Organon F 23 (1) 2016

Contents

ARTICLES

DISCUSSIONS

BOOK REVIEWS

REPORTS

Organon F 23 (1) 2016: 2-20

Logical Expressivist's Logical Constants

PAVEL ARAZIM

Department of Logic. Faculty of Arts. Charles University in Prague
Palachovo náměstí 2. 116 38 Praha 1. Czech Republic
pavel.arazim@centrum.cz

RECEIVED: 01-12-2014 • ACCEPTED: 28-04-2015

ABSTRACT: I would like to show that the problem of logical constants can be helped by treating the problem of relationship between logic and human reasoning. Thus I will present some parallels between the respective dilemmas and show that choice of a proof-theoretic answer in one case induces an expressivist choice in the other and the other way round, as well. This does not mean that other options are closed, though the two selected ones are thus given a new plausibility. Furthermore, the proof-theoretical demarcations of logical constants can provide missing details into the expressivist story, as they say which constants and why can actually perform the expressivist job.

KEYWORDS: Demarcation – logical constants – logical expressivism – model theory – proof-theory.

1. The problem of logical constants

Logic, presumably as every other discipline, should have its own vocabulary, i.e. there should be some words which belong specifically to the purview of logic. One can distinguish between meta-vocabulary and object-vocabulary. Let us illustrate the distinction using the example of zoology. Members of its meta-vocabulary are, e.g. *species, kind, family* etc. Its object-vocabulary, then, includes such words as animal, dog or dolphin. Now logic obviously also has its meta-vocabulary, including, among others, *conse-*

quence, *premise* or *contradiction*. But we are interested here in the object-vocabulary, which plausibly includes the classical quantifiers, the signs for classical connectives and perhaps other signs as well. These are the ones which make the logical form of a meaningful sentence. Now how can we decide, which linguistic items to count as logical constants in this sense?

Obviously, no logician would think of the word *dog* as of something within the purview of logic. Whether the sentence *Every dog is a mammal* is true is something vindicated by a biologist (as far as it makes sense to verify such a sentence instead of accepting something like *Dog is a kind of mammal* as a partial definition). The same holds for the fact that the sentence *Some dogs can fly* is false. The truth value would change in both cases, if we substituted the word *dog* by *bird*, which would mean to break the laws of biological discourse. On the other hand the sentence *Every dog is or is not a mammal* can be judged true by a logician. And logicians would agree that the words *or* and *not* belong, unlike *dog* or *bird*, within the purview of logic. Thus also substituting other ones for them, such as *and*, would mean breaking the laws of logic.

It can thus be said that logic is, as every other discipline, distinguished by its specific vocabulary (by which I mean, once again, the object-vocabulary). This does not necessarily mean that logic is concerned specifically with just linguistic phenomena, just as zoology's having a specific object-vocabulary does not mean it is concerned only with linguistic matters. In case of logic, unlike in that of zoology, this is of course a much more attractive way to see its subject matter.

Anyway, the boundaries of the object-vocabulary are hardly clear. Does the modal operator on sentences *necessarily* belong to the logical vocabulary? It is disputable, whether a vocabulary of any discipline is quite definite. Intuitively, it seems (especially to someone acquainted with Quine 1951) that it is actually not. Yet clearly some specification must be at hand in every case, otherwise we would have no idea about the given discipline. Actually, there can be more non-equivalent specifications and none has to make a clear-cut distinction, but they still have to serve to elucidate the character of the given discipline.

Now, should there be something special about the case of logic, should, say, its boundaries be specified more definitely, getting more close to clear-cut boundary than in the case of other disciplines? The intuition probably is that we should be somewhat more demanding about the specificity of criteria

of belonging to its vocabulary, i.e. criteria of logical constanthood.[1] Logic should, after all, be in one way or other, constitutive of our rationality and thus if anything is supposed to be definite, then it is logic. It appears that when one is somehow incompetent in logic, by which I do not mean that he lacks the academic skills and knowledge, but that (s)he is unable to master the behaviour of words such as *and* in every-day life, then (s)he cannot really understand anything else, indeed is incapable of being rational. Thus definiteness should be in general considered as a virtue of any suggested demarcation of logical constants.

2. The most important approaches

How can we try to demarcate the logical constants, then? In MacFarlane's (2009) helpful summary of the history of the present issue, we are offered quite a lot of possible approaches, yet only two of them not shortbreathed and actually alive (that is, besides the obvious possibility of scepticism, which, as I have tried to motivate in the introduction, should be at least tamed, if not refuted). One approach can be called a model-theoretical, the other one proof-theoretical. I will occasionally speak about the model-theoretic and proof-theoretic demarcation. Yet in both cases it would be more accurate to speak rather of a family of demarcations, as there are more possibilities, how to actually demarcate logical constants both model-theoretically and proof-theoretically. Yet no confusion should arise, as I will try to make clear in the subsequent sections. Let us see the two main approaches in some detail.

3. The model-theoretic demarcation

The problem of logical constants is rather significantly older than the two proposed solutions. Together with an ancestor of what I call here the model theoretical demarcation, it was identified by Bolzano. He was nevertheless generally rather sceptical about the possibility to give a non-arbitrary crite-

[1] This expression means that the members of logical object-vocabulary are treated by logic according to their specific meaning, unlike the members of the vocabularies of other discipline's, from the meaning of which logic abstracts, treats them rather as variables than as constants.

rion of logical constancy. For his contribution and further development of the issue till the time I cover in this article, see the trilogy of articles by Ladislav Koreň – Koreň (2014a; 2014b; 2014c).

The model-theoretical account was first proposed by the very founder of model theory, namely Alfred Tarski. In Tarski (1986), he proposes to demarcate the logical vocabulary, or rather the notions which can be denoted by logical vocabulary, in a way which generalizes the analogous demarcations of geometrical concepts. In what follows, a transformation is a bijection of some domain onto itself.

As the concepts of Euclidian geometry are invariant under similarity transformations of the universe (for example a triangle gets mapped on a similar triangle, though perhaps proportionally smaller or bigger), the affine geometry under affine transformations (so that a triangle gets mapped on a triangle, though possibly not a similar one) and topology under continuity transformation (thus the triangle will be mapped on something which might not be a triangle but still is a continuous figure, the transformation does not tear it apart), we can say that the respective geometries deal with concepts of increasing generality. Now, according to Tarski, there are also concepts which remain preserved under all the transformation of the universe onto itself and these are exactly the logical notions. For instance, take the existential quantifier. In model theory it can be seen as a second-order predicate, i.e. a predicate on sets. If we consider any bijection of the universe onto itself and with it the induced bijection of higher-order objects, such as the sets of primitive objects (i.e. members of the original domain), then the existential quantifier holds of any set if and only if it holds of its image under such a bijection.

This original idea had to be modified because otherwise some unwelcome concepts would also have to be counted as logical. In more recent model-theoretical demarcations, the logical notions are defined as the ones which are invariant not just under a bijection of a given universe onto itself but rather under bijections between different universes, i.e. domains of models (structures). This version of the conception, as well as its historical development, is captured by Gila Sher in Sher (1991).

