 'background of widespread beliefs' and characterised by 'naive familiarity' and inescapability (see Habermas 1981, II, 199-201).

8 As anticipated above, a non-epistemic conception of reality tends to separate truth from what it is possible to know about it. On the other hand, an epistemic conception of reality sees it as being made up of truths, putting emphasis on the link between truths and our epistemic possibilities (justifications).

After this preamble, aimed at avoiding the accusation of scientific foundationalism, the concept of the environment, interpreted as the limited realm of potential actions for *each* living species can be reintroduced (cf. Sanders 1997, 108). As I will make clear with the idea of 'niche construction', this environment's independence is only *partial* because, on the one hand, it depends on the constraints imposed by other environments (independence) and, on the other hand, it depends on our cultural or ecological intervention (dependence). It is accurate to say that we are animals moving within a reality that we share with other species and whose existence—together with the existence of chemical substances—does not depend on us. Even if *what* these substances are is described by science, the *existence* of these substances does not depend on the existence of science.[9] By considering the various determinants of existence, we arrive at the idea of the 'unity of the universe, to which humans belong as natural creatures' (Habermas 2007, 40-41). Reflection on this fact leads us to think of humans as being part of a reality that was already *out there before* the birth of human species, and that will continue to survive in the case of the disappearance of human beings. So even if we deal with a reality that is somehow made by us (environment) we also *indirectly* deal with other less depending realities.

How do we conceive of this external reality? Does it affect our own life constraints, or must we instead conceive of such constraints as depending strictly upon us? Constructivism emphasises the responsibility of human beings in creating their own reality, forgetting that not all of the constraints we encounter are easily met (see von Glaserfeld 1981). In contrast, metaphysical realism does not take into account the constructive role of human beings in building their own reality. In fact, according to metaphysical realism, there is a 'Ready-Made World', or an absolute reality that our knowledge must correspond to in order to be true; this can only be achieved if we are able to obtain a God's Eye Point of View where it is possible to describe reality in its actual form, thus developing a 'unique true theory' (cf. Putnam 1982).

The problems here are with the idea of a unique (absolutely) true description according to scientism, or the idea of a fixed and eternal 'Ready-Made World' and the presumption of accessing a God's Eye Point of View from a human context. In a nutshell, we could never be sure to have achieved this

---

[9] For instance, one could ask whether or not salt was sodium chloride before the introduction of chemistry (cf. Marconi 2007, 64).

God's Eye Point of View even if we actually had it. It seems as though the answer lies in acknowledging both answers to the question, and that arguing for the presence of an external reality that influences the features of our own 'reality', together with the influence of the individual's own construction, paint the most accurate picture of what reality itself actually consists of. There is of course an interplay between external and internal factors in characterizing what we usually call 'reality'.

In the present essay, I will show how this interplay works in determining the reality of our 'environment'. In fact, it seems to me that evolutionary biology and ecological psychology—relying on the notions of 'niche construction' and 'affordances', respectively—lead us on the right path in the argument for an external reality (not completely constructed) that stripped of all of the baggage usually associated with metaphysical realism.

2. Realism of niche construction and affordances

A preliminary step in the argument requires that a definition of both niche construction and affordances be provided. These two concepts shed light on both the dependent and independent aspects of the environment. External reality can be thought of as the framework for all human epistemic activities, consisting of the indefinite totality of all *ecological relations* between organisms and their environments, meaning that the 'products, resources, and habitats that […] organisms construct […] constitute fundamental components of their world and those of other species' (Day, Laland & Odling-Smee 2003, 87). We can consider these relations as *affordances*, defined as 'relations between abilities of the organisms and features of the environment' (Chemero 2003, 189). This concept was first introduced by J. J. Gibson (cf. Gibson 1979), who defined an affordance in terms of what the environment offers to an animal as a possibility, either good or bad: for example, 'for humans the ground is walk-on-able, chairs are sit-on-able, water affords drinking and so on' (Withagen & Wermeskerken 2010, 490).[10] However, the vagueness implicit in this first definition has made it possible for other accounts of affordances to be proffered,

[10] 'The meaning or value of a thing consists of what it affords' (Gibson 1982, 407). It is possible to talk about 'positive' and 'negative' affordances (see Kono 2009, 358). For other definitions of affordances, see Chemero (2003, 184-190).

among which I find Chemero's to be one of the most appealing. Defining affordances in terms of 'dynamic relationships' has the advantage of avoiding the question of how they could have existed before the animals, because it leads us to consider the environment as not having existed before the animals, but instead as being deeply linked to them. In other words, 'animals and their environments evolve together, and animal's alteration of the environment has a constitutive role in this co-evolution' (Withagen & Wermeskerken 2010, 497).

This alteration of the environment caused by animal activity is the so-called niche construction, the process which includes birds, ants, and spiders constructing nests, and humans (for whom cultural activities play a major role) creating houses (Withagen & Wermeskerken 2010, 499).[11] Laland, Odling-Smee & Feldman (2000) have discussed two ways in which organisms can operate in their environments: 'perturbation' (through which they change components of their environment, such as chemicals, resources, or artefacts) and 'relocation' (when they move within an environment's space, or from one environment to another). As a clear example of the first, forests 'contribute to the hydrological cycle, through the retention and evapotranspiration of water, and by doing so they may affect their own weather' (Day, Laland & Odling-Smee 2003, 85), whereas an example of the second is when animals select habitats for annual or seasonal migrations.[12] If the notion of affordance shows that the environment constrains the possibility of action (which is a fundamental point in support of metaphysical realism), the concept of niche construction can resolve this concern by showing that animals can change their own environments—in other words, they can modify what their environments 'afford' them (thus consistent with the requirements of constructivism).

An example of how niche construction counteracts the environment's natural selection process can be shown through the study of Lill & Marquis (2003), where they describe how *Pseudotelphusa* caterpillars build leaf shelters which in effect constitute environments for *other* insects to colonise. However, the classic (and maybe clearest) example of how environment is strictly dependent upon an organism's life is Charles Darwin's study of the earthworm's impact on the landscape (cf. Darwin 1881; Ghilarov 1983). Darwin

---

[11] Niche construction theory is usually associated with the name of Richard Lewontin, who presented its first formulation in 1983.

[12] Perturbatory niche construction, too, is more visible in animals (cf. Day, Laland & Odling-Smee 2003, 85).

aimed to show how 'earthworms and the vegetable mould surrounding them have co-evolved. Vegetable mould did not exist before earthworms. Their relation is *mutual*' (Costall 2001, 478).[13] In particular:

> [W]orms, through the creation of vegetable mould, are responsible for changing the structure and chemistry of the topsoil. And this has profound effects on the ecosystem. As every gardener knows, vegetable mould provides a rich structural basis from which to grow plants. However, the change in the topsoil has implications also for the worms themselves. Initially, worms were structurally not very well adapted to the topsoil. Indeed, their epidermis is very sensitive and needs to be kept warm and wet. They are better suited to live in water than in the soil. However, by changing the chemical composition of the soil, digging burrows and dragging leaves in, earthworms created an environment that better suits their physiological make-up. (Withagen & Wermeskerken 2010, 500)

We now seem to be back at the starting point: does a truly independent reality exist or not? I want to stress that affordances and niche constructions are two concepts that make it possible for us to see more deeply within the structure of the particular reality called the environment. Even if an environment does not exist before the animals living within it come into being, it is only *partially* unavailable to them in a sense that I am now going to explain. Through this kind of partial unavailability, the environment puts us on the right path to find a *minimally* independent reality (still unable to meet the requirements of the strongest realists) because even if, on the one hand, reality undergoes all of the changes that we, as living species, bring to our environment (niche constructions), on the other hand, it imposes constraints upon us that do not depend on our natural selection, instead depending on:

(a) what we have brought to life (and what we face now in life as an autonomous reality);

---

[13] Because most of the effects of niche construction are long-term ones (on scales ranging from the extremely local to the global), it is possible to speak of an 'ecological inheritance' (Laland & O'Brien 2012, 192). For example, the worms' offspring will inherit 'not only genes […] but also an environment that better suits their epidermis' (Withagen & Wermeskerken 2010, 501). Withagen & Wermeskerken (2010, 505) see ecological inheritance as an inheritance of affordances.

(b) what we have not changed (because we are still unable to change it or because it was not a priority for us to change);
(c) what other species do in their own environments.

It is thus clear that through niche construction, living organisms 're-script the pattern of natural selection' so that 'adaptation (adaptative complementarity) results from two processes (selection and construction), not one' (Laland & O'Brien 2012, 195). The environment of a particular species is made up of the sum of all of *its* affordances, which together determine the realm of all intentions that can be satisfied at a given place and time, and since affordances (and thus environments) do not exist before their species, it follows that if a species disappears, the affordances linked to it vanish as well.[14] This result leads us to consider the difference between the environment and the external reality, the first being is a product of particular affordances: the environment is made up of all of the affordances that are related to it, whereas external reality is made up of all of the existing affordances related to different environments. In other words, considering that the latter is the sum of all affordances (also made of affordances that do not depend on us but on other organisms whose existence is independent from us), it can also be conceived of as the sum of all the environments.

Affordances owe their own existence to the presence of a species able to perceive and exploit them, while their exploitability also depends on the presence of an intention (see Stoffregen 2003, 125-126).[15] However, animals can alter the environment 'in order to change what it affords them' (Kono 2009, 366), i.e. to adapt it to their intentions. In the case of humans, this niche construction occurs mostly through the use of language as a powerful instrument of change. In fact, it is language (and, more generally, communication) that

[14] Conceiving environment in terms of affordances makes it possible to explain not only what is relevant to a species but also why it is relevant (cf. Withagen & Wermeskerken 2010, 505).

[15] Chemero conceives of affordances as a *primitive* way of perceiving the environment, as a 'feature placing' mechanism: the difference between this kind of perception and the perception of an object is explained as being the difference between 'realizing that your car is dented' and 'realizing that it is raining' (Chemero 2003, 185). We can consider an affordance as an *emergent* property because since 'it is a conjunction of a property of the environment and a property of the animal, the actualization of the affordance is a property of the animal-environment system' (Stoffregen 2003, 122).

allows us to develop such a complex culture—one that is able to feed back to nature with the ultimate aim to change it according to our needs and desires, such that it is communicatively shaped and stabilised. When affordances do not allow us to realise an intention, we begin to think about different solutions to realise our goals. We may want to find even more effective ways to realise them. Through evolution, we have developed cooperative attitudes, finding that sociality makes things easier (based on our needs of survival). An example here is the division of labour that allows us to increase production in order to have more free time. A necessary requirement for this complex development is communication. Here, subjects exploit their society, and at the same time are oppressed by it, but language is the medium through which they can both influence and be influenced via the set of affordances that their particular society allows them. In other words, we can say that human niche construction operates mostly through culture—that is, through sociality and communication. In light of this, affordances are the means by which a connection is established between cultural and linguistic human activity and external reality.

This last dimension reveals the external reality's *unavailability* when it does not allow some change to happen (at least at the present time). This shows resistance from a dimension that does not depend on us: that is, what exists prior to a constructed environment, or the indefinite set of *other* animal-environment relationships that together constitute the reality in its not-constructed level (as a sum of other environmental relationships). Such a context is both *absolute* (from the point of view of the existence of a species) and *relative* (because externality is always perspectival with respect to the species' points of view within the considered time frame). The environment has a connection to an external reality, characterised by many environments that undergo changes due to both external and internal factors, such as the presence of other environments and the niche construction processes. These environments can possibly modify selection pressures (at least certain kinds).

This ecological approach to realism that I have shown here seems to be the way to avoid the risks of: (1) metaphysical realism; and (2) constructivism. In fact, regarding (1), my idea is consistent with the view that the environment is 'evolutionist without being selectionist' (cf. Chemero 2003, 190), as it is inconsistent with Reed's 'environmental determinism' (see Costall 2001, 478) where the environment, as an independent variable (or a Ready-Made World),

gives us a set of possibilities (or an ecological niche) to which we must conform in an adaptative way.[16] This traditional selectionist view, best captured in the metaphor of adaptation, forgets that animals' 'utilization, destruction and creation of affordances are central elements in evolutionary dynamics' so that 'animals do not evolve so as to fit in a pre-existing environment' (Withagen & Wemeskerken 2010, 489-490). As a result of these limitations, I have provided here a different and non-selectionist way to conceive of the environment.

Finally, regarding (2), the reason why this kind of constructive process cannot be conceived of in a strong constructivist way (anti-realism) is that 'niche construction *can* alter the evolutionary process', but it has to confront itself with constraints that can (or cannot) allow for such changes: 'Changes in the affordance layout are not exclusively the result of animal activity. Indeed […] geological and hydrological processes can also alter the affordances in an animal habitat' (Withagen & Wemeskerken 2010, 503). For example, 'A volcanic eruption can change the context of selection of many species in a certain habitat' (Withagen & Wemeskerken 2010, 502). In addition, 'niche-constructing organisms may also influence the evolution of other populations' (Day, Laland & Odling-Smee 2003, 90). For instance, forests can alter the weather, weather in general can influence the development of other plants, and 'organisms can pump abiota into physical states that the abiota could never reach on a dead planet, and these modified abiotic components of ecosystems may later become the source of modified natural selection pressures acting on other species' (Day, Laland & Odling-Smee 2003, 90).[17] There is more: the environmental change that results from niche construction feeds back again into the

[16] Reed (1996) presents a selectionist view of affordances, according to which they, as resources, exist prior to the animals, while Withagen & Wemeskerken opt rather for a conception which is more committed to niche construction theory. The fact that affordances extert some selection pressure does not mean that we cannot operate on them.

[17] This phenomenon is known as 'ecosystem engineering'—that is, organisms' creation, destruction, or modification of habitats and/or modulation of the availability of resources to other species (see Jones, Lawton & Shachak 1994). A necessary requirement for making this possible is the presence of 'engineering webs'—connective webs in ecosystems—which are caused by species' influencing energy and mass flows and which create habitats and other resources for other species (cf. Jones, Lawton & Shachak 1994). For such definitions see Odling-Smee, Erwin, Palkovacs, Feldman & Laland (2013). Abiota are non-living factors influencing the characteristics of an ecosystem.

constructive population as a modified source of natural selection that is therefore *independent*. For example:

> [T]he construction of villages, towns and cities creates new health hazards associated with large-scale human aggregation, such as the spread of epidemics. Humans may respond to this novel selection pressure either through cultural evolution […] constructing hospitals and developing medicines and vaccines or, at the ontogenetic level, developing antibodies that confer some immunity or through biological evolution, with the selection of resistant genotypes. (Laland & O'Brien 2012, 198)

Using the vocabulary of affordances, it is possible to say that 'affordances not only form the context of selection that animals *encounter*; by creating and destroying affordances, animals also *construct* this context and thereby affect the evolutionary dynamics' (Withagen & Wemeskerken 2010, 504).

In considering the idea that some external constraints can influence the niche construction process, the concept of the environment that I have provided displays a kind of realism (about an entity called the 'environment') that is both epistemic and non-epistemic because it lead us to think about different realities, some of which are independent both in their existence and in the constraints they put on us (*gradual* realism).[18] This provides us with an idea of how these two different ideas can coexist, provided that we conceive of 'non-epistemic' in the minimal and external sense of a limit that can (or cannot) influence our epistemic processes, but whose reality still exists outside of them (even if it is not completely independent).

It is the very presence a non-epistemic dimension that makes my model different from Wuketitis'. He correctly acknowledges that adaptationist approaches in evolutionary epistemology (see Vollmer 1975; 1984) do not consider the fact that what is out there exerting pressure for natural selection is also a product. In other words, organisms are not a mere product of their surroundings, passively shaped by natural selection pressures. They also positively respond to these pressures, by changing them and creating new and different pressures. This is what a non-adaptionist approach affirms (cf. Wuketitis

[18] I speak of *gradual* realism to highlight that not all realities depend on us to the same extent. In fact, environment is more modifiable than external reality: both are independent, even if in different degree.

1989).[19] Along these lines, Wuketitis develops a *functional* realism that, in my view, ignores the role of the non-epistemic dimension. He stresses the species' own world as being the product of a dynamic and interactive process, but this same world also undergoes changes and constraints from within the external dimension: the external reality (as I define it). This notion is not without use (as is claimed in Wuketitis 1999, 30), to the extent that it exerts a very specific role (a negative one) on our own environment.

3. Conclusion

Epistemic processes (the example here is the niche construction process) are both free and constrained (natural selection). We move within a framework of different degrees of reality, some more dependent upon us than others. The case of the environment clearly illustrates the presence of different realities, and in particular, the nature-universe (or external reality), which is more independent (even if it undergoes indirect revisions from all the environments that, together, constitute it), and the environment itself, which is more dependent on our niche construction processes. So, affordances still undergo the limits that our environment inherits from an external reality; this explains why not all intentions can be satisfied. However, their unavailability is not absolute; affordances can change for both internal and external reasons, i.e., because of our niche construction processes and for external non-epistemic reasons (e.g., other species' processes).

To conclude, what is realist, in the concept of environment, is its working as a reality that allows or not allows our intentions to be satisfied; what is anti-realist, is that through the niche construction processes we can modify our environment.

References

BOGHOSSIAN, P. (2006): *Fear of Knowledge. Against Relativism and Constructivism*. Oxford: Oxford University Press.

[19] He does not cut off the presence of an external reality to which organisms adapt; he simply tries to balance its weight, with respect to the constructive role of organisms. A nonadaptionist approach is not *anti*adaptionist (see Wuketitis 2005).

CHEMERO, A. (2003): An Outline of a Theory of Affordances. *Ecological Psychology* 15, No. 2, 181-195.

COSTALL, A. (2001): Darwin, Ecological Psychology, and the Principle of Animal Environment Mutuality. *Psyke & Logos* 22, No. 2, 473-484.