In her book the reader may also find in more detail which elements of language are thus identified as logical constants. Let us just briefly mention that according to this approach logic is a rather broad discipline. For instance all the quantifiers which speak about cardinality of sets are logical. For ex-

ample, we can think of a quantifier $\aleph_{131}$, which asserts that there are $\aleph_{131}$ things (satisfying the given formula). And there are many more quantifiers of quite different kinds which get counted as logical constants by these lights. Actually, even though the logic developed by Sher is actually first-order (as these quantifiers speak only of objects, not of sets of objects) we get a system, which is practically as strong as full second order logic (details about this relationship with the second-order logic can be found in Bonnay 2008; the salient properties of second-order logic are explained in Shapiro 1991). Thus logic incorporates most of the set-theory and when one accepts this demarcation, it even makes sense to say that logicism gets verified, i.e. mathematics is proven to be a part of logic.[2]

In closing this section, it is good to remark that in this orthodox form, the model-theoretic approach leads to revisionism, on the one hand, and leaves some questions about logical constants unanswered, on the other. As regards revisionism, it is a problematic charge, as it is doubtful whether we can speak of a list of logical constants the logicians generally agree on. Though, as I already pointed out, there is something like a mainstream. Perhaps the constants of classical first-order logic are the most standard list. Now, here we see that this demarcation suggest a substantial broadening of it. As regards the unanswered questions, it is clear, that modalities and in general the constants of intensional logics are not touched by this approach, as it is focused mainly on first-order quantifiers. A partial answer might, nevertheless, be still better than no answer.

4. Proof-theoretical demarcation(s)

Unlike the demarcations formulated using model-theory, the ones formulated using proof-theory form a very heterogeneous class. Actually, there are some potentially important differences in the model theoretical camp, as well, but they are less significant and less motivated (see Bonnay 2008).

But before getting into the differences among the proponents of the proof-theoretical approaches, let us first see the general idea which unites them.

[2] This is because set theory becomes a part of logic and most substantial parts of mathematics can be reconstructed in it; but it would of course need more considerations.

This time it is Gerhard Gentzen who can be considered as the originator. While presenting his sequent calculus and his calculus of natural deduction (or, better, the corresponding forms of logical calculi) in Gentzen (1935), he said that the rules for the constants function as *sozusagen Definitionen*. They express what e.g. conjunction or existential quantifier are, or at least an important part thereof.

If we concentrate now on the sequent form of these rules for, say, constants of classical or intuitionistic logic, we see that they are fully schematic. The logical constants are the only concrete linguistic items, the accompanying formulas are just placeholders which can be filled in by any sentences. This fact led authors, such as Ian Hacking or Kosta Došen, to propose that exactly the items which can be characterized in such a calculus in this schematic manner should be regarded as logical constants. This corresponds to the fact that logic is plausibly supposed to be topic-neutral.

Let us see two examples of proof-theoretic characterization. I will show a characterization of conjunction and of existential quantifier in the form presented in Došen (1994). This approach is little bit unorthodox, as it countenances the double-line rules, the double-line expressing a mutual derivability (that is we can derive also upwards and not just downwards as usual).

$$\frac{\Gamma \Rightarrow \Delta, A \qquad \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \wedge B}$$

$$\frac{\Gamma, A \Rightarrow \Delta}{\Gamma, \exists x A \Rightarrow \Delta}$$

In the case of existential quantifier, there is a special requirement on the side formulae, namely that they do not contain any free occurrence of the variable x. Thus in the case of conjunction the rule shows that its inferential properties do not anyhow depend on a specific context, i.e. on what is being talked about. Just anything can occur among the formulae of Γ and Δ. This connective is thus independent of a specific makeup of the context. Imagine a similar rule for the word *dog*. I do not want to propose it in any specific form, but besides the word *dog*, it would certainly have to include other specific words, perhaps *mammal* and others. In this sense the rule would not be truly schematic. And this corresponds to the intuition that the logical object-

vocabulary should be topic-neutral or universally applicable in rational discourse. Put otherwise, the rules governing the logical constants are special in the sense that they are independent of context – the rules apply no matter what a given discourse is about.

In the case of the rule of existential quantifier, we can say that it is not fully schematic, because of the proviso regarding the variable. This leads Ian Hacking to assert that the existential (as well as the general) quantifier is not completely schematic and thus topic-neutral, which makes it somewhat less logical than the connectives. While I am sympathetic to this stance, I would like to point out that the proviso is obviously very week and only negative, requiring something not to occur in the accompanying formulae.

Nevertheless, the case of existential (and in fact also the universal) quantifier shows that the notion of the inference rules being schematic might be in need of more specification. Perhaps being schematic is not a yes or no matter, but rather a matter of degree. Thus relaxing the requirements we get the classical quantifiers counted as logical constants and relaxing even more might lead to accept some modalities. Yet these further relaxations to allow them would be rather significant. The details might be found again in Hacking (1979).

Besides the possibilities to understand the schematic character of rules in different ways, there are more ways how the model-theoretical demarcators can legitimately differ. First of all, it is open which structural rules one accepts or whether one allows for multiple conclusions. And such choices do affect what one demarcates, as for example allowing multiple conclusions leads together with the other rules and relatively strict understanding of what makes a rule schematic to classical logic, while banning them, i.e. allowing only single conclusion, leads to intuitionistic logic.

A variation on this theme is the possibility to require the constants to preserve different structural rules. Thus Hacking (1979) demands that after an introduction of a new logical constant the structural rules be provable, i.e. they do not have to be stipulated for more complex formulae so that the new complexity does not conflict with the previously valid structural rules. Furthermore, according to Hacking, the rules also have to be conservative. On the other hand, Došen (1994) leaves open the possibility that introducing a new constant may, for instance, not be conservative (implication can make left weakening valid). These requirements on the rules are proposals to capture the notion of harmony of the rules for the logical constants, which was

introduced in Dummett (1973) and further discussed in Dummett (1991). They prevent, among other things, such constants as Prior's *tonk* to become a part of logic.[3]

These technicalities might be substantial not just for a logician but also for a philosopher but let us put them aside, as we want to discuss rather the respective merits of the proof-theoretical approach in general, as compared with the model-theoretical one. For the present purposes another dispute internal to the proof-theoretical demarcators is more relevant. Some authors claim that the rules of the respective calculi are indeed definitions in the full-blown sense. Others are more modest, saying that these rules merely characterize or somehow analyse these expressions, possibly leaving some features of their meaning aside.

The attempts to fully define the constants by proof-theoretical means do indeed have their significant problems. The proof-theoretical definitions cannot actually distinguish the more desirable logical constants from some exotic and hardly acceptable ones. For example, Harold Hodes is forced to speak about obscure properties of some rules, such as their being primitively compelling (cf. Hodes 2004). Hacking and Došen adduce, on the other hand, plausible arguments for not considering the rules as definitions in the full sense. Hacking emphasizes that in order to understand for instance the rules for conjunction one has to understand the concept already (as we speak about the first conjunct appearing in its appropriate place AND the second conjunct as well). Došen shows that the rules lack some characteristics of definition. Most importantly they do not allow eliminating the constants in every context.

But as long as our concern is with the demarcation of logical constants it is not so important whether the rules fully define the logical concepts. It is enough that they pin down some of their features which distinguish them from extra-logical ones. Thus Gentzen's remarks about *sozusagen Definitionen* serve rather just as a motivation for this approach, not necessarily as a binding programme. The proof-theoretical demarcators thus do not have to embrace the thesis that the semantics of the constants has to be provided proof-theoretically.

---

[3] The tonk connective, introduced in Prior (1960) is defined by introduction and elimination rules in a natural-deduction calculus. Given a formula A, the rules say you can infer AtonkB for any other B and subsequently you can infer B.

To conclude this section, let us remark that the proof-theoretical approaches are generally not as revisionist as the model-theoretical ones. Furthermore, they also allow discussing the modalities and the language of intensional logics in general (though we did not focus on this issue here). Yet they do not give a fully definitive answer in these cases. But perhaps this is not only a bad thing, as it reflects some natural fuzziness of the notion of a logical constant. On the other hand, when we settle on a notion of schematicity of rule, the answer is definite.

5. Connections with human reasoning

Now I would like to discuss some broader implications of the presented demarcations, especially of the proof-theoretic ones. It is clear that both the proof-theoretic and the model-theoretic account can claim some degree of plausibility. But they give quite different results,[4] so we have to inquire into the differences of the disciplines which they describe and ask which of them is more adequately described as logic. The difference is even more striking if we consider that the two approaches point to different answers to the questions about the truth of logicism.