DARWIN, CH. (1881): *The Formation of Vegetable Mould, through the Action of Worms with Observations on their Habits*. London: John Murray.

DAY, R. L., LALAND, K. N. & ODLING-SMEE, J. (2003): Rethinking Adaptation: The Niche-construction Perspective. *Perspectives in Biology and Medicine* 46, No. 1, 80-95.

GHILAROV, M. S. (1983): Darwin's Formation of Vegetable Mould – its Philosophical Basis. In: Satchell, J. E. (ed.): *Earthworms Ecology: From Darwin to Vermiculture*. London: Chapman and Hall, 1-4.

GIBSON, J. J. (1979): *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

GIBSON, J. J. (1982): Notes on Affordances. In: Reed, E. & Jones, R. (eds): *Reasons for Realism: The Selected Essays of James J. Gibson*. Hillsdale: Erlbaum, 401-418.

JONES, C. G., LAWTON, J. H. & SHACHAK, M. (1994): Organisms as Ecosystem Engineers. *Oikos* 69, No. 3, 373-386.

KONO, T. (2009): Social Affordances and the Possibility of Ecological Linguistics. *Integrative Psychological and Behavioral Science* 43, No. 4, 356-373.

LALAND, K. N. & O'BRIEN M. J. (2012): Cultural Niche Construction: An Introduction. *Biological Theory* 6, No. 3, 191-202.

LEWONTIN, R. (1983): *Gene, Organism and Environment*. In: Bendall, D. (ed.): *Evolution from Molecules to Men*, Cambridge: Cambridge University Press, 273-285.

LILL, J. T. & MARQUIS, R. (2003): Ecosystem Engineering by Caterpillars Increases Insect Herbivore Diversity on White Oak. *Ecology* 84, No. 3, 682-690.

HABERMAS, J. (1981): *Theorie des kommunikativen Handelns*, I-II. Frankfurt am Main: Suhrkamp.

HABERMAS, J. (1999): *Wahrheit und Rechtfertigung. Philosophische Aufsätze*. Frankfurt am Main: Suhrkamp.

HABERMAS, J. (2007): The Language Game of Responsible Agency and the Problem of Free Will. How Can Epistemic Dualism Be Reconciled with Ontological Monism? *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* 10, No. 1, 13-50.

MARCONI, D. (2007): *Per la verità. Relativismo e filosofia.* Torino: Einaudi.

ODLING-SMEE, J., ERWIN, D. H., PALKOVACS, E. P., FELDMAN, M. W. & LALAND, K. N. (2013): Niche Construction Theory: A Practical Guide for Ecologists. *The Quarterly Review of Biology* 88, No. 1, 3-28.

PUTNAM, H. (1982): Why There Isn't a Ready Made World. *Synthese* 51, No. 2, 141-167.

REED, E. S. (1996): *Encountering the World*. New York: Oxford University Press.

SANDERS, J. T. (1997): An Ontology of Affordances. *Ecological Psychology* 9, No. 1, 97-112.

STOFFREGEN, T. A. (2003): Affordances as Properties of the Animal-Environment System. *Ecological Psychology* 15, No. 2, 115-134.

VOLLMER, G. (1975): *Evolutionäre Erkenntnistheorie*. Stuttgart: Hirzel.

VOLLMER, G. (1984): Mesocosm and Objective Knowledge – on Problems Solved by Evolutionary Epistemology. In: Wuketits, F. M. (ed.): *Concepts and Approaches in Evolutionary Epistemology: Towards an Evolutionary Theory of Knowledge*. Dordrecht: Reidel, 69-121.

VON GLASERFLED, E. (1981): Einführung in der Radikaler Konstruktivismus. In: Watzlawick, P. (ed.): *Die erfundene Wirklichkeit*. München: Piper & Co., 16-38.

WITHAGEN, R. & WERMESKERKEN M. V. (2010): The Role of Affordances in the Evolutionary Process Reconsidered: A Niche Construction Perspective. *Theory Psychology* 20, No. 4, 489-510.

WUKETITIS, F. M. (1989): Cognition: A Non-adaptationist View. *La Nuova Critica* 9-10, 5-15.

WUKETITIS, F. M. (1999): Functional Realism. In: Carsetti, E. (ed.): *Functional Models of Cognition*, Dordrecht: Kluwer Academic Publisher, 27-38.

WUKETITIS, F. M. (2005): Evolutionary Epistemology: The Nonadaptationist Approach. In: Gontier, N., van Bendegem, J. P. & Aerts, D. (eds.): *Evolutionary Epistemology, Language, and Culture*. Dordrecht: Springer, 33-46.

Organon F 23 (4) 2016: 503-524

A Minimalist Framework for Thought Experiment Analysis

MAREK PICHA

Department of Philosophy. Masaryk University in Brno
Arne Nováka 1. 602 00 Brno. Czech Republic
mpicha@gmail.com

RECEIVED: 12-08-2016 • ACCEPTED: 29-09-2016

ABSTRACT: Thought experiments are frequently vague and obscure hypothetical scenarios that are difficult to assess. The paper proposes a simple model of thought experiments. In the first part, I introduce two contemporary frameworks for thought experiment analysis: an experimentalist approach that relies on similarities between real and thought experiment, and a reasonist approach focusing on the answers provided by thought experimenting. Further, I articulate a minimalist approach in which thought experiment is considered strictly as doxastic mechanism based on imagination. I introduce the basic analytical tool that allows us to differentiate an experimental core from an attached argumentation. The last section is reserved for discussion. I address several possible questions concerning adequacy of minimalistic definition and analysis.

KEYWORDS: Experimentalism – Galileo – minimalism – reconstruction – Rescher – thought experiments.

Despite extensive discussions on the nature of thought experiments, there have been only a few attempts to explicitly grasp the way thought experiments are constructed. In the first part of my text, I present two of those attempts: Ray Sorensen's structure of *refuter* and Nicholas Rescher's quintet of *complex hypothetical reasoning*. While Sorensen emphasizes the similarities between real and thought experiments, Rescher underlines the relevant lesson we draw from a supposition. In the second part, I argue that both approaches rely on

dubious definitions of thought experiment. To avoid their unintuitive consequences, I form a *minimalist* view where thought experiments are identified strictly with contemplation of a hypothetical scenario.

There are three questions I ask in each of the mentioned frameworks:

(i) What is a thought experiment?
(ii) What is the structure of a thought experiment?
(iii) Which discursive role does a thought experiment play?

The first question is conceptual and deals with a definition of thought experiment. The second question is logical and concerns the relation of basic elements within a thought experiment. The third question is dialectical and focuses on a connection of thought experiments and a standpoint in a critical dialogue.

The paper does not address epistemological issues of thought experimentation, at least not directly. Both Sorensen and Rescher offer clues for thought experiments' reconstruction, both elaborate their definitions into robust analytical frameworks—and there are no other major contributions to the explicit and universal *structural* analysis of though experiments. That is the sole reason why the paper focuses on their accounts; I will however discuss the relation between minimalism and other prominent approaches to thought experiments in the final section.

1. Experimentalist approach

(i) Ray Sorensen defines thought experiment as "an experiment that purports to answer or raise question rationally without the benefit of execution" (Sorensen 1992, 205). In his view, there is a broad category of general experiment. This category covers experiments that are performed by causal manipulation in the environment, and also experiments that are just contemplated. What is a general experiment? Sorensen accepts the classical explication by which an experiment is a process of the variation of the independent variable and the measurement of the dependent variable. A thought experiment is the instance of a general experiment that is conducted by speculation, the change and measurement of variables is *as–if* executed.

(ii) Further, Sorensen explicitly mentions the logical structure of thought experiments (1992, 132). He puts forward two reasoning schemes, necessity

refuters and possibility refuters, which I present here in a simplified and unified version:[1]

1. source [S]
2. extraction [if S, then I]
3. content [C]
4. conditional [if I and C, then W]
5. dismissal [non-W]

According to Sorensen, all thought experiments have a structure of paradox. One cannot hold as true all the propositions in the quintet, yet all of them seem plausible. From the argumentative point of view, they are instances of *reductio*: reasoning about an imaginary case *C* leads to an absurd result *W* that makes the provisional hypothesis *S* untenable.

Let me present the experimentalist analysis by using two examples. For the sake of clarity, the first one considers the trivial case of flying pigs, the second one is more subtle and deals with the famous Gettier example.

Case I: Flying Pigs

Suppose that pigs can fly. Wouldn't we know that since we have radars?

1. source: Radars can register flying objects.
2. extraction: Radars would register flying objects of pig size.
3. content: Pigs can fly.
4. conditional: Radars would spot a flying pig.
5. dismissal: No radar has spotted a flying pig yet.

The source statement about radars leads to a consequence that we would detect flying pigs. This consequence is dismissed as absurd—not in terms of logical or nomological impossibility, but in terms of sheer improbability. The following analysis is a very slight modification of Sorensen's own regimentation of Gettier's thought experiment (1992, 137).

[1] Sorensen proposes two variants of the same general scheme based on modalized modus tollens. They differ in few details which I find too minor to reproduce here. For the sake of simplicity I have thus omitted modal operators and distinction between "content possibility" and "content copossibility" in his reconstructions; for the whole picture see Sorensen (1992, 136 and 153).

Case II: Gettier example

Suppose that Mr. Smith has justified belief that is true by some unforeseen coincidence. Does Mr. Smith have knowledge?

1. source: The definition of knowledge is justified true belief.
2. extraction: If knowledge is justified true belief, then if any person has justified true belief that p, then the person knows that p.
3. content: It is possible for Smith to be justifiably right for the wrong reason.
4. conditional: If all justified true believers that p have knowledge that p and Smith is justifiably right but for the wrong reason, then Smith knows that p because of luck.
5. dismissal: It is impossible for anyone's knowledge to be due to luck.

(iii) What is the dialectical role of thought experiments? Sorensen presents a thought experiment as a tool for attacking a standpoint. Imaginary cases are put forward as counterexamples, they focus on the refutation of a modal statement held or implied by an opponent. In scientific discourse, a thought experiment may serve as a test of hypothesis. The experimenter attempts to falsify the hypothesis by appeal to its counterintuitive consequences.

Karl Popper famously yet vaguely mentions three possible types of thought experiments: critical, heuristic and apologetic (cf. Popper 1992, 464). Thought experiments of the first type are meant to refute a theory, heuristic experiments illustrate and apologetic experiments support a theory. As we see, Sorensen's approach does not count with an apologetic or heuristic use of a thought experiment. In an experimentalist framework, all experiments are aimed at refuting a source statement and therefore are straightforwardly critical.

2. Reasonist approach

(i) According to Nicholas Rescher, thought experiments are complex courses of hypothetical reasoning. A thought experiment is a "lesson learned by deriving a conclusion from supposition and background knowledge" (Rescher 2005, 8). Rescher describes thought experimentation as a train of thought during which we imagine a scenario, consider the circumstances, infer provisional outcomes and pick the most plausible one. However, that is not enough; there has to be a serious question we are trying to answer. Rescher

explicitly mentions answering a *larger* question and drawing a *larger* lesson. Not every conclusion reaching process based on hypothetical reasoning is thus a thought experiment—only some of them qualify as such and the rest is a *mere speculation* (cf. Rescher 2005, 6). Rescher mentions the following question as an example of mere speculation: What if one could converse with flowers? Such a question is not able to constitute a thought experiment; there is no larger problem to solve. I will discuss the distinction in more detail later.

(ii) Thought experiment is a subtype of hypothetical reasoning. By Rescher's definition, there must be five elements present in genuine thought experimentation:

1. supposition
2. context-specification
3. conclusion-deriving
4. lesson drawing
5. synoptic reasoning

The first three steps are quite straightforward: we introduce a supposition into a context to get a conclusion. According to Rescher, a supposition is a provisionary accepted proposition, i.e. the proposition we are in fact agnostic about or even the proposition we disbelieve in. Context-specification provides additional information about the scenario. Clues about who, what, where and when may be crucial for reaching the conclusions of the thought experiments. In the third step, we infer the conclusion from the given supposition and background beliefs fixed by context. However, Rescher points out that the result of a thought experiment is not the conclusion itself, but the lesson we learn from the conclusion. To illustrate his view about lesson drawing, he uses a slightly obscure example:

Case III: Yet Again Flying Pigs

Suppose that pigs can fly. Wouldn't it be an interesting thought experiment?

1. supposition: Pigs can fly.
2. context: Animals exercise their abilities on suitable circumstances.
3. conclusion: Pigs will sometimes fly.
4. lesson: Not every thought experiment is all that interesting.
5. reasoning: Yet Again Flying Pigs is an example of uninteresting thought experiment.

There is a strange self-reference in Rescher's reasoning when the first three steps are mentioned in step four. Fortunately, Rescher later offers a more straightforward example when he considers the thought experiment about the relation between morality and our ability to predict (see Rescher 2005, 13). I propose the following regimentation:

Case IV: Unanticipated Results

Suppose people have no capacity for foresight. Would they be responsible for their actions?

1. supposition: People have no capacity for foresight.
2. context: Capacity for foresight is necessary for predicting consequences of actions.
3. conclusion: People cannot predict consequences of their actions.
4. lesson: Predictive foresight is a crucial precondition of morality.
5. reasoning: Moral responsibility is crucially dependent upon foresight.

I understand the quintet as a simple structure where a conclusion derived from a supposition and context serves as an input for reasoning leading to a lesson. A context and reasoning are warrants; they capture the way a supposition is connected to a conclusion, and a conclusion to a lesson, respectively.

(iii) Considering the relation of thought experiments to standpoints, Rescher's approach is in accordance with Popper's. Thought experiments can be used as a constructive support for a standpoint; they can be used destructively as well. However, Rescher and Popper differ in their evaluation of the constructive use. Popper is highly suspicious about apologetic/constructive thought experiments and does not admit them as a legitimate part of scientific method (see Pooper 1992, 466 and 473). On the other hand, Rescher is more benevolent towards thought experiments; both destructive and constructive ones are valuable tools of inquiry. Though, he points out that constructive thought experiment does not support a standpoint as a proof but only as a plausible reasoning (see Rescher 2005, 34).[2]

[2] Rescher also mentions the explanatory use of thought experiments, but unfortunately not with much detail. As it seems, formulating an explanation of phenomenon B in the form 'if A happens, then B happens' is for Rescher a type of thought experimentation. See Rescher (2005, 16 and 51).

3. Minimalist approach

Both Sorensen and Rescher consider thought experimentation as a process in which we yield a result from the initial supposition. For Sorensen, the result is the *refutation* of the source statement; for Rescher, the result of a thought experiment is the *final lesson* we draw. The minimalist position is different: both of the mentioned results do not belong to the thought experiment itself. The refutation of the source statement and the lesson learned is extra.

(i) Thought experimentation is a particular way we accept a belief. Minimalism sees thought experimentation as a doxastic mechanism built upon imagination. Thought experiment is thus defined as "a set of instructions which specify what to imagine provided a particular cognitive goal is pursued" (Picha 2011b, 22).

There are two necessary conditions for a thought experiment. First, one is asked *to imagine a situation*, a scene, a state of a possible world. Thought experimentation is a method; it is a way of how to try solving a given task, and it requires a specific mental capacity. Let us call the person who brings up a thought experiment a *promoter* and the person who should execute the experiment an *experimenter*. The promoter instructs the experimenter to find the answer by envisaging the scenario followed by the educated guess of the result. It would not be a thought experiment if one finds the answer either by calculating the result, by the real world execution or by a public survey.

Secondly, imagination is involved in order to *find a solution* to the given task. An experimenter is imagining the scenario to guess what would probably happen under the described circumstances. Imagination driven by other motives does not count as thought experimenting—for instance, when someone fantasies about resting in a silent comfortable place to calm down tense emotions or when a child amuses herself by picturing the world where pigs can fly.

(ii) According to minimalism, thought experiments are neither paradoxes nor lessons; thought experiments are just segments of those structures. Thought experiments correspond solely to *a content possibility* in Sorensen's structure or to the first three steps of Rescher's structure where *a supposition* and *a context* lead to *a conclusion*. A thought experiment is thus just the part of complex hypothetical reasoning where one contemplates an imaginary situation and resolves the related task. The minimalist structure of a thought experiment resembles the structure of an experiment in general. There are two pairs of interconnected propositions:

1. input *i*
2. output *o*
3. modification *m*
4. result *r*

The *input* is the starting point of thought experiment. A promoter suggests a proposition as an accepted one. The input is meant to be taken for granted and there should be no discussion involved. It usually describes a familiar situation or puts up an obvious point. As an example, here is the input of Flying Pigs:

1. input: Some fairly big animals can fly.

The *output* is a proposition stating what would happen under the circumstances given as the input. More precisely, an output specifies how input should be treated. Besides natural thought experiments concerning what would happen, there are also conceptual thought experiments where we are interested in what one would say. In other cases, we may be looking for answers to the behavioral question what one would do under specified circumstances, or what one morally or instrumentally should do, and so on. I reconstruct the output of Flying Pigs as follows:

2. output: Radars would detect these big flying animals.

Both the input and output instantiate a *baseline* which is normal, usual, obvious, or at least a provisionally accepted description of a segment of our world. The baseline may express various features, causal relation between input and output, traditional evaluation of situation captured by the input, accepted definition, common policy, working hypothesis and so on. It is impossible to give an exhaustive overview of all possible relations between input and output. Generally speaking, there are two kinds of bridging principles forming two kinds of baselines: factive baseline grasping the way the world or its parts behave, and evaluative baseline describing our attitude to the facts.[3] The baseline of a thought experiment states what the experimenters should take for

[3] Tamar Szabó Gendler proposes a third kind: the conceptual thought experiments. I treat conceptual thought experiments as instances of factual thought experiments. See Gendler (2000, 25) and Picha (2011b, 25).

granted and what they should focus on in forthcoming supposition. The baseline is often implicit; therefore, its reconstructions may vary. However, the proper reconstruction is the one that grasps the bridging principle intended by the promoter of the experiment.

The *modification* of input is the only element that must be explicitly formulated in every thought experiment. A modification is the iconic step in thought experimentation—the *what-if* moment, the supposition. The promoter gives instructions what to imagine; the experimenter at least provisionally accepts the prescribed proposition. The modification of Flying Pigs is obvious:

3. modification: Pigs can fly.

And finally, the experimenter forms a belief about what would happen in such a scenario. In other words, the *result* of a thought experiment is the proposition the experimenter accepts upon consideration of the modification under the aspect fixed in the baseline.

4. result: Radars would detect flying pigs.

As we may see, the minimalist model of thought experimentation is really simple and, well, minimalistic. The baseline tells us what to *focus on* in the following scenario. The modification tells us what scene to *imagine*. The result is obtained by *focusing on the imagined* scene. To get a better picture of the minimalist reconstructions, I propose an analysis of the three thought experiments mentioned above:

Case II: Gettier example

1. *i*: A person has justified true belief that *p*.
2. *o*: That person knows that *p*.
3. *m*: Smith has justified true belief that *p*.
4. *r*: Smith would not know that *p*.

Case III: Yet Again Flying Pigs

1. *i*: Some fairly big animals can fly.
2. *o*: Those fairly big animals would fly.
3. *m*: Pigs can fly.
4. *r*: Pigs would fly.

Case IV: Unanticipated Results

1. *i*: People have capacity for foresight.
2. *o*: People are morally responsible.
3. *m*: People have no capacity for foresight.
4. *r*: People would not be morally responsible.

Minimalism holds that thought experiments themselves bring us nothing more than these almost trivial beliefs. The real struggle begins when the experimental results are implanted into arguments.[4]

(iii) Minimalism is a "narrow" conception: thought experiments are identified strictly with the process of estimating what would happen under the stipulated circumstances. The superstructure built upon the belief accepted in this way is not part of the thought experiment—minimalism thus draws the line between thought experimentation and argumentation. The scenario induces a belief in the experimenter and the belief may then serve as an accepted premise in any type of argumentation.

Thought experiments are usually foundations for paradoxes, lessons and reductio, but are not identical with either of these. It means that thought experiments are not constructive or destructive by their inner structure, i.e. baseline and modification, or by their results. Experimental results are simply used in arguments – and it depends on the relation of the argument to the standpoint whether the experimental result is deployed as a support or as a rebuttal.

Sometimes we are satisfied with the answer to the particular question and a thought experiment does not need to be a part of an argumentation. Nonetheless, more often a thought experiment is promoted as a (hypothetical) case study, i.e. the thought experiment plays a part in a persuasive dialogue in a form of an *argument from thought experiment*. The scheme of this argument is very simple—it is a variant of the basic argument from an example, the only special ingredient is the way the particular example is obtained:

1. The thought experiment induces belief that *b*.
2. *b* is an example of the principle *P*.
3. Therefore principle *P* is (sometimes/typically/mostly/always) true.

---

[4] For more minimalistic analyses of famous philosophical thought experiments see Picha & Pichová (2013).

While the first premise just sums up the result of a thought experiment and the second premise presents the result as an exemplary case, the conclusion deserves a little more attention. The principle *P* in the conclusion can have one of the four following qualifications: some x is y (exemplification); x are typically y (plausible generalization); most x are y (probabilistic generalization); all x are y (universal generalization). The qualification of the principle *P* is an important factor when evaluating the argument: since all three mentioned types of generalization can be subjected to the critique of hasty generalization, the argument from thought experiment based on exemplification is significantly less fragile. The following reconstruction captures the deployment of the Gettier example:

1. Smith would not know that *p*.
2. Smith's belief is the example of justified true belief that is not knowledge.
3. Therefore some justified true beliefs are not knowledge.

The argumentation scheme enables the thought experiment to play the role of counterexample. In this way, the argumentation scheme based on exemplification probably captures Popper's idea of the critical use of thought experiment.

To sum up, a thought experiment plays exactly the same discursive role as an observation or a real experiment. They all elicit a belief that may further serve as a basis for refuting or supporting a standpoint.

4. Discussion

I would like to face some possible questions about the mentioned analytical frameworks, especially minimalism. I will point out some important differences in the frameworks and show their not so intuitive consequences.

4.1. Is not the minimalist definition of thought experiment too broad?

According to minimalism, whenever we solve a given task by imagining the described scenario and estimating the outcome, we are conducting a thought experiment. A minimalist is thus forced to categorize even the most common imagination-based problem solving as a thought experiment. Let me

give an example: You are facing the question whether your luggage will fit into the trunk of a car. You can find the answer by measuring and comparing the volume of your luggage and the trunk; you can solve the task by real experimentation and physically manipulate the luggage inside the car trunk; or you can rotate a mental representation of the luggage to fit into the mental representation of the car trunk. Minimalism holds that the last procedure counts as a thought experimentation. Minimalist definition with its emphasis on imagination is by no means novel. Similar approach was held by Ernst Mach who supposedly defined thought experiments "as the capacity to 'imagine mentally the variation of facts'" (Engel 2011, 146); and Tamar Szabó Gendler who treats thought experiments as "reasoning about an imaginary scenario with the aim of confirming or disconfirming some hypothesis or theory" (Gendler 1998, 398). What is new, however, is the fact that minimalism takes this definition strategy seriously and makes a clear cut between thought experiments and superimposed arguments.

Of course, the minimalist definition is liberal. However, I will argue that the mentioned alternatives are either too restrictive, or they are liberal to the same extent. The experimentalist approach seems promising at first, but turns out to be very limiting when it comes to the structure and the usage of such "successful unexecuted experiments". Experimentalism treats every thought experiment as a disguised paradox with a hypothetical premise. Surely, every thought experiment can be with some extra work built into a reductio. However, the goal of many hypothetical examples is just to support a proximate claim, not to refute a distant source statement. For instance, Yet Again Flying Pigs are very unambitious in this way.

The reasonist approach is restrictive by choice. Rescher is explicit about the distinction between thought experiments and mere speculations—a speculation must be a part of some larger problem to be a thought experiment. Alas, Rescher is not clear about the "larger problem" and I see at least three possible interpretations. First, a problem is larger when it relates to something *worthy* and sublime. Thought experiments would be therefore thematically restricted. For example, speculations about conversation with flowers are not thought experiments, but speculations about conversation with newborns might be. Since there are many prototypical instances of thought experiments about unassuming problems, this interpretation seems obviously wrong.

According to the second interpretation, a problem is larger when it relates to something *general*. A speculation becomes a thought experiment when it

supports or disproves a general claim. I find this interpretation of the "larger problem" too restrictive, because the experimenter might be sometimes interested solely in a particular outcome of a hypothetical scenario; for instance, when asked how many road turns would it take to get to the nearest hospital.

The third interpretation holds that a problem is larger when there is *a question* involved in the speculation. The thought experiment would be a hypothetical scenario accompanied by a relevant question. For instance, "What if one could converse with flowers?" is not a thought experiment (cf. Rescher 2005, 6), but "What if one could converse with flowers? Would flowers have some concepts totally incomprehensible for humans?" is a thought experiment. This third interpretation of the "larger problem" is very liberal and in fact perfectly matches a minimalist approach: modification ("mere speculation" in Rescher's terms) is accompanied by a baseline which determines what in modification should be of the experimenter's interest (the "larger problematic issue"). Therefore, if the "lesson learned" in a reasonist definition truly means the "question answered", then the reasonist and the minimalist definitions are equally liberal.[5]

Finally, let me point out the most important distinction between minimalist and reasonist definitions. As was mentioned before, the reasonist considers thought experiments as lessons learned by the combination of supposition and contextual information. Rescher holds that at least some *mathematical problems* are thought experiments. He gives two numerical examples: the first one is "If 4 were a prime, there could be five prime numbers between 2 and 12", the second one is a set of three equations with two unknown variables (see Rescher 2005, 4-5). The reasonist definition forces such categorization because the mathematical examples are lessons learned by supposing a numerical value in the context of a set or an equation. On the other hand, the minimalist definition draws a strict line between thought experiments and mathematical examples. There is a difference when one is asked to imagine a situation and when one is instructed to do the math. The methodological instructions may be implicit and contextual, but it would definitely be a misunderstanding when one

[5] I am not convinced though that Rescher is liberal to that extent. The following passage suggests that a lesson is more than just a basic answer: "The larger lesson involved in a thought experiment does not lie in the substances of the consequences that follow from its launching suppositions, but rather in the wider ramifications of the higher–level circumstances that, given the wider cognitive context at issue, those consequences ensue from the supposition" (Rescher 2005, 13).

solves the Flying Pigs in a philosophy class by calculating the surface and the reflective properties of the average pig.

Is the minimalist definition too broad? I hope not. The minimalist definition is not arbitrarily restrictive as with the experimentalist one. However, it allows us to distinguish a thought experimentation from other types of problem solving, namely from a calculation.

4.2. What are the identity conditions for thought experiments?

Experimentalism and reasonism are "wide" approaches to thought experiments. They identify a thought experiment not only with a speculation, but also with the way the speculation is further used. For Sorensen, the identity conditions of a general experiment include the experimenter's intention (cf. Sorensen 1992, 133). Two identical experimental setups may constitute two different experiments depending on the hypothesis they are supposed to prove or disprove. To give Sorensen's own example, ringing a bell before feeding the animals and observing their behavior may constitute at least two different experiments—one testing the hypothesis that animals can hear and the other testing the hypothesis that animals can be conditioned. Since thought experiments belong to a class of experiments, the same identity condition applies: the setup (i.e. the absurd result *W* obtained from combination of content *C* and extraction *I*) can constitute multiple thought experiments. For instance, in Sorensen's reconstructions of the Gettier example the result that *Smith knows that p because of luck* rebuts the classic analysis of knowledge. However, it would be a different experiment, if the intention of the experimenter was, for instance, to support a claim about compatibility of knowledge and epistemic luck.

The reasonist approach leads to the same consequences as experimentalism. Rescher claims that a thought experiment is constituted not only by a supposition, a contextual background and a derived conclusion, but by a lesson as well. Therefore, a change of a lesson means a change of the identity of a thought experiment. Let us stay with the Gettier example: the conclusion that *Smith does not know p* may yield a lesson that *the classic analysis of knowledge is wrong* or a lesson that *we should be cautious about testimonial justification*. The two different lessons mean two different thought experiments.

Wide approaches have troubles with open-ended thought experiments, i.e. scenarios without explicit or implicit result. A prominent example is Lucretius' spear where one is asked what would happen to a spear thrown out at the edge

of the universe. Different answers mean different conclusions, different conclusions mean different lessons, and different lessons mean different thought experiments.

Minimalism holds that the identity conditions of a thought experiment are fully captured by *the modification and the baseline*. The result of the same thought experiment may be used in different arguments, it may change in time and it may be relative to a belief system, but none of that has any influence on the identity of the thought experiment itself. Therefore, the Gettier example may be included in many different arguments about the definition of knowledge or the justification condition or the role of certainty. Lucretius' spear is still the same experiment whether we guess that the spear would hit something or whether we guess that the spear would continue its movement.

4.3. Which level of analysis is appropriate?

Every thought experiment can be reconstructed in several ways. The reconstructions usually differ with respect to the superficial details of the scenario (e.g. "Smith believes that Jones will get the job" vs. "Smith believes that p"), but there could be more profound difference as well. Let me illustrate the point by two possible minimalist reconstructions of Unanticipated Results. The first reconstruction is simple and was already presented above:

i: People have capacity for foresight.
o: People are morally responsible.
m: People have no capacity for foresight.
r: People would not be morally responsible.

That is all. The question about moral responsibility is reconstructed as an internal part of the thought experiment. The second minimalist reconstruction of Unanticipated Results is more elaborate and much closer to Rescher's own analysis:

i: People have capacity for foresight.
o: People are able to predict the consequences of their actions.
m: People have no capacity for foresight.
r: People would not be able to predict the consequences of their actions.

The result is then qualified through argument from thought experiment to the principle constituting the following argumentation:

1. *Principle:* Foresight is a necessary condition for the ability to predict consequences of action.
2. Ability to predict consequences of action is a necessary condition for moral responsibility.
3. Therefore, people would not be morally responsible.

Which of the analyses is the proper one? The first one where a thought experiment leads us to a moral claim, or the second one where the experimental result is a psychological assessment?

Minimalism holds that there is one major rule of interpretation: the adequate reconstruction of a thought experiment reflects *the intention of the promoter*. Sometimes, the description of a thought experiment is complete or we can even inquire the honest promoter himself, thus we have an excellent source of information about the promoter's intentions. Other times, we have to cope with the limited contextual clues about the intended purpose of the thought experiment.

Let us say that in the case of Unanticipated Results we have nothing more to work with than the explicit instructions: *Suppose people have no capacity for foresight. Would they be responsible for their actions?* There is no mention about the ability to predict the consequences of actions. I would therefore suggest the first minimal reconstruction with a moral assessment. However, if the instructions were as follows: *Suppose people have no capacity for foresight. Would they be responsible for their actions, since they would not be able to predict the consequences of their actions?* I would prefer the second reconstruction with the argumentative extension.

4.4. What about other approaches to thought experimentation?

Minimalism is not a theory of epistemic value of thought experiment. Its purpose is to offer a framework for concise structural analyses based on widely accepted assumption that thought experimentation is a kind of imaginative reasoning. The key element is the careful distinction between doxastic process of belief elicitation through imagination, and argumentation based upon this belief. According to minimalism, thought experiments should be identified with the doxastic process regardless the superimposed argumentation. I will point out relations between minimalism and some distinct contributions to the debate about nature of thought experiments, namely

Norton's eliminativism and Gendler's constructivism. Let me demonstrate their approaches by the famous debate on Galileo's example with falling bodies.

Case V: Pisa Experiment

> Then if we had two moveables whose natural speeds were unequal, it is evident that were we to connect the slower to the faster, the latter would be partly retarded by the slower, and this would be partly speeded up by the faster. … But if this is so, and if it is also true that a large stone is moved with eight degrees of speed, for example, and a smaller one with four [degrees], than joining both together, their composite will be moved with a speed less than eight degrees. But the two stones joined together make a larger stone than the first one which was moved with eight degrees of speed; therefore this greater stone is moved less swiftly than the lesser one. But this is contrary to your assumption. So you see how, from the supposition that the heavier body is moved more swiftly than the less heavy, I conclude that the heavier move less swiftly. (Galileo 1974, 65)

John D. Norton advocates the radical view that all thought experiments are only "disguised arguments", i.e. arguments with premises about hypothetical states of affairs of particulars (Norton 1996, 336). His reconstruction of Pisa Experiments therefore emphasizes the elements of Galileo's general argumentation (Norton 1996, 341-342):

Norton's reconstruction of Pisa Experiment

1. The speed of fall of bodies in a given medium is proportionate to their weights.
2. If a large stone fall with 8 degrees of speed, a smaller stone half its weight will fall with 4 degrees of speed.
3. If a slower falling stone is connected to a faster falling stone, the slower will retard the faster and the faster speed the slower.
4. If the two stones of 2 are connected, their composite will fall slower than 8 degrees of speed.
5. The composite of the two weights has greater weight than the larger.
6. The composite will fall faster than 8 degrees.
7. Conclusion 4 and 6 contradict.
8. Therefore, we must reject Assumption 1.
9. Therefore, all stones fall alike.

The reconstruction is motivated by Norton's view on epistemology of thought experiments. Premises about particulars (i.e. premises 2, 4 and 6) are irrelevant to the conclusion in the sense that they can be replaced by non-particular premises without a loss of demonstrative power. Epistemic value of thought experiments is thus equal to the general arguments behind the particular cases.

Tamar Szabó Gendler refuses Norton's claims about epistemic irrelevance of particular details in two steps. At first, she offers her own stark argumentative reconstruction of the experiment (Gendler 1998, 404):

Gendler's reconstruction of Pisa Experiment

1. Natural speed is mediative.
2. Weight is additive.
3. Therefore, natural speed is not directly proportional to weight.

At second, she claims that the reconstruction does not grasp everything we learn from the experiment. By contemplating the scenario we get to see the proper „way out"—the adjustment of stipulated theory needed to resolve the paradox. The stark argumentative reconstruction leaves many ways out open, whereas the thought experiment specifies which claim needs to be abandoned.

What can minimalism add to this debate concerning epistemology of thought experiments? I will reconstruct Pisa experiment and show that both Norton and Gendler might be right. In fact, I will present *three* possible minimalistic reconstructions; let me start with the most opulent one:

Minimalistic reconstruction of Pisa Experiment I

(The first thought experiment)
i1: The faster stone is connected to the slower stone.
o1: The slower stone will slow down the faster stone.
m1: The faster falling stone is connected to the slower falling stone.
r1: The slower falling stone would slow down the faster falling stone.

(The second thought experiment)
i2: Two objects are connected.
o2: Their composite is heavier then its part.
m2: The falling stone is connected to the other falling stone.
r2: The composite of two falling stones is heavier than its part.

(Argument)

1. A composite is heavier than its part. *[from r2]*
2. Speed of fall is proportional to weight. *[assumption]*
3. The speed of falling composite would be higher than the speed of its falling part. *[from r2 and 2]*
4. The speed of falling composite would be lower than the speed of its falling part. *[from r1]*
5. Therefore speed of fall is not proportional to weight.

As said before, minimalism draws a strict line between imagination-based belief acquisition and argumentation. The reconstruction above stresses the role of imagination in Galileo's case. There are two thought experiments which serve as a basis for *reductio*. The reconstruction supports Gendler's critique of eliminativism: these experiments are crucial for backing up premises 1 and 4 in the argument, thus leaving only the premise 2 open for rebuttal.