Logic obviously has to be connected to human reasoning in some way or another. The basic question is how much it can be independent of it. I would like to distinguish three basic approaches. My list is not supposed to be exhaustive, though I believe that the most relevant accounts are basically variations of them.

The first approach is psychologism. According to psychologism, logic is a discipline which studies how we human beings actually think or perhaps argue. (There are several ways of how to specify it.) Logic is thus a descriptive discipline and a given system is refuted when its disagreement with real praxis is demonstrated. It should be noted that although psychologism has been largely discredited by Frege's and Husserl's criticism, there are authors who try to revive it, for example Susan Haack in Haack (1978), Robert Hanna in Hanna (2006) or Johann van Benthem in van Benthem (2008).

[4] Remember that the model-theoretic demarcation leads to the logic of generalized quantifiers, which is not much short of full second-order logic. The proof-theoretical ones, on the other hand, can end up demarcating just the classical first-order logic.

The second approach is Platonism, which makes logic practically independent of actual praxis. Logic has got a domain of specific entities, abstract objects, which exist independently of our discourse and which it studies. Discrepancies between its (correct) claims and actual human reasoning are to be ascribed to defects of our everyday use of reason. Frege often seems to be a Platonist. There has been, of course, a heated dispute about the adequacy of such a prima facie reading, but certainly many of his passages suggest Platonism, see e.g. Frege (1884).

The third variety of approaches is distinguished from the first two by taking logic to be a normative, not descriptive discipline. Neither the praxis of reasoning, nor the realm of abstract objects is described. Instead, logic determines which reasoning is and which is not correct, i.e. how one should reason. This vague idea can be concretized in different ways. It has to be explained why such norms are instituted in the first place and what their roles are. Furthermore, it should be clear whether the norms which are stated by logic are its original creation or whether they are rather codifications of norms which are already acknowledged in a reasoned argumentation. Do we have to decide for one of these two radically different forms of normativism, if we want to be normativists? Well, the difference can hardly be explained away, but I think a viable version of normativism has to have it both ways, though it might emphasize one of these aspects more. Yet a normativism which is based purely on codification of preexisting norms ceases to be a normativism just like the normativism which ignores the norms which actually live in our daily argumentation praxis. The first one would be end up being just a variety of psychologism, while the latter just a variety of Platonism (for whence would the norms stated by a logician derive their legitimacy?). Thus any normativist approach should somehow correspond to rules of reasoning which are actually adopted, yet it has to transcend them. Classical example of a normativist would be Kant (1954) and, at least according to some ways of reading some passages, Frege (1884).

But let us postpone these considerations about possible versions of normativism, because here I would like to consider the species which can be called expressivism and which was developed by Robert Brandom (a clear statement of it can be found in the first chapter of Brandom's 2000). But for the time being, let us leave normativism unspecified and briefly reflect on the compatibility of its general shape, as well as with that of psychologism and

Platonism, with the two dominant approaches to demarcating logical constants.

Of course the issue of logical constants and the issue of relationship between logic and reasoning are divided and many connections between them are possible, but still it will be readily acknowledged that psychologism hardly seems to be suggested either by the model theoretic or by the proof-theoretic approach. Their sheer abstractness seems to make them steer far away from actual practices, which are notoriously replete with fuzziness and heavily context-dependent. The case is perhaps more clear for the model-theoretic approach. It is simply given which expressions (of course, when they have their standard meaning) are invariant with respect to the aforementioned bijections. The source of this definiteness still may be our practices but once they establish that some words are invariant (again, with their meanings, i.e. the abstract objects), then they behave independently and it is up to us to discover their properties. And the development consists mainly in finding the logical concepts and attaching names to them. Psychologism may not be lost completely, but it gets in a very difficult position, when the model-theoretic approach is accepted. Truth of logic depends on us as little as the truths of mathematics, if not as truths of natural sciences.

How about the proof-theoretical approach and psychologism? First of all, it is highly disputable whether psychological observations about human reasoning can be formulated very well in the framework of a calculus of one of the described forms. And granted that, which structural rules should be admitted? Perhaps allowing or banning some structural rules can be said to provide for mapping different areas of human reasoning. But they can also be taken as different models of the same set of practices. Nevertheless, it does not seem very plausible that in our every-day reasoning we distinguish some rules as formal in the sense of sequent calculi. Thus psychologism, though not ruled out completely, does not interact very smoothly with the proof-theoretical account of logical constants.

The marriage of Platonism with model-theoretical demarcation has the best prospects to be a happy one. They share the strong sense of objectivity of logic. Indeed, it suggests itself that logical constants denote some quite specific objects which belong to a different realm than the more mundane ones (the members of a given domain of a model). One of the specifics of modern model theory is of course its relativization of ontology in the sense that there is not one universe of what there is, the domain of each model con-

taining only some entities. But the entities which are supposed to be denoted by logical constants are nevertheless invariant over all the domains. Strictly speaking, though, the denotation of logical constants, i.e. the real shape of logical concepts, remains dependent on which models there are. By this I mean that Tarski's logical notions are identified with extensions induced over the models. And in a way this seems a little bit strange and undermines the notion of logic preceding all other knowledge and being independent of it. Though it might be legitimate to relax these foundationalist views, the Platonism just described seems to make logic dependent on the assumption that the Tarskian models represent all the possible discourses, all the possible ways our reasoning can be about something.

Keeping in mind the particularly good fit between Platonism and model-theoretic demarcation, we can expect that the relationship between the proof-theoretic demarcation and Platonism will be somewhat less harmonious. This time not some set-theoretical construct over a given domain but inferential steps (transitions from some propositions to other ones) or rather types of inferential steps are hypostatized. And these would be rather strange entities. Thus it seems to me that should one interpret the Gentzenian rules governing logical constants in a Platonic way, then they would have to be regarded just as different ways of introducing the logical notions described by model-theoretists. And since the Gentzenian demarcations demarcate weaker systems than the Tarskian ones, they are actually incomplete. Platonists can see them as different ways of illuminating the systems which should however be demarcated in the Tarskian manner.

Now what about the family of conceptions of logic which regard it as a normative discipline and though they differ from psychologism significantly, still take logic to be intimately (in a sense in which the connection is not intimate for Platonism) connected with human practices of reasoning? In this case, the proof-theoretical demarcations should square better but let us begin with the model-theoretical ones.

I do not claim that it is impossible to be a normativist and favour the model-theoretic approach at the same time. It has to be supposed that our practices are guided by rules which specify how to operate with certain logical notions. These notions should not be understood as existing independently of our practices (or more generally, of us), on pain of falling back to Platonism. The problem is that it is hardly intelligible how the extensionally, model-theoretically understood logical notions can be actually taken to be an

object of human manipulation. I am not sure whether this would be only psychologically implausible or whether it would be downright impossible. But it seems that one has to be able to individuate a given logical notion with its infinite extension. And it should be clear that inferential practices should be finite, i.e. always consist of a finite number of acts. In case, say of conjunction, the situation does not have to be so grave, as we can envisage rules which specify the correct usage of it, that is according to the abstract object it denotes, i.e. its truth function. Yet if we consider the generalized quantifiers, such as $\aleph_{131}$, it is much more difficult to make sense of it as something which can be intelligibly referred to in a specification of a rule. First of all, it is clear that the concept of such infinite cardinality has no place outside a relatively narrow context of mathematics and thus cannot be regarded as universally applicable. Furthermore, the concept of $\aleph_{131}$ is extremely unsharp, we cannot say that we understand what we say when we use it in the way we know what we say when we use conjunction or the existential quantifier. Every concept contained in a rule is bound to be vague to some degree, but it seems that we move to a new level when we envisage rules, which rely on the concepts expressed by many of the generalized quantifiers.