Another possible reconstruction simply omits the second thought experiment. Let me therefore just repeat the first thought experiment and show the slight change in the argument:

Minimalistic reconstruction of Pisa Experiment II

(Thought experiment)

i1: The faster stone is connected to the slower stone.
o1: The slower stone will slow down the faster stone.
m1: The faster falling stone is connected to the slower falling stone.
r1: The slower falling stone would slow down the faster falling stone.

(Argument')

1. A composite is heavier than its part. *[assumption]*
2. Speed of fall is proportional to weight. *[assumption]*
3. The speed of falling composite would be higher than the speed of its falling part. *[from 1 and 2]*
4. The speed of falling composite would be lower than the speed of its falling part. *[from r1]*
5. Therefore speed of fall is not proportional to weight.

This second reconstruction is probably the most mundane one. There is only one thought experiment involved, the *reductio* is established partly upon experimental result, partly upon assumption about additive nature of weight.

The third possible reconstruction is radical:

Minimalistic reconstruction of Pisa Experiment III

(Argument'')

1. A composite is heavier than its part. *[assumption]*
2. Speed of fall is proportional to weight. *[assumption]*
3. The speed of falling composite would be higher than the speed of its falling part. *[from 1 and 2]*
4. The slower falling object would slow down the connected faster falling object. *[assumption]*
5. Therefore speed of fall is not proportional to weight.

According to this view, there is no imagination involved in Pisa Experiment. We are dealing with general argument based upon two assumptions presented as obvious (premises 1 and 4) and one theoretical postulate (premise 2). There would be no thought experimentation in Galileo's example under this reconstruction. This perspective supports Norton's approach: details about falling objects (i.e. their weight and the fact that they are stones) are irrelevant to the conclusion. Particulars play no epistemically relevant part in Pisa Experiment.[6]

Minimalist reconstructions of Pisa Experiment show us how the debate betwen Norton and Gendler boils down to the interpretation of two sentences:

> *Then if we had two moveables whose natural speeds were unequal, it is evident that were we to connect the slower to the faster, the latter would be partly retarded by the slower, and this would be partly speeded up by the faster.*
>
> *But the two stones joined together make a larger stone than the first one which was moved with eight degrees of speed.*

Does the protagonist instruct us to imagine something or is he just explicitly pointing out shared background knowledge? Minimalist framework itself does not tell us which interpretation is appropriate; it does, however, show how

---

[6] Gendler (1998, 408) claims that „no austere argumentative reconstruction will be able to do this, because part of the thought experiment's function is to bring the Aristotelian to accept certain premises". I have shown that such reconstruction is in fact possible. For more discussion see Norton (2004) or Picha (2011a).

interpretation of the two sentences translates into Norton's view and Gendler's critique. Personally, I am inclined to interpret these two sentences as postulates. They are evident and so simple that further appeal to imagination would be redundant. I favor the third reconstruction of Pisa Experiment; there is no thought experiment involved, Galileo presents a general argument with minor illustrations and Norton's eliminativist approach to Pisa Experiment is sound. However, I strongly disagree with Norton's transfer of the succesful elimination of Pisa Experiment to other thought experiments. In many cases we are explicitly instructed to *imagine* a scenario and then transform the result into a premise. Thoughts experiments frequently provide support for less obvious premises and arguments would be significantly weaker without them. On a larger scale, I therefore adopt Gendler's claim on the indispensability of thought experiments.

Let me again point out the fact that minimalism provides an analytical framework and as such is neutral to the debate about justificatory power of thought experimentation. Minimalism has therefore no necessary connections to Mach's conception of instinctive knowledge, Kuhn's conceptualism, Brown's platonism, Williamson's deflationary account and so on.

5. Summary

Thought experiments are hard to deal with. They involve bold suppositions formulated in vague language with distracting details and implicit elements prone to misinterpretation. I have presented three attempts to overcome such difficulties by informal reconstruction of a thought experiment. *Sorensen's experimentalism* emphasizes the compliance between real and thought experiments. Hypothetical scenarios are means of refutation with the structure of paradox. *Rescher's reasonism* focuses on inferring a broad conclusion from a supposition. I argue that both attempts are based on a wide analysis of the notion of thought experiment. Both experimentalism and reasonism therefore motivate reconstructions that include not only the thought experimental core, but the argumentative superstructure as well.

I propose a minimal model for reconstruction, which draws a strict line between the thought experiment itself and the way the experimental result is used. There are two key components: *modification* and *baseline*. A modification instructs an experimenter what to imagine; a baseline tells him what to do

and what to pay attention to. According to minimalism, the modification and the baseline fully capture all the dialectically relevant features of a thought experiment. I further argue that a dismissal of thought experiments is more likely a disagreement with the argument built upon a trivial experimental result. Thought experimentation is more frequent and less suspicious than expected.

References

ENGEL, P. (2011): Philosophical Thought Experiments: In or Out of the Armchair? In: Ierodiakonou, K. & Roux, S. (eds.): *Thought Experiments in Methodological and Historical Contexts.* Leiden: Brill, 145-164.

GALILEI, G. (1974): *Two New Sciences.* Madison: University of Wisconsin.

GENDLER, T. S. (1998): Galileo and the Indispensability of Scientific Thought Experiment. *The British Journal for the Philosophy of Science* 49, No. 3, 397-424.

GENDLER, T. S. (2000): *Thought Experiment: On the Power and Limits of Imaginary Cases*. New York: Routledge.

MACH, E. (1976): *Knowledge and Error. Sketches on the Psychology of Enquiry*. Dordrecht: D. Reidel.

NORTON, J. D. (1996): Are Thought Experiments Just What You Thought? *Canadian Journal of Philosophy* 26, No. 3, 333-366.

NORTON, J. D. (2004): On Thought Experiments: Is There More to the Argument? *Philosophy of Science* 71, No. 5, 1139–1151.

PICHA, M. (2011a): How to Reconstruct a Thought Experiment. *Organon F* 18, No. 2, 154-188.

PICHA, M. (2011b): *Kdyby chyby: epistemologie myšlenkových experimentů* [*What-If Errors: An Epistemology of Thought Experiments*]. Olomouc: Nakladatelství Olomouc.

PICHA, M. & PICHOVÁ, D. (2013): *100 myšlenkových experimentů ve filozofii* [100 *Thought Experiments in Philosophy*]. Praha: Dybbuk.

POPPER, K. R. (1992): *The Logic of Scientific Discovery*. London: Routledge.

RESCHER, N. (2005): *What If?* New Jersey: Transaction Publishers.

SORENSEN, R. (1992): *Thought Experiments.* Oxford: Oxford University Press.

Organon F 23 (4) 2016: 525-538

The Rejection of Fatalism about the Past

GAL YEHEZKEL

Department of B.A. in Liberal Arts and Sciences. The Sapir Academic College
D.N. Hof Ashkelon 79165. Israel
Department of History, Philosophy and Judaic Studies. The Open University of Israel
The Dorothy de Rothschild Campus. Ravutski Street 108. P. O. Box 808, Raanana 43107. Israel
Yehezkel.G@gmail.com

RECEIVED: 27-04-2016 • ACCEPTED: 03-10-2016

ABSTRACT: In this paper I defend the rejection of fatalism about the past by showing that there are possible circumstances in which it would be rational to attempt to bring about by our decisions and actions a necessary and sufficient condition, other things being equal, for something which we see as favorable to have occurred in the past. The examples I put forward are analogous to our attempts to bring about the occurrence of future events, and demonstrate the symmetry between the past and the future in this respect.

KEYWORDS: Fatalism – Future – Past – Richard Taylor – Symmetry.

1.

According to a common sense view there is a fundamental asymmetry between the past and the future. This asymmetry is reflected by the fact that people usually reject fatalism about the future while adopting fatalism about the past. In this paper I defend the rejection of fatalism about the past. My rejection of fatalism about the past relies on the rejection of fatalism about the future. I shall not try to defend the rejection of fatalism about the future in the scope of this paper, and assume that it is rational to attempt to bring about the occurrence of future events. I shall merely argue that the rejection of fatalism about

the future implies the rejection of fatalism about the past. While there is an asymmetry between the past and the future in respect of their relative independence of our decisions and actions, for the past is more independent of our decisions and actions than the future, this is a difference of degree rather than a difference of kind.

The term "fatalism" is used in different ways by different writers and in different contexts. Taylor defines fatalism about the future as the view that one cannot do anything about the future (cf. Taylor 1962a, 56). If this view is not vacuous, it must have a practical implication on the way we conduct ourselves. Hence, fatalism about the future is the view that there is no point in attempting to bring about, by our decisions and actions, future occurrences. In analogy, fatalism about the past is therefore the view that it is irrational to attempt to bring about, by our decisions and actions, past occurrences. Hence, rejecting fatalism about the future implies rejecting the analogous attitude towards the past. (For a similar claim, see Markosian 1995).

I should stress that I do not argue that we have the power to change the past, that is, to make happen what did not in fact happen, as Taylor sometimes carelessly describes the rejection of fatalism about the past (see Taylor 1962b, 26). This involves a contradiction, just like the analogous claim about the future, which states that we can change the future (cf. Makepeace 1962, 29).

Furthermore, I do not argue that it is possible to influence the past—a claim which involves a reversed causality. The possibility of reversed causality was interestingly suggested by Taylor himself in an earlier paper (see Chisholm & Taylor 1960). However, I believe that the idea of a reversed causality involves a contradiction. This is not due to any inherent asymmetry between the past and the future, but because of an asymmetry in the concept of causality, according to which a cause always precedes its effect (cf. Hume 1978, 76). I shall not attempt, in the scope of this paper, to defend this claim. Due to the fact that the conceptual possibility of backwards causation is debatable, I shall assume for the sake of the argument that backwards causation does not occur.

Without committing to a general account of causation in terms of necessary and sufficient conditions, in many circumstances in which we attempt to bring about a future event we consider our action as a necessary and sufficient condition, other things being equal, for the occurrence of this event. Hence, I argue that just as it is rational in certain circumstances to attempt to bring about by our decisions and actions a necessary and sufficient condition for something to

occur in the future (that is, to affect the future), there are possible circumstances in which it is rational to attempt to bring about by our decisions and actions a necessary and sufficient condition for something to have occurred in the past. Notice that I do not argue that it is possible to bring about, by our decisions and actions, past occurrences. This claim involves backwards causation, due to the fact that "bring about" is a causal term, which in this paper I assume is impossible. Rather, I argue that it is possible to *bring about future* occurrences, whether they are our actions or effects of our actions, that are a necessary and sufficient condition for something to have occurred in the *past*.

This analogy may seem disappointing in light of previous attempts to establish the possibility of backwards causation. For an attempt to establish the symmetry between the past and the future which discards the element of causality seems to relinquish from the analogy an important aspect of our attempt to influence the future, and thus is in danger of reducing into triviality.

I agree that without the element of causality some of the examples found in the literature of alleged backwards causation reduce into triviality. However, this is not due to the lack of the causal element, but rather because they neglect the most important aspect of our attempts to influence the future. That is, they fail to show that there are possible circumstances in which one should *take into account*, while deciding how to act, the need to bring about a necessary and sufficient condition for the occurrence of past events. Nerlich, for example, who claims that we actually are continually doing things that make earlier things have happened, flatly denies that actions can ever be *intended* in order to make things have happened (cf. Nerlich 1994, 247; 269; 271). Lacking this element, all these claims remain purely theoretical, devoid of any implication on our attitudes towards past events or any practical implication on the way we conduct ourselves in the world.

I show in what follows that there are circumstances in which one *should take into account*, while considering what to do, the need to bring about by our decisions and actions a necessary and sufficient condition for a favorable past event. That is, in this paper I defend the rejection of fatalism about the past by showing that there are possible circumstances in which it would be rational to attempt to bring about by our decisions and actions a necessary and sufficient condition, other things being equal, for something which we see as favorable to have occurred in the past. The examples I put forward are analogous to our attempts to bring about the occurrence of future events, and demonstrate the symmetry between the past and the future in this respect.

In section 2, I argue that Taylor (1962a) has relied on a false analogy in order to support the seeming absurdity of rejecting fatalism about the past. I show that once the analogous scenario for our attempt to shape the future is correctly formulated, not only it does not describe an absurdity, it actually describes a surprisingly trivial scenario. In section 3, I present examples for possible circumstances in which it would be rational to attempt to bring about by our decisions and actions a necessary and sufficient condition for something to have occurred in the past. If there is an asymmetry between the past and the future in this respect, it is simply a matter of contingent fact that the future is more susceptible to our decisions and actions than the past. I summarize the conclusion of my paper in section 4.

2.

Let us begin by considering a classic example, which originates from Aristotle, regarding the occurrence of a future naval battle (cf. Aristotle 1963, 19a23). Suppose that a naval commander examines two possibilities. The possibilities are either to attack the enemy tomorrow or to prevent the occurrence of any naval battle. These possibilities depend on his actions. If he issues an order to the fleet to move ahead and attack the enemy a naval battle will occur tomorrow. If he does not issue the order the battle will not occur. He decides that it would be best to attack tomorrow. In order to ensure that a naval battle will occur tomorrow, he issues an order to the fleet to move ahead and attack the enemy.

Assuming determinism, it has been argued by Taylor (1962a) that the case should be symmetrical for the past. If one rejects fatalism about the future, one should also reject fatalism about the past. However, rejecting fatalism about the past seems untenable, as Taylor demonstrated by the use of an example. Hence, Taylor adopts fatalism about both the past and the future.

Taylor's argument for fatalism about the future was criticized extensively, and is widely agreed to be unsound (see, e.g., van Inwagen 1983, 43-50). On the other hand, Taylor's criticism of rejecting fatalism about the past is widely accepted. Most philosophers have denied the alleged symmetry between the past and the future (for a recent attempt, see Diekemper 2005). Nevertheless, there are philosophers who argue for the rejection of fatalism about the past (see, for example, Nerlich 1994, 251-253). To defend this option, I show in

what follows that Taylor has relied on a false analogy in order to support the seeming absurdity of rejecting fatalism about the past.

Taylor relies on the following example for demonstrating the alleged absurdity of rejecting fatalism about the past (Taylor 1962a, 59). Suppose that Taylor (Taylor uses the first person in his example) is about to open a newspaper. It is assumed that the headline would declare the occurrence of the battle only if it occurred yesterday, and declare otherwise if it did not occur. Let us call him seeing the first headline act *S*, and him seeing the second headline act *S′*. It is not in his powers, Taylor argues, to do *S*, or it is not in his powers to do *S′*. That is, it is not up to him what headline he will read. Taylor explains this inability by the fact that the sort of headline he would see depends on the occurrence of the naval battle, and that is not up to him (cf. Taylor 1962a, 60).

Obviously, one example cannot prove the fatalistic view of the past. There may be other examples which show that it is up to us, in certain circumstances, to influence what has happened in the past. I address this possibility in the next section of my paper. My current interest is to show that the example Taylor uses in order to demonstrate our inability to influence the past is not the correct analogy for an attempt to influence the future.

To begin with, it should be noted that Taylor's description of his actions actually insures that it is not up to him whether he does *S* or *S′*. For it is up to an agent to decide whether to read the newspaper or not, but it is not up to the agent what headline he reads. Our use of the senses depends (to some degree) on our will. We use the senses, however, to gather information about our environment, and in this respect we are merely passive observers. In this sense, it is not up to anyone what headline he reads. It is also not up to the naval commander what headline he reads about the occurrence of tomorrow's naval battle, even if it is up to him whether the naval battle occurs tomorrow.

Obviously, there is a sense in which it is up to us what we perceive. That is, we can influence what we perceive if we can influence our environment. If I can paint my walls white, there is a sense in which it can be said that it is up to me what I see when I watch the walls.

According to this interpretation of Taylor's example, it is not up to Taylor what headline he will read simply because it is not up to him whether the naval battle occurred yesterday. This is not because the past is determinate, but rather because he is not, for example, in command of the fleet. Hence, it is also not up to Taylor what headline he will read even if it is assumed that the headline

correctly predicts whether the naval battle will occur *tomorrow*, simply because it is not up to him whether the naval battle will occur tomorrow. It is up to the naval commander, on the other hand, what headline he will read, in the sense that he can directly influence the occurrence of the naval battle tomorrow, by issuing an order to the troops for example.

Furthermore, Taylor's example about the occurrence of the naval battle yesterday is not analogous to his example about the occurrence of the naval battle tomorrow in another important respect. When one attempts to influence future occurrences, one wishes at time t_1 to perform at time t_2 an action A, which is a necessary and sufficient condition, other things being equal, for the occurrence of O at time t_3, which one sees (at t_1) as favorable, such that $t_1 < t_2 < t_3$.

Taylor's example for demonstrating the fatalistic view of the past, on the other hand, describes a scenario in which one wishes at time t_1 to perform at time t_2 an action A, which is a necessary and sufficient condition, other things being equal, for the occurrence of O at time t_3, such that $t_3 < t_1 < t_2$. In analogy, this is just like pointing to the present inability of the naval commander to have performed a past action which is a necessary and sufficient condition for a future occurrence.

After recognizing the shortcomings of Taylor's example for demonstrating the fatalistic view of the past, let us try to describe the correct analogy for the naval commander's attempt to influence occurrence of the naval battle tomorrow: The naval commander wishes now (time t_1) to have performed at time t_2 an action A, which is a necessary and sufficient condition, other things being equal, for the occurrence of O yesterday (time t_3), such that $t_3 < t_2 < t_1$. This action, for example, could be issuing an order this morning (t_2) for the troops to regroup following their victory.

Now that the correct analogy has been formulated, it is clear that it does not involve any contradiction or absurdity. Notice that the epistemic ignorance of the naval commander, about the occurrence of the naval battle yesterday and his actions this morning, is necessary for the analogy. For it is meaningful for the naval commander to attempt to act in order to bring about the occurrence of the naval battle tomorrow only if he does not already know that these occurrences will happen.