Why should the union with proof-theoretic demarcations be more feasible? The first simple point is that, once again, these demarcations have weaker results, i.e. they pick fewer constants and thus make logic narrower. This means that they would be less demanding on rational creatures. Therefore the problems with psychological possibility decrease. More importantly, though, the form of Gentzenian rules suggests that they codify inferences, i.e. activities of certain kind. On the other hand, these activities do not necessarily correspond to the real practices. And, as we have seen, there are more possible proof-theoretic demarcations. Which one is then to be picked out as the correct one? More specifically, which structural rules should we accept for our logical calculus? These issues are nevertheless not my principal concern, as I am only trying to argue in favour of the proof-theoretic approaches in general. What matters now is that the potential discrepancies between such calculi and actual reasoning do not have to be as troubling as they were in the case of model-theoretical approaches. Even if the inferential steps codified in the calculi do not correspond directly to the actual inferences we make on daily basis, this does not mean that they could not serve to regulate actual inferential practices as norms. But of course not every norm should be regarded as relevant to a given practice. It is therefore important to understand what

kinds of norms should be relevant to our reasoning and how they should relate to it. And, most importantly for us, whether the inferences codified by Gentzenian calculi can be regarded as such.

6. Logical expressivism

So far we have only sketched the normative account of logic. As a sketch, it can be elaborated in various ways. Here I would like to concentrate on one particular elaboration and see how it fits the proof-theoretic approach to logical constants.

The expresivism I would like to present here was presented by Brandom as a part of his inferentialism. According to inferentialism, meaning of a given word is constituted by the inferential properties of sentences containing it. Meanings of sentences, then, are determined by the sets of sentences they can be inferred from and by sentences that can be inferred from them (possibly with further premises). This has the consequence that not only formal but also material inferences are legitimate, i.e. the formal inferences are not seen as corrections of the material ones (understood as enthymemes). Nevertheless, they enable to correct our inferential practices, as they make explicit the implicit inferential commitments (by introducing conditionals) or incompatibility between statements (by introducing negation). Brandom thus explicitly provides a rationale for at least the conditional and the negation but it is not clear which one, whether e.g. the classical or the intuitionistic or still other ones.

MacFarlane (2009) takes Brandom's approach to logic to be pragmatic, contrasting it, among others, with the principled approaches, among which the proof-theoretical ones belong. It is certainly true that Brandom puts much more emphasis on the overall purpose of logic and its place in the overall epistemology than the authors who actually came up with the proof-theoretical demarcations, such as Hacking or Došen. Yet I would like to show that the proof-theoretic approach does not have to be seen as a different one but rather as compatible with logical expressivism. While Brandom gives a more broadly philosophical account of what logic is, the proof-theoretic demarcations give a more technical, clear-cut account. I suppose that Brandom's theory possesses its own appeal and may thus legitimize the proof theoretic account, while the proof-theoretic account has much more indirect attractions.

First of the attractions of proof-theoretical approach is the fact that the demarcation gets relatively unsurprising results. This is of course only a weak virtue, though it is not so unimportant that it corresponds to general intuitions and practice. Yet, of course, it is relative which logics one expects to be demarcated. Still, it is true that the generalized logics of proof-theoretic demarcators cannot be said to be standard in the way the classical first-order logic is, which is clear at least from the fact that students are first introduced to it and perhaps much later, if ever, to their generalized quantifiers and further developments.

More important is that this approach specifies the formality and therewith the generality of logic. The Gentzenian rules governing the use of a logical constant are formal in that they leave everything else unspecified, i.e. they do not depend on context.[5] Brandom requires of logical constants that they be conservative, for otherwise they could not really be used to perform the role of logic, namely to make our implicit inferential rules explicit. The constants demarcated by their formal Gentzenian rules are in general conservative (this depends also on the structural rules allowed in the calculus). Moreover, I am inclined to regard them as indispensable for our discourse, since it would not be rational without them, we could not do anything about the implicit rules and therefore would merely follow them similarly as physical objects follow the laws of physics. Here I am probably at odds with Brandom himself who would rather say that there might be discursive practices based just on implicit rules. To be a rational being, it is not enough to follow rules, but also to make them explicit and thus also to be able to question them.

In the same vein it holds that the constants specified by the proof-theoretic demarcation occur necessarily in any rational discourse. This is because of their formality. Once we engage in the enterprise of reasoning, the door for these constants simply gets open. The development of language presumably starts with some more concrete (thereby material) rules of inference and then by making the rules more and more general, we arrive at the completely formal ones, i.e. the ones governing the logical constants. For example, we can say that some expressions are more general than others, e.g. *dog* is more general than *Cocker Spaniel*, since its use is governed by more general rules. The rules for dog are presumably more schematic. For example its

[5] This is due to their being schematic, as was explained in the exposition of the proof-theoretical approach.

introduction rules demand less. Here I clearly do not mean only discursive rules, i.e. inferential rules in a narrow sense, but also the language-entry and language-exit transitions, envisaged in Sellars (1974) (a further discussion of the relationship of these transitions and inferential rules in the narrow sense in the framework of logical expressivism can be found in Peregrin 2014). Language then develops, among other things, by adding more general rules, which make more and more other rules explicit. Logical rules are somewhere at the end of this development. Ultimately it does not matter where, i.e. in which areas of reasoning our generalization gets started (if we forget about the possible controversies regarding the structural rules, as the differences on this issue can very well be motivated by applying logic to different areas). Logic is general and formal precisely because the logical constants can be arrived at from the standpoint of every specific discourse. This shows that logic is something which is bound to accompany rational discourse at its every step, i.e. it is a necessary, though of course not sufficient, condition of truth. Trying to assert something which contradicts logic is like playing a game which is lost at the very beginning. The development of language of course never stops and is not driven just by the increasing generalizations of the rules that are made explicit. And as logic can be seen as a final-point of the process of generalization, it is implicitly present in our discourse from the beginning, exactly because its rules can be arrived at no matter with which material rules we begin. Let us also not forget that there might be a legitimate discussion about which rules should be deemed as truly formal (recall the issue about the schematicity of the rules for existential quantifier).

But let us make the relationship between the formal and the material inferential rules more precise. I will try to tell the story beginning with the situation when we have only the material inferential rules. In such a language people can start to disagree about the inferences they make, some deem a given inference legitimate, others not. Such a situation can be solved only by stipulating the allowed inferential step, which amounts to stating the conditional (in case the rule is accepted), i.e. using a logical constant. This does not mean that the inference is now for ever taken to be legitimate only in its formalized shape (allowing to infer e.g. *Thunder will be heard soon* from *Lightning is seen now* only when we have the relevant conditional *If lightning is seen now, then thunder will be heard soon*), but it is possible that people might temporarily reason in a cautious mode, requiring the inference rules used to be stated explicitly.

The other point, however, is that such a process of explication and formalization of our inferential practices has to accompany our reasoning all the time because that is what makes it rational. And this process has to be guided by some norms which are given a precise form in the Gentzenian calculi. Yet to make the process of explication intelligible we have to study how the explicators, such as the conditional, behave. And to do this we consider them isolated from our actual material inferential practices. This isolation is effected by planting them into an artificial niche where no material inferences are valid, i.e. in the sequent calculus. Why do we need such a general ideal? I suppose it is so because the distance it thus acquires with respect to our actual inferential practices enables it to work as an impartial device, it constitutes its objectiveness (Wittgenstein's *Härte des logischen Muss*). And furthermore, by taking the rules governing the logical constants to be fully formal, i.e. that all the accompanying premises are only schematic, we guarantee that they do not actually change the inferential framework they are supposed to explicate, because the logical constants are the only actual linguistic items occurring in the rules. Thus by introducing them we do not make any other linguistic items enter into new inferential relationships and thus do not change their meaning. Or we do, as they enter into relationships with logically complex sentences, but equally so for all of them, thus not distinguishing any single one. The change of meaning which all the sentences and thus also all of their constituents undergo can be compared to the change of size, e.g. doubling, underwent by all the physical objects. As is well known, such a change would actually be unrecognizable and would in fact be as good as no change at all.