Admittedly, it looks strange that the subject does not know what he did this morning, but this is surely not an impossibility. This is clear if we consider the case in which an aging naval commander is reflecting on his long military career. He does not remember whether a certain naval battle occurred, about 50

years ago, and tries to remember his own actions. He seems to remember that he has issued an order for the troops to regroup following their victory, and hopes that he did issue that order.

If there is an asymmetry between the past and the future in this respect, therefore, it is merely an asymmetry in our knowledge of past and future occurrences, and especially of our own actions. We simply know more about the past than we do about the future. I return to this point in what follows.

It might, however, seem that something is missing from this analogy. When the naval commander contemplates the possible occurrence of a future naval battle, he wishes to bring about the occurrence of this battle with the help of his act. He *intends* to perform an action that would bring about the occurrence of the naval battle tomorrow. On the other hand, when the naval commander contemplates the possible occurrence of a past naval battle, he merely *wishes* that it had occurred.

Admittedly, the term "intends" cannot be applied equally to past actions. The naval commander intends to perform a future action, but cannot "intend to have performed" a past action. However, rather than reflecting an asymmetry between the past and the future, this impossibility reflects an asymmetry in the concept of causality. The analysis of intentions goes beyond the scope of this paper, but what is relevant in our case is a certain "pro" attitude, perhaps a wish, towards a certain action, which will be fulfilled by causing this action.

As I have already made clear, I believe that reversed causality contradicts an inherent asymmetry in our concepts of cause and effect. One cannot therefore "intend to have performed" a past action. This asymmetry, however, does not reflect any asymmetry in our conception of the past and the future. It simply reflects an asymmetry in the concept of causality. The previous examples show that just as events can depend on previous actions and wishes of an agent, events can also depend on subsequent actions and wishes of an agent.

3.

The conclusion of the previous section may seem disappointing, in light of the seeming triviality of the analogy for our attempts to influence the future. Although it sheds light on the symmetry between the past and the future

it does not imply any change in our attitude towards the past. More specifically, the analogy which is revealed so far does not show that there are circumstances in which one should take into account, while considering what to do, the need to bring about a necessary and sufficient condition for a favorable past event.

Furthermore, it might be expected, given the symmetry between the past and the future, that this possibility would be available. For if later actions can constitute necessary and sufficient conditions, other things being equal, for previous occurrences, why is it not the case that we are taking into consideration, while planning our future actions, the need to bring about the necessary and sufficient conditions for the occurrence of favorable past events?

In this section I show that there is an analogy for our ability to influence future events, which is directed towards past events. Just as it is rational in certain circumstances to attempt to bring about, by our decisions and actions, a necessary and sufficient condition for something to occur in the future (that is, to influence the future), there are possible circumstances in which it would be rational to attempt to bring about, by our decisions and actions, a necessary and sufficient condition for something to have occurred in the past.

Let us begin by describing an example for an attempt to bring about, by the agent's decisions and actions, a necessary and sufficient condition for the occurrence of future events, and then attempt to describe an analogous example, in which it would be rational to attempt to bring about, by the agent's decisions and actions, a necessary and sufficient condition for the occurrence of past events.

Consider the following example. Naomi visits the doctor, who informs her that she has a high level of cholesterol and she must keep a strict diet from now on if she wants to increase her chances of living a long and healthy life. Naomi, who read philosophy for her B.A., points out to the doctor that the future is determinate, and therefore it is irrational on her part to attempt to do anything about the future. As it is already determined that she will, or will not, live a long and healthy life, she does not need to change her diet. It is irrational on her part to start a diet.

However, the doctor points out to Naomi that the fact that the future is determinate does not imply that her decisions and actions are useless. The claim that her future is determined causally, for example, implies that future occurrences are determined at least partly by her choices and actions. It is a fact that people who keep a strict diet tend to live longer and enjoy better health, while

those who do not suffer from a variety of diseases. The fact that her decisions are themselves determinate is neither here nor there. Deciding to begin a strict diet and keeping to it is a necessary and sufficient condition for a better future, and so this is the rational choice for her to make. Furthermore, although her decision may itself be determinate, it is not something that is independent of her own considerations. If she is convinced by the fatalist argument, she is doomed, if not, she can expect good health. The rational thing for her to do is therefore to make the decision and keep her diet. I think most of us would agree that the doctor is right, and so the rational thing for Naomi to do is to accept the doctor's advice and begin to diet.

Consider now another scenario. Naomi goes to the doctor. The doctor informs her that in light of a history of high cholesterol levels in her family she herself is more likely to suffer from a high level of cholesterol. Hence if she wants to live a long and healthy life she must follow a strict diet. The doctor also tells her about a recent medical discovery. Biologists have discovered that the correlation between a high level of cholesterol and diet is not the result of a direct link between the two. A third element is involved. A gene has been identified, which is at once associated both with a high level of cholesterol and a general lack of care for health. Research has shown that people who are concerned about their health and who do manage to follow a strict diet over a long period do not carry this gene, and therefore also enjoy low levels of cholesterol and good health. People who are not concerned about their health, and fail to follow a strict diet over a long period, do carry this gene, and therefore also have high levels of cholesterol. Hence, deciding to begin a strict diet and keeping to it is a necessary and sufficient condition for a better future. The doctor therefore recommends Naomi to keep a strict diet, in order to ensure that she will enjoy a long and healthy life.

Naomi might object that the doctor is advising her to do something impossible, that is, to affect the past. Surely, if she carries this gene from birth, it is irrational of her to attempt to avoid inheriting this gene by her present decision and future actions. Either she has already inherited this gene, and therefore it is pointless for her to keep a strict diet, or she did not inherit this gene, and therefore it is pointless for her to keep a strict diet. The most that deciding and keeping her diet can do is indicate whether she carries this gene or not, but she cannot do anything about the past. That is, the only function that her attempt to keep a strict diet can serve is epistemological, that is, it can teach her that she does not carry this gene, and that she can expect to live a long and healthy life.

In response, the doctor tells Naomi that backwards causation is not involved here. The relevant factor is that there is a natural law which establishes a link between caring about one's health by adhering to a strict diet and the absence of this gene. If Naomi decides to the contrary, regardless of her reasons (for example, she might be persuaded by fatalistic considerations about the past), she carries the gene, and her future is not bright. If she does decide to keep a strict diet, she does not carry this gene, and will enjoy a long and healthy life. If the rational thing for Naomi to do in the previous example is to accept the doctor's advice and begin to diet, by analogy it is also the rational thing to do in this example.

Some may object that her decision to keep a strict diet is merely a way of finding out whether she carries the gene or not (Dummett, for example, discusses a similar objection—see Dummett 1954, 35-37). In this respect this example seems asymmetrical with attempts to influence the future, for in these latter cases we are trying to bring about an occurrence, rather than merely discover whether it is about to occur or not.

In order to tackle this objection, the difference between attempting to bring about something and merely trying to find out whether something occurs or not should be clarified. To begin with, if the naval commander wishes to bring about the occurrence of the naval battle tomorrow, he attempts to perform an action which is a necessary and sufficient condition for the occurrence of the naval battle. For example, he can issue a command for the troops to attack the enemy fleet. If the naval commander merely wants to know whether the naval battle will occur tomorrow, he will usually perform an action which is a sufficient condition for *knowing* whether the naval battle will occur or not. For example, he can read intelligence reports on the position and possible intentions of the enemy. Notice that his action is not a sufficient condition for the battle to occur or not.

Similarly, if Naomi only wants to find out whether she carries the gene or not she can perform a genetic test. Her taking the genetic test is neither a sufficient condition for her carrying the gene, nor for her not carrying the gene (notice the analogy for reading the headline in order to find out whether a naval battle occurred yesterday or not). If Naomi keeps a strict diet, on the other hand, it is a necessary and sufficient condition for not carrying the gene. Hence, her attempt to keep a strict diet is not merely an attempt to discover whether she carries the gene or not, and is symmetrical with an attempt to bring about future occurrences.

In many cases, therefore, when one attempts to find out whether an event is about to happen, one's actions are not a sufficient condition for the occurrence or the non-occurrence of this event. There are, however, cases in which one attempts to perform an action which is a sufficient condition for the occurrence of an event, in order to find out whether this event occurred or not. For example, one can attempt to raise one's right hand in order to find out whether the surgery which was supposed to recover the mobility in the right arm was successful. Surely, in this case one is not trying to secure the success of the surgery that has been undergone, but merely to find out whether the surgery was successful or not.

The difference between the last example and Naomi's attempt to keep a strict diet is that Naomi is trying to provide a *necessary* and sufficient condition, other things being equal, for not inheriting this gene. The raising of the arm, on the other hand, is merely a sufficient condition for the success of the surgery. Hence, if one does not want to know whether the surgery was successful, and refuses to attempt to raise one's arm, one's decision and actions do not imply anything regarding the success of the surgery. In this case it is justified to say that an attempt to raise the arm is merely a way to find out whether the surgery was successful or not.

Naomi's attempt to keep a strict diet is therefore unlike an attempt to discover whether she carries a gene which is associated, for example, with the Tay-Sachs disease. The existence of the Tay-Sachs gene does not depend on her decision to examine its existence, or indeed on any decision or action she makes. Hence, her only consideration is epistemological, that is, whether she wants to know if she carries the gene or not. However, Naomi's attempt to keep a strict diet is different, and she should consider the fact that her decision and action can constitute a necessary and sufficient condition for her not carrying this gene, which would ensure that she lives a long and healthy life, or for her carrying this gene. It is an attempt to bring about by our decisions and actions a necessary and sufficient condition, other things being equal, for something to have occurred in the past.

So far I have described a scenario in which we aim to secure favorable future occurrences, that is, a long and healthy life. For this purpose, Naomi attempts to bring about, by her actions and decisions, a necessary and sufficient condition for something to have occurred in the past. This is not the only possible scenario in which one may be required to take into consideration, while deciding how to act, the need to bring about a necessary and

sufficient condition for something to have occurred in the past. Another possible scenario is one in which the subject's goal is a favorable past occurrence.

Consider the following scenario. Lucy is seeing a psychologist. She is afraid that she was abused as a child, but does not remember clearly what happened. In response, the psychologist tells her of a recent discovery she made. While attempting to use bungee jumping as a method for teaching her patients how to control their fears she uncovered a surprising correlation between refusal to engage in bungee jumping, for any reason at all, and a background of abuse in childhood. All and only the people who were abused in childhood, whether they had recollection of this abuse or not, either rejected the suggestion outright or withdraw at the last minute from jumping, relying on countless excuses for why they should not engage in this activity. Although she does not know how to explain this correlation, it is an established correlation. The psychologist then suggests that Lucy should bungee jump, just in order to ensure that she was not abused as a child.

Lucy might object that the psychologist is advising her to do something impossible, that is, to affect the past. She obviously prefers not to have been abused as a child. However, this is not up to her now. Although her bungee jumping may have an epistemological significance, that is, it can teach her that she was not abused, it cannot affect the past.

In response, the psychologist denies that reversed causality is involved here. The only relevant factor here is a correlation between her current decisions and actions and her past. Furthermore, it is not merely a way of discovering whether she was abused or not. An x-ray of her hand is a way of discovering whether she broke her hand as a child or not. In this latter case, her taking the test, or her refusal to be examined, does not imply anything about whether she broke her hand as a child or not. The only relevant consideration is whether she wants to know if she broke her hand or not. However, her refusal to bungee jump would imply that she was abused as a child, while her bungee jumping would imply that she was not abused as a child. The reason she should bungee jump is not epistemological, but rather to ensure that she was not abused as a child. The rational thing for Lucy to do, if she wants to ensure that she was not abused as a child, is to decide to do the bungee jump, that is, to attempt to bring about a necessary and sufficient condition for favorable past occurrences.

The two examples I have put forward in this section show that there are possible circumstances in which it would be rational to attempt to bring about,

by our decisions and actions, a necessary and sufficient condition for something to have occurred in the past, just as it is rational to attempt to bring about, by our own decisions and actions, a necessary and sufficient condition for something to occur in the future. Hence it seems that the only asymmetry between the past and the future in this respect lies in a contingent fact, according to which the past is more independent of our decisions and actions than the future.

4.

The conclusion of this paper is that there are possible circumstances in which it would be rational to attempt to bring about, by our decisions and actions, a necessary and sufficient condition, other things being equal, for something to have occurred in the past. Although these attempts do not involve backwards causality, they are symmetrical with our attempts to bring about, by our own decisions and actions, a necessary and sufficient condition for something to occur in the future. Although the past is more independent of later intentions and actions than the future is independent of previous intentions and actions, this difference is a difference of degree rather than a difference of kind. That is, the difference between the past and the future in this respect rests on a contingent fact, rather than a genuine asymmetry between the past and the future.

References

ARISTOTLE (1963): *Categories and De Interpretatione*. Oxford: Oxford University Press.

CHISHOLM, M. R. & TAYLOR, R. (1960): Making Things to Have Happened. *Analysis* 20, No. 4, 73-78.

DIEKEMPER, J. (2005): Logical Determinateness, Fixity, and the Symmetry of Time. *Philosophical Papers* 34, No. 1, 1-24.

DUMMETT, M. (1954): Can an Effect Precede Its Cause? *Proceedings of the Aristotelian Society, Supplementary Volume* 28, 27-44.

HUME, D. (1978): *A Treatise of Human Nature*. Oxford: Oxford University Press.

MAKEPEACE, P. (1962): Fatalism and Ability II. *Analysis* 23, No. 2, 27-29.

MARKOSIAN, N. (1995): The Open Past. *Philosophical Studies* 79, No. 1, 95-105.

NERLICH, G. (1994): *What Spacetime Explains*. Cambridge: Cambridge University Press.
TAYLOR, R. (1962a): Fatalism. *The Philosophical Review* 71, No. 1, 56-66.
TAYLOR, R. (1962b): Fatalism and Ability I. *Analysis* 23, No. 2, 25-27.
VAN INWAGEN, P. (1983): *An Essay on Free Will*. New York: Oxford University Press.

Organon F 23 (4) 2016: 539-550

Δ-TIL and Problems of Deontic Logic

VLADIMÍR SVOBODA

Institute of Philosophy. Czech Academy of Sciences
Jilská 1. 110 00 Prague 1. Czech Republic
svoboda@site.cas.cz

In two articles which recently appeared in *Organon F*, Daniela Glavaničová outlined Δ-TIL—a theory aiming at extending the apparatus of Transparent Intensional Logic (TIL) so that it addresses (and ideally solves) problems that arise within deontic logic. The ambition of the first paper (published in Slovak) is to present a suitable logical analysis of *deontic modalities* in terms of TIL. Its main contribution, according to Glavaničová, consists in "offering a semantically based distinction between implicit and explicit deontic modalities" (Glavaničová 2015, 211). The goal of the second article is "to amend the former analysis of deontic modalities in terms of TIL to incorporate both the standard (relativistic) view and the minimal semantics of TIL" (Glavaničová 2016, 204).

The general point of the articles can be viewed from two somewhat different perspectives. If we are primarily interested in the logical theory designed by Pavel Tichý and developed by his followers,[1] we can view the articles as an attempt to extend the analytical potential of TIL so that the theory allows for an illuminating analysis of sentences that have not yet been the focus of attention of TIL adherents. We can thus see the papers primarily as aiming at advancing TIL as an analytical tool. If we are mainly interested in deontic logic, we can view the articles as an attempt to address some problems that trouble deontic logic and to outline their solution. Of course, these two perspectives are not mutually exclusive and we can conjecture that Glavaničová's aspiration

[1] The crucial reference works here are Tichý (1988) and Duží, Jespersen & Materna (2010).

was twofold—to contribute to the development of TIL as well as to the development of deontic logic.

In this discussion paper, I would like to briefly consider to what extent the goal of the papers has been fulfilled. My worry is that the contribution of the paper to the development of TIL is less significant than Glavaničová thinks and that her conviction that the analytical insights which she outlines open a pathway to solving some (or perhaps even most) paradoxes of deontic logic is based on a misperception of the nature of the problems that arise within deontic logic.

At the very beginning, I should stress that Glavaničová limits her attention to the analysis of deontic modalities—sentential modifiers that occur in sentences of the form "It is obligatory that φ", "It is forbidden that φ", "It is permitted that φ" (symbolically $O\varphi$, $F\varphi$, $P\varphi$) in which φ stands for sentences or statements. The resulting compound sentences are interpreted descriptively. This means that they have certain truth-conditions, i.e. they are supposed to be true or false (in a given context). It is thus important to remember that when Glavaničová speaks about deontic logic what she has in mind is a narrowly conceived deontic logic—the logical theory of sentences which don't have prescriptive (action-guiding) meaning but which describe—adequately or not—the (or a) normative situation.

The central contribution of the first article consists in the suggestion that we should treat sentences of the form

F1 "It is obligatory that φ"

as systematically ambiguous. Sentence form F1 can be interpreted as synonymous with the sentence form

$F1_{imp}$ "It is implicitly obligatory that φ"

or as synonymous with the sentence form

$F1_{exp}$ "It is explicitly obligatory that φ".

The operators "it is implicitly obligatory that" (we can concisely write "obligatory$_{imp}$") and "it is explicitly obligatory that" ("obligatory$_{exp}$") are to be carefully distinguished as they denote objects of different types. Let us take for example the sentence

S1 "It is obligatory that Pavel is silent"

The sentence can be—correctly—analysed either by the formula

$\text{TILS1}_{\text{imp}}$ $[{}^0O_{wt}\,[\lambda w \lambda t\,[{}^0Silent_{wt}\,{}^0Pavel]]]$

or by the formula

$\text{TILS1}_{\text{exp}}$ $[{}^0O^*_{wt}\,{}^0[\lambda w \lambda t\,[{}^0Silent_{wt}\,{}^0Pavel]]]$.