The change regarding the whole of language thus happens on a metalevel. Let us say that it is rather the way in which we treat language than the language itself. And it has to be said that language has its implicit logic even before the introduction of logical constants in the strictest sense. The material inferences can be said to constitute a sort of implicit logic, which nevertheless cannot ever be fully identified with its explication. Actually, it is very well thinkable that there are more explications in this sense. And therefore it does not make sense to speak of one true logic, though there are limits, and so not every system can form a logic.

And actually such an opinion is captured very well by the proof-theoretic approaches, because besides providing demarcations, they typically allow for pluralism. This pluralism, as was already mentioned, is engendered by the variety of possible structural rules that are or are not accepted. Logic is thus

presented as something which is not independent of human activities and so it also is not definite as an independent entity, but it still has some boundaries, as by far not everything can be accepted as logic.

7. Conclusion

I hope to have given some plausibility to the thesis that the logical constants (and with them the bounds of logic as a discipline) can in fact be determined. This can be done by two different argumentation strategies using different theoretical backgrounds, which nevertheless support each other. One is the proof-theoretic approach with its technical clarity but in need of further philosophical motivation and the other one is Brandom's inferentialism and expressivism which is philosophically appealing but does not say very much about the problem of pluralism in logic; Brandom (2000) also does not give us any details on what the list of logical constants should be. The details of the proof-theoretical specification can be seen as formal regimentation (with the typical virtues and also vices, of course) of Brandom's philosophical insights, which can be seen, as he himself suggests, as developments of thoughts much older. When we ask, from the expressivist position, which logics can serve the expressive role, the proof-theoretical approach gives us a good, though not fully conclusive, answer. Or at least, it provides a framework for looking for such an answer.[6]

References

BONNAY, D. (2008): Logicality and Invariance. *Bulletin of Symbolic Logic* 14, No. 1, 29-65.

BRANDOM, R. (2000): *Articulating Reasons*. Cambridge: Harvard University Press.

DOŠEN, K. (1994): Logical Constants as Punctuation Marks. In: Gabbay, D. M. (ed.): *What Is a Logical System?* Oxford: Clarendon Press, 362-381.

DUMMETT, M. (1973): *Frege: Philosophy of Language*. London: Duckworth.

DUMMETT, M. (1991): *Logical Basis of Metaphysics*. London: Duckworth.

FREGE, G. (1884): *Grundlagen der Arithmetik*. Jena: hrsg. von Christian Thiel.

[6] The present article was supported by Program for Development of Sciences at the Charles University in Prague no. 13 Rationality in the Human Sciences, section *Methods and Applications of Modern Logic*.

GENTZEN, G. (1935): Untersuchungen über das logische Schliessen. *Mathematisches Zeitschrift* 39, 176-210.
HAACK, S. (1978): *Philosophy of Logics*. Cambridge: Cambridge University Press.
HACKING, I. (1979): What is Logic? *Journal of Philosophy* 76, No. 6, 285-319.
HANNA, R. (2006): *Rationality and Logic*. The MIT Press.
HODES, H. (2004): On the Sense and Reference of a Logical Constant. *The Philosophical Quarterly* 54, No. 214, 134-165.
KANT, I. (1954): *Kritik der reinen Vernunft*. Hamburg: Felix Meiner.
KOREŇ, L. (2014a): Quantificational Accounts of Logical Consequence I: From Aristotle to Bolzano. *Organon F* 21, No. 1, 22-44.
KOREŇ, L. (2014b): Quantificational Accounts of Logical Consequence II: In the Footsteps of Bolzano. *Organon F* 21, No. 3, 303-326.
KOREŇ, L. (2014c): Quantificational Accounts of Logical Consequence III. The Standard Model-theoretic Account: Quantificational Accounts Triumphant? *Organon F* 21, No. 4, 303-326.
MACFARLANE, J. (2009): Logical Constants. In. Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/archives/fall2015/entries/logical-constants/.
PEREGRIN, J. (2014): *Inferentialism: Why Rules Matter*. Palgrave Macmillan.
PRIOR, A. (1960): The Runabout Inference Ticket. *Analysis* 21, No. 2, 38-39.
QUINE, W. O. (1951): Two Dogmas of Empiricism. *The Philosophical Review* 60, No. 1, 20-43.
SELLARS, W. (1974): Meaning as Functional Classification. *Synthese* 27, No. 3, 417-437.
SHAPIRO, S. (1991): *Foundations without Foundationalism*. Oxford: Oxford University Press.
SHER, G. (1991): *The Bounds of Logic: A Generalized Viewpoint*. Cambridge: The MIT Press.
VAN BENTHEM, J. (2008): Logic and Reasoning: Do the Facts Matter? *Studia Logica* 88, No. 1, 67-84.

Organon F 23 (1) 2016: 21-31

The Constitution Theory of Intention-Dependent Objects and the Problem of Ontological Relativism

MOHAMMAD REZA TAHMASBI

307-9088 Yonge Street. Richmond Hill
Ontario, L4C 6Z9. Canada
mohammadreza.tahmasbi@uqconnect.edu.au

RECEIVED: 16-06-2015 • ACCEPTED: 28-08-2015

ABSTRACT: According to Lynne Rudder Baker's constitution theory, all artworks and artifacts are constituted intention-dependent (ID) objects which are irreducibly real and cannot be reduced to the collections of particles which make them up. The constitution theory of ID objects is based on Baker's theory of practical realism according to which our everyday life-world is a resource for metaphysics. This paper will focus on the problem of ontological relativism entailed by the constitution theory of intention-dependent objects. I will argue, by way of an example, that the constitution theory of intention-dependent objects entails ontological relativism. That is because everyday life worlds vary from culture to culture. Finally, I examine if there is any possibility for the constitution theorist to avoid the problem of ontological relativism. I discuss Baker's idea of a thin commonsense framework.

KEYWORDS: Constitution – everyday life-world – intention-dependent objects – ontological relativism – practical realism.

1. Introduction

Lynne Rudder Baker's constitution view is a metaphysical theory of everyday objects. Ontologically speaking, the constituted object could not be reduced to its constituter. Both the constituted object and the constituter are individual objects individuated by their different primary-kind properties which

are described by substance sortals (cf. Baker 1997, 602). Baker's paradigmatic example of a constituted object is Michelangelo's *David* made from a piece of marble. *David* is an irreducible real entity which is neither identical to the marble piece (hereafter as Piece) nor is independent of it (see Baker 1999, 145-147; and Baker 2000, 33). According to Baker, whenever the constituting object *x* (e.g. Piece) with its primary-kind property F (e.g. being a marble piece) is in certain circumstances required for instantiating the constituted object *y* (e.g. *David*) with its primary-kind property G (being a statue) the constituted object will come into existence. Baker has dubbed those certain circumstances as "G-favorable circumstances". For example, if anything that has 'being a piece of marble' as its primary-kind property is presented as a three-dimensional figure in an art-world, given a title, and put on display at *t*, then there is something that has 'being a statue' as its primary-kind property which is spatially coincident with the piece of marble at *t*. As Baker explains, the constitution view can be employed in order to explain the ontological status of two groups of objects; "intention-dependent objects," or "ID objects" and "non-intention-dependent objects" or "non-ID objects".[1] Everyday objects, artworks and artifacts are ID objects while a piece of stone which is constituted by a sum of molecules is a non-ID object (see Baker 2000, 35).