The operator O employed in the first analysis represents a property of propositions (this is within the technical notation of TIL suggested by the following type analysis $O/(\text{oo}_{\tau\omega})_{\tau\omega}$).[2] The operator O^* employed in the second analysis represents a property of propositional constructions $(O^*/(\text{o*n})_{\tau\omega}$ where *n is the type of constructions—the objects which have the central place in the semantics provided by TIL).

The truth conditions of the formulas representing the two alternative analyses of sentences like S1 are stated in analogous ways: Let ${}^0T$ constructs the truth-value True and ${}^0F$ constructs the truth-value False. Let, furthermore, C be a construction of a proposition and let the expression $\alpha : \beta$ is true if and only if α construes the same object as β (with respect to a valuation). The truth-conditions of sentences formed by means of O and O^* are then as follows:

${}^0T : [{}^0O_{wt}\,C]$ iff $C \in O_{wt}$
${}^0F : [{}^0O_{wt}\,C]$ otherwise.
${}^0T : [{}^0O^*_{wt}\,{}^0C]$ iff ${}^0C \in O^*_{wt}$
${}^0F : [{}^0O^*_{wt}\,{}^0C]$ otherwise.[3]

These definitions in effect say that if, e.g., the sentence "It is obligatory that Pavel is silent" is disambiguated as saying "It is implicitly obligatory that Pavel is silent", then it is true in the actual world just and only in case that the *prop-*

[2] We should note that the term "proposition" is within the present discussion consequently used as a technical term denoting any function from world courses to truth values.

[3] See Glavaničová (2016, 213). The definitions, in fact, seem somewhat suspicious to me. I, for example, don't see how a *construction* of a proposition could be a member of the set of (in this case obligatory) propositions.

osition denoted by the sentence "Pavel is silent" is among those which are implicitly obligatory in the actual world. Similarly, if the sentence "It is obligatory that Pavel is silent" is disambiguated as saying the same as "It is explicitly obligatory that Pavel is silent", then it is true in the actual world if and only if the *construction* constructed by the sentence "Pavel is silent" is among those which are explicitly obligatory in the actual world.

Of course these definitions would be entirely uninteresting if they were not supplemented by some logical principles. These are provided by the following four rules[4]

(R1) $[{}^0O*_{wt}\ {}^0c] \vDash [{}^0O_{wt}\ c]$
(R2) (i) $\lambda w\lambda t\ [{}^0O*_{wt}\ {}^0c]$, (ii) $[{}^0{=}i\ {}^0c\ {}^0c'] \vDash \lambda w\lambda t\ [{}^0O*_{wt}\ {}^0c']$
(R3) (i) $[{}^0O_{wt}\ [\lambda w\lambda t\ [c_{wt} \to d_{wt}]]]$, (ii) $[{}^0O_{wt}\ c] \vDash [{}^0O_{wt}\ d]$
(R4) $[\forall^\omega w\forall^\tau t\ c_{wt}] \vDash [{}^0O_{wt}\ c]$

The first rule dictates that whenever some propositional construction is among those which are explicitly obligatory, the proposition constructed is among those which are implicitly obligatory. The second rule states that if some propositional construction is explicitly obligatory then so are all those which are procedurally isomorphic. (R3) and (R4) are analogues of two principles of *Standard Deontic Logic* (SDL). The first is a deontic version of *modus ponens* and the second is a deontic version of the modal rule of necessitation.

Now it is time to assess what kind of interesting insights the framework presented by Glavaničová provides. She tries to demonstrate its virtues by what she calls *Russell's test*.[5] Somewhat surprisingly, the whole testing consists in a discussion of how the inferential scheme called the *Ross paradox*

---

[4] In these rules $=i$ represents the relation of procedural izomorfism, and $\vDash$ represents the entailment relation among constructions. c, c' and d represent propositional constructions. Classical predicate logic is accepted as a background theory.

[5] She quotes the following passage from *On Denoting*: "A logical theory may be tested by its capacity for dealing with puzzles, and it is a wholesome plan, in thinking about logic, to stock the mind with as many puzzles as possible, since these serve much the same purpose as is served by experiments in physical science" (Russell 1905).

fares with respect to her distinction between implicit and explicit obligation.[6] The main result can be shortly presented in the following way:[7] while the inferences

RP1 "It is obligatory$_{imp}$ that Pavel is silent"

"It is obligatory$_{imp}$ that Pavel is silent or kills Richard"

and

RP2 "It is obligatory$_{exp}$ that Pavel is silent"

"It is obligatory$_{imp}$ that Pavel is silent or kills Richard"

are shown to be valid in Δ-TIL, the inference

RP3 "It is obligatory$_{exp}$ that Pavel is silent"

"It is obligatory$_{exp}$ that Pavel is silent or kills Richard"

is claimed to be invalid.[8] The point of Glavaničová's deliberations on the Ross paradox seems to consist in the claim that though RP2 is valid there is nothing paradoxical about this as Pavel can fulfil the implicit command described by the sentence "It is obligatory$_{imp}$ that Pavel is silent or kills Richard" but that this still does not mean that he fulfils the command which is described by the sentence "It is obligatory$_{exp}$ that Pavel is silent" (cf. Glavaničová 2015, 225).

---

[6] The original version of Ross paradox was presented in the form of the inference *Mail this letter!,* hence *Mail this letter or burn it!*, which was valid according to the prevailing accounts of the logic of imperatives (cf. Ross 1941). It was thus not straightforwardly relevant for *statements* about obligations.

[7] I won't follow Glavaničová's way of presenting the Ross paradox as I find it quite misleading. She introduces the paradox in the form of an argument which she labels as "intuitively invalid" but then she says that the argument consists in an unproblematic application of *modus ponens* (which is in fact the case).

[8] This conclusion is, in Glavaničová's text, presented in the form of the claim that we cannot prove the sentence "If it is obligatory$_{exp}$ that Pavel is silent then it is obligatory$_{exp}$ that Pavel is silent or kills Richard" within her system.

This outcome is somewhat puzzling. First, it is not clear what is meant by the phrase "implicit command described by the sentence …" The most plausible explanation is that Glavaničová presupposes that there exists a logic of commands which allows us to derive implicit commands from explicit ones and that these implicit commands can be described by sentences which are analysed by means of the "obligatory$_{imp}$" operator.[9] Unfortunately, the papers don't provide clues shedding light on the relationship between (explicit and implicit) commands and statements speaking about (explicit and implicit) obligations. Generally, we don't learn anything as to how the sets O^*_{wt} and O_{wt} are formed. This is a serious problem—most model situations against which we test acceptability of the principles of different systems of deontic logic involve discussion on morally relevant obligations, but it seems quite strange to distinguish between explicit and implicit obligations in moral discourse, so we seem to lack an intuitive grounding for these (quite essential) considerations.[10]

Another problem is that it is not at all clear why the fact that Pavel can fulfil the command in conclusion without fulfilling the one in the premise should guarantee that the inference leading to "It is obligatory$_{imp}$ that Pavel is silent or kills Richard" is not paradoxical. Pointing out that coping with the obligation described in the conclusion does not automatically exempt the obliged person from coping with the obligation mentioned in the premise is, of course, relevant,[11] but it would, as it seems, keep being relevant even if RP3 were a

---

[9] Glavaničová remarks that inferences consisting of commands are inherently problematic as commands are neither true nor false (Glavaničová 2015, 201), but she does not explicitly deny the possibility of a logic of commands.

[10] We might suppose that all sentences speaking about moral obligations are to be generally treated as speaking about implicit obligations. Then, however, Δ-TIL clearly leads us to conclude that "It is morally obligatory to kill a drowning child or to save it" is a true moral statement as soon as we suppose that "It is morally obligatory to save a drowning child" is true. This is likely to be hard to swallow for those who view the Ross paradox as a serious problem of the logic of deontic statements. (The more that statements like "It is morally obligatory to give the starving beggar food or to give him money so that he can buy some" seem perfectly reasonable as 'choice offering' moral claims.) For my analysis of the Ross paradox cf. Svoboda (2004; 2013).

[11] Arguments stressing this point were presented in the early eighties (at the latest) (cf. Castañeda 1981).

valid argument. Yet Glavaničová seems to presume that validity of RP3 would be highly problematic.

The general problem with Δ-TIL is that it is very weak. Let us suppose that the monastery code which is several times mentioned in Glavaničová's examples contains the sentence:

S2 "It is obligatory that monks fast and keep silent".

We may, moreover, have reasons to suppose that it is correct to disambiguate the sentence as saying

S2* "It is obligatory$_{exp}$ that monks fast and keep silent"[12]

Now, we can imagine Glavaničová's hero Pavel asking the truthful custodian Richard whether the code explicitly asks the monks to fast. Intuitively, it is quite obvious that the answer should be positive. But under Glavaničová's conception of explicit obligations we don't have any reason to affirm such an answer—the truth of S2* surely doesn't provide any substantiation for the claim that

S3 "It is obligatory$_{exp}$ that monks fast"

is true. (R2) is the only rule that governs inferences which have as conclusions statements about explicit obligations, and the constructions which follow "obligatory$_{exp}$" in the formal explication of S2* and S3 are obviously *not* procedurally isomorphic. This quite clearly indicates that what Glavaničová means by explicit obligations is remote from what we normally mean when we say that something is explicitly required or explicitly obligatory.[13]

---

[12] It seems very strange to admit that the code could contain the statement "It is obligatory$_{imp}$ that monks fast and keep silent". This, in fact, rather clearly suggests that 'statements' contained in such a code are in fact not to be seen as describing a code but as forming it (i.e. as—covertly—action guiding). Glavaničová, however, does not explain what she means by a code, so there is no ground for a discussion.

[13] What Glavaničová says in section 10 of her earlier article suggests that she is aware of the problems associated with her account of explicit obligations. She mentions the possibility of introducing semi-explicit modalities (obligations). The somewhat controversial advantage of introducing such modalities is that it would allow us (and force us) to disambiguate deontic sentences in three different ways. The main problem with this

Δ-TIL presented by Glavaničová (2015) is also weak in another respect. The sets O^*_{wt} and O_{wt} can include all propositional constructions, or all propositions respectively. In fact, it is easy to demonstrate that as soon as two statements of the forms $[{}^0O^*_{wt}\ {}^0c]$ and $[{}^0O^*_{wt}\ {}^0\neg c]$ are true (for example the statements "It is obligatory$_{exp}$ that Pavel is silent" and "It is obligatory$_{exp}$ that Pavel is not silent") then the set O_{wt} unavoidably contains all propositions, i.e. any proposition is implicitly obligatory and we face an 'explosion of obligations'. In most systems of deontic logic, the problem of 'dangerous explosion' is precluded by adoption of a principle assuring that contradictory obligations cannot arise—typically in the form of the axiom: $\neg(OA \wedge O\neg A)$. Glavaničová, however, explicitly refrains from adopting a principle of this sort and so the threat of explosion is an urgent problem for her Δ-TIL.

In her second article, Glavaničová seems to take a somewhat different stand. She says that if we don't assume that there is just one set O^*_{wt} and one set O_{wt} of explicit obligations, i.e. if we accept what she calls "deontic relativism", then the situation becomes different. Here she notes that "it is quite reasonable to demand that normative systems be internally consistent" (Glavaničová 2016, 206). Unfortunately, it is not clear how this requirement is to be reflected in Δ-TIL. As the initial step towards the relativist framework, she suggests that axioms and rules are "decorated with subscripts". Thus, instead of formulas of the form OA and O^*A we may use formulas O_xA and O^*_xA where x refers to certain normative system. In fact, the core of the second article consist in justifying of this kind of relativism.

My general impression here is that Glavaničová has been somewhat misled by the metaphysics behind the apparatus which she employs. The fact that in the first article she considers only one set O^*_{wt} and one set O_{wt} would normally be taken as being just a reasonable simplification—it is quite natural to suppose that whenever we speak about something being obligatory or permitted we talk within a given context. The fact that adherents of SDL don't explicitly relativize statements saying that something obligatory to a particular normative system surely is not to be taken as a testimony that their theory concerns (only) talk on 'absolute obligations'. It is thus surprising to see that Glavaničová felt pressed to interpret deontic modalities discussed in the first article as absolute

idea, of course, is that it is entirely unclear how a useful concept of semi-explicit modalities might be introduced.

(non-relative) in the suggested sense.[14] From an 'external' point of view her argumentation in favour of deontic relativism (a significant part of the second article consists of this argumentation) appears to be close to trivial and the adherents of objectivism that she mentions appear to be mere straw men. Moreover, what she says reveals a serious confusion on some issues. She suggests that adopting the relativized deontic modalities solves the paradox of contrary-to-duty obligations (Chisholm's paradox), which she presents in the following way (cf. Glavaničová 2016, 208):

(P1) Sophie shall not kill.
(P2) It ought to be that if Sophie does not kill, she is not punished for killing.
(P3) If Sophie kills, she ought to be punished for killing.
(P4) Sophie kills.

The statements presented in (P1)–(P4) describe a situation in which three claims containing deontic modalities are complemented by a factual claim. The situation described is unfortunate as the last statement suggests that Sophie did not meet her obligation and killed, but what has been said does not seem inconsistent. However, under its most plausible formalisation in SDL[15]

(P1') $O\neg A$
(P2') $O(\neg A \rightarrow \neg B)$
(P3') $A \rightarrow OB$
(P4') A

the set of sentences is inconsistent—the formula $OA \wedge O\neg A$ as well as the formula $OA \wedge \neg OA$ are derivable. This, of course, raises the question how the problem might be explained away. Glavaničová's solution is quite striking. She points out that the paradox can be solved "via deontic relativism"—we can treat the deontic modality contained in (P1) as referring to other normative

[14] In practice, however, she relativizes deontic modalities from the very beginning of the first article.

[15] Glavaničová, in fact, discusses a somewhat specific version of the paradox. Its original version and the related discussion can be found in Chisholm (1963) and Hilpinen (1971).

system than those deontic modalities contained in (P2) and (P3).[16] This is a bizarre solution. Glavaničová is surely right that if we treat the deontic statements as referring to different normative systems the inconsistency is avoided, but she does not provide a single reason why should we do that. The whole problem of Chisholm's paradox arises from the fact that all the mentioned sentences speaking about obligations seem reasonable from an intuitive point of view—in this case from a viewpoint of a person taking a coherent moral standpoint. Maybe Glavaničová has some substantiation for presuming that it is reasonable to treat the statements within the paradox in a relativistic way, but she does not mention any and I can't imagine any.

Acceptance of a kind of deontic relativism together with the requirement that normative systems are internally consistent seems to be a reasonable position. I am, however, afraid that Glavaničová will face problems if she wants to take seriously the idea that statements speaking about explicit obligations occurring in one normative system should not be in conflict. While formulating the consistency rule for statements on implicit (relative) obligations is unproblematic,[17] it is quite unclear how we could formulate a useful consistency rule assuring (deontic) consistency of statements about explicit obligations. By a "useful rule" I mean a rule that would, for example, allow us to identify the following three deontic statements as inconsistent:

S4 "It is obligatory$_{exp}$ that Pavel is silent"
S5 "It is obligatory$_{exp}$ that if Pavel works as the monastery porter, he is not silent"
S6 "It is obligatory$_{exp}$ that Pavel works as the monastery porter"

It is quite obvious that, if we analyse the sentences in the way suggested by Glavaničová, we will have to conclude that they are perfectly consistent. My conjecture is that if we analyse these sentences by means of the operator O^*

[16] She literally says "We can solve CTD problem via deontic relativism treating primary and secondary subsystems of certain normative systems as different normative systems". (Even if we had some reason to make the step she suggests, we should ask why inconsistency between norms belonging to two subsystems of a certain normative system should be tolerable.)

[17] For example, the principle $[^0O_{wt}\ c] \vDash \neg[^0O_{wt}\ \neg c]$ will do for the non-relativistic case.

where the type analysis is $O*/(o*n)_{\tau\omega}$ then the only way to identify the statements as conflicting is through the inconsistency of the implicit obligations derived by (R1). But, if the conflict among statements on explicit obligations is identified by means of the relation of conflict (inconsistency) among statements on implicit obligations, then it is difficult to imagine how the logical properties of the two kinds of statements could be so different as Δ-TIL in its present form suggests—the relation of normative conflict and logical inconsistency are obviously two sides of one coin in SDL and they are in an obvious way interconnected with the relation of entailment. Maybe the theory can be developed so as to provide a reasonable account of conflicts among statements speaking of explicit obligations, but the fact that this central issue has not been addressed at all by Glavaničová suggests how scanty Δ-TIL is.

The just outlined controversial features of Glavaničová's proposal (as well as some others which I am leaving out for brevity's sake)[18] lead me to conclude that Δ-TIL in its present form does not open very promising way to enhancing the analytic potential of TIL. Glavaničová has apparently been misled by the fact that the leading figures of TIL distinguish between explicit and implicit attitudes and within common normative discourse people speak about something being explicitly or implicitly ordered (ordered, requires) or permitted. This association between the two kinds of explicitness is, however, rather a terminological coincidence than a clue that deserves be taken seriously.[19] At the same time, I don't find in the paper any argument which would convince me that the apparatus of TIL is a suitable tool for addressing and solving the problems that have been discussed in deontic logic during the decades of its development. I am not saying that TIL cannot serve as a basis for developing a framework that will allow for illuminating analyses of deontic modalities, I

[18] I, for example, find Glavaničová's (2016, 215) proposal to treat normative system (provisionally) as individuals (along with humans, mountains and pieces of furniture) as highly problematic. It is difficult to imagine how, e.g., a moral code of some society could reasonably be explicated as an individual. Her other suggestion, namely, to add a further atomic type—a type for normative systems—to the basis is more interesting. She, however, mentions this possibility only in passing.

[19] The terms "explicit attitudes" and "implicit attitudes" were introduced as technical terms in TIL (probably in Duží 2004), and this terminology is in my view not ideal. It would be, perhaps, more suitable (though less concise) to distinguish between "attitudes to the construction of an object" and "attitudes to the constructed object" or "coarse-grained attitudes" and "fine-grained attitudes".

am only suggesting that Δ-TIL is not a very promising first step towards the formation of such a theory.[20]

References

CASTAÑEDA, H-N. (1981): The Paradoxes of Deontic Logic: The Simplest Solution to All of Them in One Fell Swoop. In: Hilpinen, R. (ed.): *New Studies in Deontic Logic*. Dordrecht: D. Reidel, 37-85.

CHISHOLM, R. M. (1963): Contrary-to-Duty Imperatives and Deontic Logic. *Analysis* 23, 33-36.

DUŽÍ, M. (2004): Intensional Logic and the Irreducible Contrast between de dicto and de re. *ProFil* 5, No. 1, 1-34, available at: http://profil.muni.cz/01_2004/duzi_de_dicto_de_re.pdf

DUŽÍ, M., JESPERSEN, B. & MATERNA, P. (2010): *Procedural Semantics for Hyperintensional Logic. Foundations and Applications of Transparent Intensional Logic*. Berlin: Springer.

GLAVANIČOVÁ, D. (2015): K analýze deontických modalít v Transparentnej intenzionálnej logike. [On the Analysis of Deontic Modalities in Transparent Intensional Logic.] *Organon F* 22, No. 2, 211-228.

GLAVANIČOVÁ, D. (2016): Δ-TIL and Normative Systems. *Organon F* 23, No. 2, 204-221.

HILPINEN, R. (ed.) (1971): *Deontic Logic: Introductory and Systematic Readings*. Dordrecht: D. Reidel.

JØRGENSEN, J. (1937/1938): Imperatives and Logic. *Erkenntnis* 7, No. 1, 288-296.

ROSS, A. (1941): Imperatives and Logic. *Theoria* 7, No. 1, 53-71.

RUSSELL, B. (1905): On Denoting. *Mind* 14, No. 56, 479-493.

SVOBODA, V. (2004): Rossův paradox a (deontická) logika. [Ross' Paradox and (Deontic) Logic] In: Zouhar, M. (ed.): *Používanie, interpretácia a význam jazykových výrazov*. Bratislava: Veda, 56-66.

SVOBODA, V. (2013): *Logika pro Pány, Otroky a Kibice. Filosofický průvodce světem deontické logiky*. [*Logic for Masters, Slaves and Kibitzers. A Philosophical Guide through the World of Deontic Logic*] Praha: Filosofia.

TICHÝ, P. (1988): *The Foundations of Frege's Logic*. Berlin – New York: de Gruyter.

[20] I am grateful to Pavel Materna for very useful discussions on topics related to the theme of this paper as well as to more general problems of TIL. I am also indebted to Czech Science Foundation for its support of my work in the area of deontic logic through funding of the project 13-20785S.

Organon F 23 (4) 2016: 551-560

Nissim Francez: *Proof-theoretic Semantics*
College Publications, London, 2015, xx+415 pages

During the second half of the twentieth century, most of logic bifurcated into model theory and proof theory. Model theory, as established by Tarski & Co., was considered as a matter of "semantics": it investigated the relationship between formal languages and the domains of entities about which the languages were supposed to be. Proof theory, on the other hand, was taken to be a matter of "syntax": not concerning what the formulas of formal languages are about, but about certain relations among them. Hence, when there appeared the term *proof-theoretic semantics* (PTS), it sounded quite paradoxical: how could there be a "syntactic semantics"?

The solution to this quandary lies, I am convinced, in the elucidation of the misleading role the term "syntax" has played within modern logic (and philosophy). Primarily, syntax is a theory of "well-formedness"—of the delimitation of the range of expressions which make up a given language. In this sense, syntax indeed has nothing to do with semantics and it would be futile to try to base a semantics on it. However, the term "syntax" has also been used to refer to inferences, derivations and proofs, and if considered in *this* sense, it is no longer so clear that it is unrelated to semantics. Indeed, the second half of the twentieth century also witnessed the rise of the so-called use-theories of meaning, at least some of which identified meaning of an expression specifically with the role conferred on it by the inferential rules governing its proper employment.[1]

Proof-theoretic semantics is also closely connected with the search for the proper semantics of intuitionistic logic. While classical logic has the natural truth-functional semantics, there was, for some time, no such canonical semantics for intuitionist logic. What was subsequently to become accepted as its adequate semantic account was its so-called BHK-interpretation (see Troelstra & van Dalen 1988): the idea that the semantics is based on the concept of proof. This idea is usually incorporated into PTS in such a way that the meaning of a sentence is considered as the set of all its proofs (or all its "canonical" proofs) and that logical connectives express ways to combine proofs of components into proofs of a compound. The term "proof-theoretic semantics" was introduced by Schroeder-Heister

[1] See Peregrin (2006a) for a discussion.

(1991) and its development is often associated with Prawitz (2006). Francez's book presents its elaboration not only for the formal languages of logic, but also for natural languages.

PTS built on this basis seems to pose two problems. Firstly, the association of the meaning of a statement with the set of its proofs appears to be epistemologically unrealistic: do we want to say that whoever understands a sentence is bound to know all the ways to prove it? And secondly, even if we accept this, the theory gives us meanings of sentences, but what about those of sub-sentential expressions?

Compare the situation with the well-known origin of model-theoretic semantics (MTS) for the languages of logic and its extension to natural language initiated by Montague (1974). Here, the starting point is the Frege's (1891) idea concerning explicating meanings of predicative expressions as functions from objects to truth values, which led to the standard truth-functional semantics for logical operators. This then led on to the general idea that the meanings of expressions of all other categories, save sentences and names, are functions built on the basis of the denotations of sentences (truth values) and names (objects). (For a very general formal language this was proposed, for the first time, by Church 1940.) Montague then provided his elegant semantic treatment of a fragment of English, which secured MTS a place on a philosophical pedestal.

But note an important feature of the Montagovian MTS: it was not extensional because the denotations of sentences were not their truth values; they were rather functions from possible worlds to truth values (or the sets of possible worlds that the functions characterize). Despite this, logical connectives could, in effect, retain their truth-functional denotations. Thus, even if we were to explicate the denotations of sentences as the sets of their proofs, it need not follow that if PTS were to follow in Montague's footsteps, the denotations of logical connectives would have to be something as monstrous as functions from pairs of sets of proofs to sets of proofs. And indeed it is the BHK-interpretation that indicates to what such denotations could be reduced: to methods of combining proofs that could be quite simple.

In view of this, we can put, and this is something Francez shows very explicitly in the book, the whole Frege-Church-Montague functional machinery into the services of PTS. On the level of propositional logic, we have just to assume that we have denotations of statements (the sets of their proofs) and derive the denotations of sentential operators as corresponding functions. Then we can try to reduce them to something simpler; in the case of conjunction it could, for example, be combining proofs of the two conjuncts into the proof of the conjunction, which could be nothing more complicated than putting the two proofs beside each other. Thus, the

meaning of "∧", for example, may be the function which maps two sets of proofs, P_1 and P_2, on the set of proofs that contains, for every proof D_1 of a formula A_1 that belongs to P_1 and every proof D_2 of a formula A_2 that belongs to P_2, the proof

$$\frac{D_1 \quad D_2}{A_1 \wedge A_2}$$

The situation is a little more complicated on the level of predicate logic, where Church made use of one more category the denotations of which were primitive, namely names (which, according to him, denote individuals—elements of a universe). Here Francez's approach diverges from the model-theoretic one: the category of expressions he chooses as the other primitive one are (individual) variables. A variable, according to him, denotes itself. As a result, a quantifier takes the denotation of a sentence (the set of its proofs) plus that of a variable (the variable itself) to the denotation of a quantified sentence. This, of course, presupposes that the category of sentences includes open formulas. In this way, the meaning of "∀", for example, is a function which maps a set of proofs P and a variable *v* on the set of proofs that contains, for every proof D of a formula A that belongs to P (and such that *v* is not in any premise or undischarged assumption of D), the proof

$$\frac{D}{\forall v A}$$

It turns out, however, that this version of PTS has a property that may be seen as problematic: the meanings of sentences, *viz*. the sets of their canonical proofs, turn out to be overly fine-grained. (For example, the meaning of $A_1 \wedge A_2$ comes out as different from that of $A_2 \wedge A_1$. This may make some sense for a natural language, but much less for a logical language in which the two sentences are provably equivalent and provably intersubstitutive w.r.t. logical equivalence.) Hence it would seem that what would fare better in this respect would be the identification of the meaning of a sentence with the set of *grounds* of the sentence: sets of all sets of formulas from which the formula is derivable. (Also we might think about including only *maximal* grounds, which would then be not so far from possible worlds, and PTS would come slightly closer to MTS.) It is a pity that Francez does not elaborate on this idea.

Francez uses the concept of ground also for the definition of proof-theoretical consequence: *A* is a consequence of *X* iff everything that is a ground for *X* is a ground for *A*. Again, it seems to me to be a pity that Francez does not tell us more about the proof-theoretical relation of consequence defined in this way. (Usually it

is noted that there is a gap between *derivability*, as a proof-theoretical matter, and *consequence*, which must be defined model-theoretically (see, e.g., Etchemendy 1990). Carnap (1934) tried to account for this gap in purely proof-theoretic terms (see Peregrin 2014, Ch. 7); and it would be nice to learn what ambitions Francez has using his definition.)

The second part of Francez's book applies PTS to natural language, thus creating an antipode to the Montagovian MTS. Some of the ideas already embodied into PTS for the languages of logic can be straightforwardly transferred to natural languages, but in some respects natural languages are different. In particular, we can treat at least some of the connectives on a par with their logical counterparts; but the way quantification operates in natural language is very different from the standard Fregean quantifiers embraced by logic.

Francez, nevertheless, approaches the situation analogously to that of the formal languages. He enriches the fragment of natural language by using "individual parameters", which play a role somewhat analogous to that played by variables in formal languages. (Thus the whole language he works with is comprised of natural language plus "open" sentences that can be assembled from elements of natural language and parameters.) And though the mechanism of quantification is different, Francez's way of coping with it proof-theoretically is quite similar: the proof of a general statement builds on the proof of the corresponding statement with an indeterminate individual parameter.

One of the crucial features of Montagovian formal semantics was that it accounted for intensional contexts, that by engaging possible worlds it surpassed the limits of extensional semantics (see Peregrin 2006b). The proof-theoretic account of Francez has no lesser ambitions: it, too, aspires to account for the intensionality of natural language. However, here the method differs greatly from the model-theoretic one. What does the work here is nothing like possible worlds. Francez introduces a new kind of individual parameters, which he calls *notional parameters*. These parameters have inferential properties different from ordinary individual parameters. For example, while *John finds a unicorn* is introduced on the basis of *John finds x* and *x is a unicorn* (hence it follows that there is something that is a unicorn), the grounds of the introduction of *John seeks a unicorn* are different: they are *John seeks n* and *n is being a unicorn* (where *n* is a notional parameter) and it has no existential import.

Francez's book is literally packed with information; it is, in fact, multiple books in one. It contains a concise introduction into Gentzenian proof theory; it contains an elaboration of the semantic ideas of both Gentzen and the BHK-people, taking them forward into an explicit theory of semantic values; and it contains—and this is the most original part—also an elaborated sketch of PTS for a fragment of natural

language, parallel to the celebrated MTS of Montague. Thus it shows that proof-theory is not syntax—at least not in any sense that would prevent it from conferring meanings on expressions. In this way it is, aside of presenting a wealth of new results, usable also as a handbook of structural proof theory. And given that College Publications, who published the book, do not overcharge their customers, buying it is a true deal!

Jaroslav Peregrin
jarda@peregrin.cz

References

CARNAP, R. (1934): *Logische Syntax der Sprache*. Vienna: Springer.

ETCHEMENDY, J. (1990): *The Concept of Logical Consequence*. Cambridge (Mass.): Harvard University Press.

FREGE, G. (1891): *Funktion und Begriff. (Ein Vortrag, gehalten in der Sitzung vom 9.1.1891 der Jenaischen Gesellschaft für Medizin und Naturwissenschaft)*. Jena.

CHURCH, A. (1940): A Formulation of the Simple Theory of Types. *Journal of Symbolic Logic* 5, 56-68.

MONTAGUE, R. (1974): *Formal Philosophy: Selected Papers of R. Montague*. New Haven: Yale University Press.

PEREGRIN, J. (2006a): Developing Sellars' Semantic Legacy: Meaning as a Role. In: Wolf, P. & Lance, M. N. (eds.): *The Self-Correcting Enterprise*. Amsterdam: Rodopi, 257-274.

PEREGRIN, J. (2006b): Extensional vs. Intensional Logic. In: Jacquette, D. (ed.): *Philosophy of Logic. Handbook of the Philosophy of Science 5*. Amsterdam: Elsevier, 831-860.

PEREGRIN, J. (2014): *Inferentialism: Why Rules Matter*. Basingstoke: Palgrave.

PRAWITZ, D. (2006): Meaning Approached via Proofs. *Synthese* 148, 507-524.

SCHROEDER-HEISTER, P. (1991): Uniform Proof-Theoretic Semantics for Logical Constants (Abstract). *Journal of Symbolic Logic* 56, 1142.

TROELSTRA, A. S. & VAN DALEN, D. (1988): *Constructivism in Mathematics: An Introduction. Vol. I*. Amsterdam: North-Holland.

Nikolay Milkov and Volker Peckhaus (eds.): *The Berlin Group and the Philosophy of Logical Empiricism*. *Boston Studies in the Philosophy and History of Science 273*. Springer, Dordrecht, 2013, x+332 pages

During the last few decades, historians and philosophers of science radically changed our perspective on logical empiricism in general, and on the Vienna Circle in particular. Though there are still some members of the Circle who did not get much attention (Victor Kraft, Richard von Mises, Felix Kaufmann, Josef Schächter etc.), we are in a quite good position to judge many of their efforts. On the other hand, our historical understanding of logical empiricism *in general* leaves something to be desired due to the circumstances that the so-called "Berlin Group" is underestimated in the literature.

The Berlin Group and the Philosophy of Logical Empiricism, edited by Nikolay Milkov and Volker Peckhaus, is meant to bring attention to the German wing of logical empiricism, thus doing justice to that forgotten projects and figures who had (in)directly an important influence on the philosophy of science in the United States after World War II. Among the most important members, one finds Hans Reichenbach, Kurt Grelling, Walter Dubislav, Paul Oppenheim, Olaf Helmer, Kurt Lewin, and Carl Gustav Hempel. The collection is devoted to their ideas and context in the European philosophy of science scene.

Part 1 is an introductory chapter composed of two papers: a longer article by Nikolay Milkov about the 'affinities and divergences' between the Vienna Circle and the Berlin Group. Though Milkov provides many important details and notions, his explanations are lacking sometimes, but I will come back to that later. The second paper is Nicholas Rescher's personal memories about his "interactions and collaborations with members of the Berlin Group" (p. 33). Rescher focuses on Helmer, Hempel, and Oppenheim (usually called as the 'H_2O philosophers'), discussing their role in the RAND corporation, and while it is always illuminating to read personal recollections about the less known sides of history, his paper is just five pages long, so one can get only a slight hint about the historical events.

The second part of the collection aims to explore the historical and philosophical context of the Berlin Group. Helmut Pulte describes in a lucid fashion those nineteenth-century roots of the Group which goes back to Jakob Friedrich Fries, an important critic of Kant. The ideas of Fries were continued by E. F. Apelt (referred to by Reichenbach in his dissertation) and later by Leonard Nelson who founded the so-called New Friesian School, inspiring such scholars as Grelling (who published with Nelson), Dubislav and Reichenbach. The second paper of this section

is
Jeremy Heis' work on the connections between Ernst Cassirer, Lewin, and Reichenbach—a topic which surfaces again in Milkov's paper who will argue that Hempel "more closely followed Lewin's Cassirer-inspired project than he did that of Hans Reichenbach" (p. 298).

Given that the Berlin Group is associated with individual figures rather than commonly shared theses, the collection treats the main figures in separate chapters. Part 3 considers Reichenbach's life and work. Flavia Padovani touches upon Reichenbach's time in the so-called Jugendbewegung (German Youth Movement), the famous Erlangen-conference, and his work on the radio. The paper, otherwise, is devoted to the conceptions of time and 'genidentity' as it was worked out by Lewin and Reichenbach. Michael Stöltzner's paper discusses the question of quantum mechanics and indeterminism. He achieves some important and strange conclusions; both Reichenbach and (based on his claims) the literature stated that the distinguishing feature of the Berlin Group (contrasting it with the Vienna Circle) is their member's continuous contact with the actual works of scientists. Nonetheless, Stöltzner shows that this idea requires some qualification since Reichenbach "did not involve himself into the details of the physical discussions, but pursued a genuinely philosophical agenda" (p. 146). Finally, Andreas Kamlah presents Reichenbach's involvement in the *Jugendbewegung*, which was a reform movement, originated from the early years of the twentieth century. He was an important figure in the so-called *Freistudenten* [Free Students] movement in Berlin—due to its explicit socialist leanings Reichenbach had different times later in pursuing academic jobs. Kamlah argues quite convincingly that Reichenbach's involvement (and leading role) in the voluntarist, pluralist, and tolerant movement had an important effect also on his philosophy (as it was the case actually with Carnap too).

Part 4 is devoted entirely to Dubislav—we got to know his logical works (Christian Thiel), his ideas on transcendental arguments (Temilo van Zantwijk), and his relation to Bernard Bolzano (Anita Kasabova). Two things emerge from these articles: Dubislav was involved in many up-to-date projects, acknowledged by many important figures of the history of logic and philosophy in the twentieth century. The other is that the fact that Dubislav is quite forgotten among philosophers might be due to the fact that, for example, his logical project was partly a failure since his work did not provide a decision procedure for classical monadic quantificational logic, or more precisely, "it yields only a sufficient criterion of validity, and not a necessary one" (p. 187).