Accepting Leibniz's Law, Baker states that there is a property we can use to differentiate between *David* and Piece. Therefore, they are not identical (see Baker 2000, 31; and Baker 1997, 601). *David* has the property of being a statue at any possible world and at any time, while Piece does not have such a property. There could be a world in which there is no concept of art. In such a world Piece exists but *David* does not. So, there is a possible world in which Piece does not have the property of being a statue. Piece could have existed without having the property of being a statue while *David* could not since Piece could have existed in a world in which there is no convention and practice required for instantiating art-works (cf. Baker 2000, 30). *David* has the relational property of being related to an art-world or artist's intention in virtue of which it exists whereas Piece does not have this kind of relational property; it exists in virtue of its properties which are not the same as *David*'s

[1] In her earlier works, like *Persons and Bodies*, Baker uses the expression "intentional object" to refer to ID objects. But in her other book, *The Metaphysics of Everyday Life*, she employs the "intention-dependent object" to refer to this kind of objects. See Baker (2009, 11, footnote 13).

properties (cf. Baker 2000, 39). A consequence of this idea of difference is that the existence of some entities depends on their relational properties (see Baker 1997, 603; and 2009a, 151). For every *x* and *y* if they do not have the same relational properties they do not belong to the same kind (cf. Baker 1997, 605). Thus, by virtue of relations to the artist's intention, the art-world or the history of the medium, an entity can be constituted. Baker labels entities like *David* as "Intention-dependent" objects. An ID object then is an object which exists in virtue of its relational properties which involve intentions (see Baker 2009, 10). As another example, consider one of the Dadaistic artworks made by Marcel Duchamp (1887-1968) dubbed *Fountain*. It was an ordinary porcelain urinal exhibited in 1917. It constituted a new entity when hung on a wall in an art gallery, and given a title.

According to the constitution view, constituting an ID object requires a mind. When *x* is in a world in which there is no mind *x* cannot constitute any ID object. Although the presence of a mind is a necessary condition for constituting of an ID object, it is not sufficient. In order to constitute any kind of ID object the constituting object must be intended by a mind to constitute that kind of ID object, and the mind must be in a world – or in a society – that has the situation needed for constituting that kind of ID object (see Baker 1997, 602). Hence, in regards to constituting a kind of ID object *y*, whenever a constituting object *x* is intended by a mind which is in a world – or in a society (in our actual world) – without the special conventions and practices required for instantiating that kind of ID object *y*, the constituting object *x* would have existed and there will not be any ID object. Piece in a context without conventions and practices required to produce artworks is just a piece of marble.

To sum up, the necessary and sufficient conditions for constituting an ID object are: (1) a mind; (2) a world – or a society (in our actual world) – with the conventions and practices required for instantiating that kind of constituted object; (3) the relation involving intention between the mind and the constituting object both of which are in that world; and (4) something of a suitable primary kind. In regards to constituting an ID object the existence of special conventions and practices are as necessary as the existence of a mind. Therefore, we can call this kind of object a context-dependent object.

Baker's theory of ID object depends on her theory of Practical Realism according to which what makes something the thing that it is, is determined by what it does in relation to other things and intrinsic properties have no

special authority to determine the nature of a thing. The gist of practical realism denies the idea that we must take metaphysics to be exclusively informed by science. According to the reductionist view in metaphysics, everyday objects are not irreducibly real, and we can reduce the ontological status of the ordinary things to the ontological status of particles and their sums (cf. Baker 2009, 10). By contrast, practical realism holds that ordinary objects (e.g. hammers, cars, tables, and etc.) are irreducibly real. According to practical realism, our everyday life-world is a resource for metaphysics (see Baker 2000, 22, 24; and 2009, 15). The reason for the epistemic legitimacy of practice-based claims goes back to the fact that they are used successfully in practice. Although *David* and Piece are intrinsically the same and they have the same fundamental particles – atoms or quarks – they differ in their relational properties, so they are not identical.

One problem with this theory arises when we consider the fact that the everyday life-world varies from culture to culture. If there is no "one and the same everyday life-world" for all human beings, the constitution theory and practical realism entail ontological relativism. In this article, by way of a factual counterexample, I will argue that the constitution view of ID objects entails ontological relativism unless we maintain that the theory can only be used to explain the ontological status of objects which exist in the part of our everyday life-world which is common to all cultures and societies. By "ontological relativism" I mean the view that the existence of some entities depends on cultural frameworks such that it is possible that an entity x exists at time t for a group of people with a special cultural framework C and at the same time it does not exist for the other group of people whose cultural frameworks are different with C.

2. Intention-dependent objects and the problem of ontological relativism

According to the constitution theory, our everyday-life world is the foundation for metaphysics. Now we can ask "What if the everyday life-world varies from culture to culture?" In fact, there are some examples showing that the everyday life-world varies from culture to culture. In what follows, I will discuss an example showing that the constitution theory entails ontological relativism where two different life-worlds clash.

Suppose that in a museum there is a glass box with an exhibit inside which is a marble Lord Krishna statue from the 12th century. This statue is an idol for Hindu believers and at the same time it belongs to India's history and civilization. Imagine that at time *t* the statue is considered by a Hindu believer and a serious Christian believer simultaneously. The piece of marble constitutes the statue which in turn constitutes an idol for the Hindu believer. For the Christian, however, the piece of marble constitutes only the statue and there is no idol constituted by the statue. Here we have a case of ontological relativism.

To see this, firstly, note that Krishna statues are worshipped by Hindu believers as manifestations of a Hindu God and they worship them. According to Baker, the Krishna statue constitutes an idol at *t* for the Hindu believer. Following Baker, we can say that if the idol (the Lord Krishna) and the statue are identical, then, by a version of Leibniz's law, there would be no property borne by the statue but not borne by the idol and no property borne by the idol but not borne by the statue, but the idol has a property – the property of being essentially an idol that the statue lacks. That is because, the statue could have existed in a society or world without the conventions and practices required for producing sacred things and idols while the idol could not exist without these conventions and practices. The statue could have existed in a society as an artwork and not as a sacred object. The statue is not essentially an idol and the Lord Krishna is essentially an idol. They differ in their modal properties, hence they are not identical.

(a) The Lord Krishna is essentially an idol
(b) The Krishna statue is not essentially an idol
∴ The Krishna statue ≠ The Lord Krishna

According to Baker, a statue and an idol are two primary kinds, since their primary-kind properties are different with each other. This difference rests on the fact that the causal properties of an idol are different from the causal properties of a statue. If the idol (the Lord Krishna) and the statue are identical, then, according to Baker, their causal properties must be the same. In order to understand Baker's idea of the causal properties of the constituted objects consider an example given by her where a piece of stone constitutes a monument (see Baker 2000, 33). According to Baker, ontologically speaking, the constituted object – the monument – could not be reduced to the con-

stituting objet – the piece of stone – because the causal properties of the constituted objects are irreducible to the causal properties of the constituter. Similarly, the causal properties of the idol are irreducible to the causal properties of the statue. The causal property of the idol is that it causes the Hindu believers take this object as a manifestation of their god and they worship it while the statue lacks those causal properties. It causes that the sense of the sacred arises when a Hindu believer stands in a relation with it.

The statue is in idol-favourable circumstances and based on the constitution theory necessarily the idol is constituted. The necessary and sufficient conditions for constituting the idol are met; (1) there is a mind; (2) the mind is in a society with Hindu beliefs (required conventions and practices); (3) the mind stands in a relation with the Krishna statue while the relation involves intentionality and both the statue and the mind are in the same world and the mind takes this statue as the idol; (4) we know that the statue is something of a suitable primary kind.

Now, secondly, consider that the serious Christian believer stands in front of the statue. The mind of the Christian believer stands in relation with the statue while the relation involves intentionality. The piece of marble constitutes the statue which does not, for him, constitute the idol. The Christian believer may be aware of the fact that the statue is an idol for the Hindu believers but, obviously, he is not ready to accept that there exists a real thing dubbed the Lord Krishna. This is because he does not behave as if the statue is an idol. He does not worship the statue. In fact, for the Christian believer the necessary and sufficient conditions for constituting an idol are not met.

As shown, intention-dependent objects are context-dependent objects, ontologically depending on an everyday life-world. In this example, two different contexts – two different everyday life-worlds – clash. On the basis of the constitution theory, a real thing exists for the Hindu believer at *t* which does not exist for the Christian believer at the same time. Thus ontological relativism follows.

According to Baker's view, the identity of a constituter is determined on the basis of the identity of the constituted object. That is because the constitution relation is a unity relation (cf. Baker 2000, 46). Baker state that constitution is unity without identity. She writes:

> For when *x* constitutes *y*, there is a unitary thing – *y*, as constituted by *x*… As long as *x* constitutes *y*, *x* has no independent existence… During the

> period that x constitutes *y* the identity of "the thing" – *y*, as constituted by *x* – is determined by the identity of *y*. (Baker 2000, 46)

For example, when we face the Lord Krishna we face a unified individual thing which is an idol-constituted-by-a-Krishna statue (cf. Baker 2009, 166). Baker again: "As long as *x* constitutes *y*, *x* has no independent existence" (Baker 2000, 46). In this example, at time *t*, the identity of the piece of marble is determined by the identity of the idol for the Hindu believer while the identity of the piece of marble is determined based on the statue for the Christian believer. Baker (2000, 33) writes: "The identity of the constituting thing is submerged in the identity of what it constitutes." Now the question is which one of the two constituted objects the identity of piece of marble will submerge in. Is this object – the thing which both Hindu believer and Christian believer encounter – a statue or an idol? It is noteworthy that, ontologically speaking, the constituted object is more important than the constituting object. "The constituted thing has ontological priority over its constituter" (Baker 2009, 166). This is because the constituted object has greater causal powers. Baker again:

> If we suppose that the greater a thing's causal powers, the greater its ontological significance, then a constituted thing ontologically more significant than what constitutes it. If *x* constitutes *y*, then *y* has all the causal powers that *x* has plus some new kinds of causal powers of its own. (Baker 2000, 25)

Moreover, according to Baker (2009), constitution is a transitive relation. Therefore, what exist for Hindu believer is an idol which encompasses the identities of the Krishna statue and the piece of marble and what exists for Christian believer is just a statue which encompasses the identity of the piece of marble. So, in regards to our question mentioned above we will have two different answers. "What is this?" The Hindu believer says "This is the Lord Krishna" and the Christian believer says "This is the Krishna statue". The Lord Krishna does not exist for the Christian believer. Is this object (ultimately) a statue or an idol? Given Baker's view about what ultimately determines the identity of an object, we will have different answers 'a statue' and 'an idol.'

One way to avoid the problem, based on the constitution theory, might be to maintain that the idol really exists but the Christian believer is not aware of its existence. Such a claim might be made about the existence of artworks.

For example, it is quite possible that someone who is not familiar with the history of art and Dadaism cannot recognize that the urinal – *Fountain* – is an artwork, but once he knows about Dadaism he will be able to recognize the *Fountain* as a real constituted entity. Baker states:

> If you went into a gallery of the Louvre that is lined with works by Antonio Canova and you identified them only as pieces of marble, you would be missing what is there. The constituted thing has ontological priority over its constituter. (Baker 2009, 166)

Can we say that in our example the Christian believer is missing something real? The answer is no. For Christian believer idol does not exist, even if he is well aware of the fact that the Hindu believers worship this object and they take it as an idol. As long as the Christian believer believes in Christian doctrines it is definitely not the case that there exists a real thing called the idol.

This case differs from the case of Duchamp's *Fountain*. This is because if you describe Dadaism to a person who cannot recognize *Fountain* as an artwork the person would be able to discern the artwork. But the Christian believer here is aware of Hindu beliefs. In fact, the difference is about the fact that in the case of *Fountain* the person is not aware of the conventions and practices based on which an ID object is constituted while in this case the Christian denies the existence of such conventions and practices as real things. If he accepts the conventions and practices based on which the idol is constituted, then he has already converted to Hinduism. So, from a Christian point of view the idol does not exist even if the Christian is well aware of Hindu's beliefs. In other words, in Christian's context there isn't a necessary condition which is required to constitute the idol. The Christian believer might show respect to this object if he knows that this is a sacred object for a group of people, but he does not worship the statue.

Thus, the constitution theory of ID object entails the problem of ontological relativism where the ID objects are religious objects. As another example, imagine that in a museum there is a glass box with an exhibit inside which is an upside down cross from the 14^{th} century. The cross is an evil object for serious Catholic believers and at the same it belongs to western history and civilization. If at time *t* the cross is considered by a Hindu believer and a serious Catholic believer simultaneously, then for the Catholic believer the piece of metal constitutes the cross which in turn constitutes a demonized object (with its causal powers). On the other hand, however at the same time,

for the Hindu-believer the piece of metal constitutes only the cross and there is no demonized object that it constitutes.

But what is the importance of knowing that constitution theory of ID objects entails ontological relativism? Besides the fact that so far there has been no discussion on this consequence of the constitution theory of ID objects in the literature, I believe there is more to discuss. Here, I do not intend to argue that any theory in metaphysics which implies ontological relativism must be avoided. One might accept the idea that the theory of ID entities entails ontological relativism, yet still believes in the constitution theory of ID entities as a sound theory in metaphysics. However, if one accepts the constitution theory as a sound theory while the theory entails ontological relativism, another problem arises.

Ontological relativism follows that there is no universal reality for all human beings, that is, there are different *realities*. According to the constitution theory, cultures are resources for reality. Hence, cultural relativism entails the relativity of reality. Based on the constitution theory of ID objects people with various cultures do not have one and the same reality to share. Hence, no one is in a situation to make any strict ontological judgment about the existence of entities in a culture to which she does not belong. In contrast to Baker, if someone who does not belong to western culture goes into a gallery of the Louvre that is lined with works by Antonio Canova and she identifies them only as pieces of marble, she would *not* be missing what is there. The constitution theory of ID objects entails that the reality for a person (A) whose culture is different with person B's culture can be inaccessible for person B. Consequently, the possibility of communication or interaction becomes blocked.

However, on the other hand, in our everyday life people from various cultures communicate with each other. This communication is possible only on the basis of the assumption that reality is one and the same for all human beings. At least, the reality for a person (A) whose culture is different with another person's culture (B) can be accessible for person (B). As mentioned, the constitution theory of ID objects has been based on the theory of practical realism which is, in turn, based on our everyday life experiences. According to practical realism, our everyday life-world is a resource for reality.

Now we can see the problem of ontological relativism for the constitution theory of ID objects. The theory entails that the reality for a person (A) can be inaccessible for another person (B) whose culture is different with the first

person (A). While the constitution theory of ID entities has been grounded on our experiences of everyday life, it entails a consequence that is at odds with our experiences of everyday life.

3. A suggestion to avoid the problem of ontological relativism

It seems that in order to avoid the relativism problem, constitution theorist can ground ontology only in what is common to all everyday life-worlds. If the ID objects are ontologically dependent on an everyday life-world, and if the everyday life-world varies from culture to culture, then the only way to avoid the problem of ontological relativism is to maintain that the constitution theory of ID objects can only be applied to a part of each everyday life-world which is common to all different cultures. If all different everyday life-worlds overlap each other there might be an area which is common to all culture. If the ID objects, ontologically speaking, depend on such a common area, then the problem of the ontological realism would be removed.

Nevertheless, the above suggestion would only be helpful on the basis of the idea that necessarily for all human cultures there is a common area. Therefore, the ontological problem will be removed if and only if we maintain two conditions. First, necessarily for all human life-worlds there is a common area. Second, the constitution theory of ID objects can only be applied on the common area.