Kurt Grelling (discussed in Part 5), though played in important role in the Berlin Group, he was always a third member behind Reichenbach and Dubislav as it is argued in Volker Peckhaus' contribution. Peckhaus mentions Grelling's ideas on

formal ontology (which is also the subject of Arkadiusz Chrudzimski's paper), his Russellian leanings (he translated four important books of Russell), and his involvement in psychology. Nevertheless, what emerges in these sections for sure is that Grelling was "a valuable collaborator" (p. 241).

The final, sixth part of the collection contains one paper on Oppenheim (Paul Ziche with Thomas Müller), and two on Hempel (Nikolay Milkov and Erich H. Reck). Ziche and Müller take Oppenheim as a co-author of many important philosophers of science and claims that he was "the greatest philosophical co-author of the twentieth century" (p. 265). On the other hand, and more importantly, they also view Oppenheim as an individual scholar who was interested in the order of the sciences, holding some unique position among logical empiricists. The Hempel-papers consider him in relation to others: Milkov argues against Michael Friedman's thesis that Hempel was influenced more by Carnap than by any Berliner. Finally, Reck takes the late Carnap's ideal of explication (on which we have now a flourishing secondary literature, partly due to Reck), and compares it to Hempel's ideas on the Covering Law Model of explanation; on the base of this Reck is able "to get clearer about why exactly [Hempel's] texts were so influential and, more basically, what their philosophical significance is" (p. 312).

Finally, a few words need to be said about the general narrative of the volume. In his introduction, Milkov claims that there is a certain asymmetry in the reception of the Berlin Group and the Vienna Circle in favor of the latter, though in some cases the Berliners have a priority claim. After that, he tries to show the reasons of the general neglect of Reichenbach's group. He discusses one theoretical and three external factors which purported to explain the asymmetry in the reception-history. The theoretical factor is that what "made the Vienna Circle's activities the more visible was Ludwig Wittgenstein's philosophy of language" (p. 5), and that when the philosophical debates of the Circle got public it "called attention to themselves in ways not seen in the Berlin Group" (p. 5). Contrary to this, the Berlin Group was occupied with dialogues of working scientist, keeping their eyes on the concrete scientific developments instead of inner-type philosophical debates.

The problem is that which Vienna Circle does Milkov talk about? The Vienna Circle members indeed shared many commitments, mainly connected to Wittgenstein, and debated only about philosophical matters (external to the actual scientific problems) *according to the received view*. Thanks to such volumes as Milkov's and Peckhaus', however, we are now aware of that fact that the received view was false—or at least misleading and oversimplified. So Milkov's story could be true in the received view, but false in the rehabilitated picture—the question is, which story was the story about the Circle in the 1920s and 1930s. Anyway, another question

emerges—namely that how did actual scientists respond to the claims and theories of the Berlin Group? Were they taken seriously? We do not get an answer to that.

The external factors behind the asymmetry are these: (i) the manifesto's radical program made the Circle recognized worldwide, (ii) the members of the Berlin Group had quite a peculiar and tragic careers without becoming as mainstream and known scholars as Schlick or Neurath, (iii) "[w]hereas Hitler came to power in Berlin in January 1933, he did not force Austria into the German Reich for more than 5 years (in March 1938)" (p. 7).

Regarding (i) one might point out that (a) the manifesto raised important and deep controversies inside the Circle; (b) the papers of Eino Kaila, Åke Petzäll and Feigl/Bloomberg did not originate from the Circle's manifesto—Kaila went to the Circle's meetings in 1929 but knew their program much earlier, and Feigl was one of those students of Schlick who persuaded him to gather a group of scholars around him to discuss philosophical problems. On the other hand, Reichenbach already published many important books and papers (also popular ones) already in the early 1920s. It is a further question (perhaps connected to the Austrian and German philosophical scene) why his efforts were less successful in the forming periods.

The second point is also problematic: it is true that it was only Reichenbach who "fully developed his philosophical program" (p. 6)—while Grelling organized some discussion groups even in the internment camp in South (Vichy) France in 1941, he died in Auschwitz one year later, still in his productive years. Dubislav, after a short imprisonment in Berlin, went to Prague where "he killed in jealousy first his girl-friend then himself on 16 September 1937" (p. 237). Many of the members of the Vienna Circle indeed had a much fruitful and longer career. On the other hand, Hahn died already in 1934, and Schlick was murdered in 1936; Zilsel committed suicide in 1944. While all the members of the Berlin Group worked in Berlin between 1926 and 1933 (when Reichenbach immigrated to Turkey), the Vienna Circle's most important members left Vienna quite early: Philipp Frank took over Einstein's position in Prague in 1912, Carnap got a positions there too in 1931, and after 1934 Neurath went to Hague. Somehow the Circle still managed to work together and built up the narrative of a successful discussion group.

Finally, though Hitler came to power in Berlin in 1933, the national socialist's revolution in Vienna caused the dissolution of the Ernst Mach Society already in 1934 and made it impossible (even earlier) for many members to get a job at the University.

Though Milkov did a great job to explore the reasons behind the asymmetry between Berlin and Vienna, some more details and inquiries are required since in

themselves the above-mentioned reasons are insufficient to explain the historical phenomena. Even if the book could be considered to be only a starting point for the later philosophical debates, *The Berlin Group and the Philosophy of Logical Empiricism* is an important collection of fine-grained and thought-provoking essays; they show some possible paths from Vienna to Berlin and back.[1]

Adam Tamas Tuboly
tuboly.adam@btk.mta.hu

[1] Supported by the ÚNKP-16-4-II New National Excellence Program of the Ministry of Human Capacities.

Organon F 23 (4) 2016: 561-565

Issues on the (Im)Possible IV
August 30-31 2016, Bratislava

In August 30-31, 2016, the Institute of Philosophy of Slovak Academy of Sciences, metaphysics.sk research group and Slovak Philosophical Association hosted the fourth instalment of Modal Metaphysics: Issues on the (Im)Possible conference. As in the previous years, this year's session brought together researchers and graduate students from all around the world to discuss contemporary moves in analytic metaphysics in general, modal logic, metaphysics and epistemology in particular. A specific feature of this conference is that accepted and presented papers were assigned commentators who read the papers in advance and provided critical comments on them. Invited speakers were Gideon Rosen (Princeton University) and Richard Woodward (University of Hamburg).

The conference commenced with two parallel sessions with talks given by Robert Michels (University of Geneva) (commented by Nathan Wildman) and Zsófia Zvolenszky. In his "Is 'Metaphysical Necessity' Ambiguous?", Michels argued that the correct conceivability-based definition of metaphysical possibility is equivalent to a sort of essentialism. The goal of Zvolenszky's talk entitled "Fictional Names, Rigidity, and the Inverse-Sinatra Principle" (commented by Vladislav Terekhovich) was to answer the question as 'how exactly Kripkean views on proper name reference are supposed to extend to the fictional names like "Holmes"?'. The next pair of talks consisted of "Counterfactuals as Property Relations" by Meagan Phillips (Northern Illinois University) (commented by Lorenzo Azzano) and "Explaining Essence and Modality" by Jakob Schieder (Humboldt Universität zu Berlin) (commented by Antonella Mallozzi). In the former, Phillips explored the possibility of an analysis of counterfactuals that she extrapolated from Jubien's analysis of modality. In the latter, Schieder argued that essence can be fruitfully explained in terms of what it takes to be the referent of a representation. Zuzanna Gnatek's (Trinity College Dublin) "Object Dependency in Timothy Williamson's Deductive Argument for Necessitism" (commented by Alexander Roberts) focused on one premise of necessitists' argument which states that necessarily, if the proposition that Socrates is nothing is something then Socrates is something and "Exploring the Contingent Fundamentality Thesis" by Nathan Wildman (University of Hamburg) (featuring Joachim Horvath as a commentator) responded to four objections to the contingent fundamentality thesis, as well as examined how it interacts with

various claims about the modal existential status of the fundamentalia themselves. Alexander Roberts (University of Oxford) in "Modal Expansionism" proposed a novel conception of metaphysical modality, according to which it fails to be the maximal objective modality. Alex Steinberg's (University of Zurich) "Saving Strict Adequacy" argued against the idea that Lewis's scheme is not strictly adequate. To do so, he provided new translations for the allegedly problematic extraordinary modal sentences. The last parallel session of the first day ended up with "Aristotle's Modal Ontology – Overcoming Potentiality-Actuality Reading" by Kei Chiba (Hokkaido University) and "More Than Impossible: Negative and Complex Probabilities and Their Interpretation" given by Vasil Penchev (Bulgarian Academy of Sciences). While the purpose of the latter was to introduce negative and complex probability relevant to special and general relativity, the former investigated Aristotle's modal ontology by distinguishing and relating each of his three modal notions the power-ability (dunamis), the completeness (entelecheia) and the at-work-ness (energeia). The last talk of the day—the keynote address—was given by Gideon Rosen (Princeton University). Entitled "Modality in the Metaphysics of Ethics", the talk argued that pure moral principles are best to be understood as modalized generalizations of the form 'It is normatively necessary that whatever is PHI is F' (where PHI is non-normative and F is normative), though not every such fact is a principle; and that given plausible assumptions, most such principles are metaphysically contingent.

The second day of the conference started with Alex Kaiserman's (University of Oxford) "A Real Definition of Token Physicalism" (commented by Jakob Schieder) and "Dispositional Arrays" by Lorenzo Azzano (Scuola Normale Superiore) followed by Meagan Phillips's comments. Kaiserman suggested a different approach to token physicalism, one which appeals to the essentialist concept of 'real definition'. On this approach, token physicalism turns out to be a much more substantive and interesting view than previously thought. The core idea behind Azzano's talk was that opposition between powers approach and possible worlds approach is unwarranted. Namely, he presented a power-based ontology of possible worlds, which in turn offers a power-based applied possible worlds semantics for modal discourse. In her "Conceivability, Possibility, and the Inconsistent Triad. The Kripkean Challenge to Modal Rationalism" (commented by Alex Kaiserman), Antonella Mallozzi (CUNY – The Graduate Center) showed that Chalmers's Modal Rationalism involves an inconsistent triad composed of (1) Two-Dimensionalism, (2) Modal Monism, and (3) a Kripkean Metaphysics. As she demonstrated, only two of those can be true at a time, while the project needs all of them in order to succeed. Dirk Franken's (University of Marburg) "In Defence of Modal Monism" (commented by Zuzanna Gnatek) made a claim that the Modal Monist is

in the better position to block her opponents master argument than the Modal Dualist. "Possible Worlds and Substances" (commented by Vasil Penchev) by Vladislav Terekhovich (Saint-Petersburg State University) concluded that the activity of substances changes from the possible modality of being to the actual modality of being in a form of existence of the worlds. Joachim Horvath (University of Cologne) presented "Philosophical Analysis: The Concept Grounding View" (commented by Frances Heather Fairbairn) in which he proposed the concept grounding view as a promising account which meets the challenge that the success conditions of philosophical analysis are unclear. Cristina Nencha (Northwest Italy Consortium) asked a question "Was David Lewis a Necessitist?", investigating what she took to be Williamson's main reason for saying that Lewis is a necessitist (followed by David Mark Kovacs's comments). Frances Heather Fairbairn (Cornell University) re-opened "The Problem of Advanced Modalizing" (commented by Robert Michels) and suggested a solution that leaves genuine modal realism, its translation schema, and its ontology intact. The last talk in a parallel session was given by Philipp Berghofer (University of Graz). In his "Unknowable Truths and Limits of Knowledge: What Conclusions Can We Draw from Fitch's Paradox of Knowability?" (commented by Robin Neiman), Berghofer discussed the impact of Fitch's argument on the question of whether there are limits to (human) knowledge and showed that there is no impact at all. The conference ended up with the second keynote address entitled "The Questions of Ontology". In it, Richard Woodward aimed to defend a broadly Quinean picture of ontology by arguing that both Fine's criticisms of Quine, and the conception of ontological inquiry that Fine subsequently develops, are problematic.

Issues on the (Im)Possible series keeps attracting researchers from all around the world(s). And although it has been a while we started organising the event, we still feel an optimism to continue in this activity. It is partly due to the fact that problems of modality are not limited to a narrowly conceived analytic metaphysics. They occur also in other branches of philosophy. Partly, it is the very community which makes the conference both intellectually intensive, yet essentially relaxed.

(Video of *Gideon Rosen's talk* is available at: http://archive.tp.cvtisr.sk?9158683.)

Martin Vacek
martinvacekphilosophy@gmail.com

5th Slovak Philosophical Congress Report

The 5th Philosophical Congress "(Meta)Philosophy – Practice" took place in Bratislava on October 21-23, 2015. It was 20 years ago that the 1st Slovak philosophical congress, also held in Bratislava, commenced the tradition of Slovak philosophical congresses. As its name suggests, the congress addressed the questions of self-understanding in philosophy, the "practical" and "theoretical" philosophy, the relationship between philosophy and practice, but also the current issues of contemporary man and society emerging in various spheres of individual and social life.

The congress was organized by Slovak Philosophical Association at the Slovak Academy of Sciences and the Institute of Philosophy at Slovak Academy of Sciences, under the auspices of the president of the Slovak Republic, Andrej Kiska. The event was generously supported by Holiday Inn Hotel in Bratislava, Bratislava Tourist Board, Slovak National Museum – Ľudovít Štúr Museum in Modra and REA printing company, Nitra.

The congress was opened by the *President* of the Slovak Philosophical Association at Slovak Academy of Sciences (SPA, SAS), Andrea Javorská, followed by the ceremonial speeches given by the director of the Institute of Philosophy, Slovak Academy of Sciences, Tibor Pichler, and the President of SPA, SAS. Among other things, Tibor Pichler pointed out to new trends in the Slovak philosophical thought.

The ceremonial speeches were followed by plenary lectures by the following participants: Vladimir Vladimirovič Varava (*Peculiarities of Russian Philosophical Language as a Language of Meta-Literary Discourse*) from the University of Voronezh, Martin Kusch (*Scepticism and Relativism*) from the University of Vienna, Emil Višňovský (*Richard Rorty and the Mirror of Philosophy*) from the Faculty of Arts, Comenius University, Marián Zouhar (*Philosophy and Conceptual Space*) from the Faculty of Arts, Comenius University, and Miroslav Marcelli (*Two Kinds of Philosophical Thinking*) from the Faculty of Arts, Comenius University in Bratislava. The lectures presented important and current issues in philosophy and reflected the heterogeneity of both problems and approaches. They triggered interesting and valuable discussions.

After the plenary lectures, the conference hosted five individual sessions in three days. The sessions were dedicated both, to the continental and analytic philosophical traditions. The talks focused on the history of philosophy, different metaphysical, epistemological and ethical issues, as well as on the perspectives coming from other scientific fields. It is evident that the variety of topics reflected (not

only) a variety of interests, contexts and different philosophical approaches along with an openness for inter- or trans-disciplinary approach.

On the first day, Andrea Javorská, on behalf of the Organising Committee of the congress afforded the honorary membership to two Slovak philosophers, namely Miroslav Marcelli (Department of Philosophy and History of Philosophy, Faculty of Arts, Comenius University in Bratislava) and Jozef Sivák (Institute of Philosophy, Slovak Academy of Sciences). It was a gesture of gratitude for their philosophical work and a long-standing cooperation with the Slovak Philosophical Association. The first day's evening reception was also attended by the Bratislava Tourist Board Executive Director, A. Mikulová.

The congress also commemorated the 200th anniversary of Ľudovít Štúr's birth, an important Slovak intellectual, by a panel discussion called "*Ľudovít Štúr in an Intersection of the Past and the Present"* which took place during the second day. The following specialists took part on the discussion: Tibor Pichler, Beáta Mihalkovičová (Slovak National Museum – Ľudovít Štúr Museum in Modra), a historian of culture and art and publicist Viliam Jablonický (Slovak P.E.N. Centre member), and Vasil Gluchman (Faculty of Philosophy, University of Prešov), who also chaired the discussion. They emphasized that Štúr's biography and intellectual work are still underappreciated.

The 5th philosophical congress hosted around 80 participants, a considerable number of them coming from countries such as Poland, the Czech Republic, Ukraine, Russia and Austria. The congress brought a professional benefit in for the participants and confirmed the view that the tradition of organising philosophical congresses in Slovakia is a worthwhile project. We believe that this event embodied a unique and precious opportunity to experience philosophy, promote creative dialogue and promote academic contacts.

Sabína Gáliková Tolnaiová
sgalikovatolnaiova@gmail.com

Institute of Philosophy, Slovak Academy of Sciences

Modal Metaphysics: Issues on the (Im)Possible V

August 17-18, 2017
(Bratislava, Slovakia)

Keynote speakers

PHILLIP **BRICKER** (University of Massachusetts, Amherst)
SEAHWA **KIM** (Ewha Womans University, Seoul)

We invite submissions for a 30 minute presentation followed by 10 minute comments and a 15 minute discussion. Areas of interest might include any aspect of analytic metaphysics, epistemology and logic of modality.

A paper of approximately 3000 words should be prepared for blind review and include a cover page with the full name, title, institution and contact information.

Papers can be submitted in pdf or doc(x) and should be sent to **modalmetaphysics@gmail.com.**

Deadline for submission: **March 15, 2017**
Notification of acceptance: **April 30, 2017**

If you wish to submit a paper, comment on an accepted paper or would need any further details, contact Martin Vacek (martinvacekphilosophy@gmail.com) or visit the conference website **www.metaphysics.sk**.

Contents

ARTICLES

DISCUSSIONS

BOOK REVIEWS

REPORTS