If we look to Baker's (1995), we see that she has discussed just such an idea of such a common realm to all everyday life-worlds, though she has not argued that necessarily all cultures have such a common area. Talking of the concepts of thick and thin commonsense frameworks, she writes:

> I use the term 'commonsense framework' to refer to any set of concepts expressed by non-logical terms occurring in sentences understood by almost everybody in a linguistic community. The 'commonsense conception' refers to the sentences containing terms expressing those concepts. (Baker 1995, 221)

The commonsense conception reflects the everyday life-world. Everybody learns the commonsense conception of his culture when he learns the natural language of his community (cf. Baker 1995, 221). This is because the everyday life world is embodied in natural language (see Baker 1995, 223).

According to practical realism, different cultures have different thick conceptions of common sense. But there is a commonsense framework in a thin sense which does not vary from culture to culture. It provides a common background for all cultures which includes concepts of medium-sized objects. Baker (1995, 222) writes: "In the broadest terms, the thin commonsense framework is not restricted to some particular outlook that may vary from culture to culture. Rather, it provides a common background against which differences among cultures become visible." If, as Baker says, there is a thin conception of common sense framework for all human beings, then we can state that the constitution theory can avoid the problem of ontological relativism if we maintain that the constitution view of ID objects is about the part of everyday life-world of each culture which can be reflected in the thin commonsense framework. But, then, the constitution theorist may face another problem. How can she recognize which part of an everyday life-world is the part that can be reflected in the thin conception of commonsense framework? Is there any possibility to provide a criterion to enable us to know that the constitution theory is about which part of our everyday life-world? I think this is a hard task to do for the constitution theorist, if she wants to avoid the problem of ontological relativism.

Acknowledgment

I am extremely grateful to Associate Professor Deborah Brown and Dr. Dominic Hyde of the University of Queensland who read the manuscript and made critical comments and valuable suggestions.

References

BAKER, L. R. (1995): *Explaining Attitudes: A Practical Approach to the Mind*. Cambridge University Press.

BAKER, L. R. (1997): Why Constitution Is Not Identity. *The Journal of Philosophy* 94, No. 12, 599-621.

BAKER, L. R. (2000): *Persons and Bodies: A Constitution View*. Cambridge University Press.

BAKER, L. R. (2009): *The Metaphysics of Everyday Life: An Essay in Practical Realism*. Cambridge University Press.

BAKER, L. R. (2009a): Unity Without Identity: A New Look at Material Constitution. In: French, P. A. – Wettstein, H. K. (eds.): *New Directions in Philosophy: Midwest Studies in Philosophy*. Malden: Blackwell Publishers, 144-165.

Organon F 23 (1) 2016: 32-49

Do Emotions Have Directions of Fit?

SUNNY YANG

Institute of Human, Environment and Future. Inje University
#705 Illgangwon. Inje-ro 197. Gimhae City Gyeongnam. South Korea
sunnyyang@inje.ac.kr

RECEIVED: 16-06-2015 • ACCEPTED: 02-11-2015

ABSTRACT: My aim in this paper is to provide a series of arguments against the conception of emotional truth. If we accept the idea that emotions are eligible for being truth-apt, then we are conceding to the view that emotions are capable of having epistemic warrant. Many contemporary writers regard this kind of warrant as the concept of appropriateness or fittingness that is taken to be analogous to truth in the emotional realm (e.g. D'Arms – Jacobson 2000a, Nussbaum 2001, de Sousa 2002, Morton 2002, Goldie 2004). Yet, if we allow an analogy between appropriateness and truth, it would seem to allow that emotions are capable of being true or false. However, I argue against the concept of truth in the emotional realm, for there are some emotions that cannot be reduced to propositional attitudes which are eligible for being truth-apt, unlike beliefs, thoughts, and judgments. I shall demonstrate these cases in terms of recalcitrant emotions. Especially, I argue that some emotions are not eligible for being truth-apt by utilizing the notion of 'direction of fit'. I argue that emotions have neither directions of fit, since emotion is only embedded in belief or desire. Finally, I conclude that appropriateness of emotions differs from truth or satisfaction by demonstrating that the norms of belief/desire differ from norms of emotion. Hence, I argue, it is a mistake to give an account of all these, namely, belief, desire and emotion in terms of rationality.

KEYWORDS: Appropriateness – belief – desire – direction of fit – emotion – satisfaction – truth.

1. Belief/desire and emotion: a Humean functionalism

In explaining emotion, there are three kinds of strong view, which reduce emotion to feeling, evaluative judgment/belief or desire respectively. In this paper I shall argue against only the latter two types: a conception of emotion which attempts to view it essentially in cognitive terms, in particular in terms of belief and desire. The first position I want to discuss in this section is a Humean functionalism, according to which emotion is reducible to desire. In order to illuminate this, in what follows, I shall utilize the concept of 'direction of fit'. The concept of direction of fit is often attributed to Anscombe, for example by Platts (see Anscombe 1957; see also Platts 1979, 256).[1] In Anscombe's story, there are two men, a man going shopping with a shopping list, and a private detective who tries to write down what the man buys. Anscombe contrasts a mistake in the performance of someone buying from a shopping list, with a mistake in a record which a detective is keeping of the shopping purchased (cf. Anscombe 1957). Then she asks "What is the difference between the man's shopping list and the detective's list?" Her answer is as follows:

> It is precisely this: if the list and the things that the man actually buys do not agree, and if this and this alone constitutes a *mistake,* then the mistake is not in the list but in the man's performance… whereas if the detective's record and what the man actually buys do not agree, then the mistake is in the record. (Anscombe 1957, 56)

In this passage, Anscombe suggests that the difference between belief and desire is similar to that between the detective's list and the shopping list. As Danto remarks, the shopper may make a mistake by putting in the basket something that is not on the list – beer when the list says 'beef'; the detective may make a mistake in writing 'beef' when the basket contains beer. The shopper corrects the mistake by replacing the beer with beef; the detective by writing 'beer' instead of 'beef' (see Danto 1999, 56). This way of contrasting beliefs and desires in terms of 'direction of fit' is, as Platts in Platts (1979, 257) rightly claims, highly metaphorical, but many critics believe that the metaphor is very useful in illuminating the role of beliefs and desires in ex-

[1] L. L. Humberston suggests that it originate with J. L. Austin; see Humberston (1992).

plaining action, especially, reasons for action (see, e.g., Smith 1994, chap. 4; see also Searle 2001). According to this idea, desire is aimed at some end, whereas belief plays a role in achieving that end. This is what Smith calls teleological explanation of action, according to which an agent's reasons are explained by the agent's attempt to achieve some purpose or goal. Smith argues that this is so because desires have 'world-to-mind' direction of fit.[2]

What then is 'world-to-mind' direction of fit? Smith believes that 'having a motivating reason just is, *inter alia,* having a goal' and 'the having a goal is being in a state with the direction of fit of desire' (Smith 1994, 116). But the trouble with this account is that it faces a circularity problem. For in order to explain having a motivational reason, it appeals to the idea of 'having a goal'. Then having a goal is defined as 'being in a state with which the world must fit'. Now in order to explain the concept of direction of fit, one appeals to 'having a goal'. So the argument seems to be circular. In order to avoid this difficulty, Smith appeals to a functional account of direction of fit, dropping the idea of goals or purposes.[3] He writes:

> For the difference between beliefs and desires in terms of directions of fit can be seen to amount to a difference in the functional roles of belief and desire. Very roughly, and simplifying somewhat, it amounts, *inter alia*, to a difference in the counterfactual dependence of a belief that *p* and a desire that *p* on a perception with the content that not *p*: a belief that *p* tends to go out of existence in the presence of a perception with the content that not *p*, whereas a desire that *p* tends to endure, disposing the subject in that state to bring it about that *p*. (Smith 1994, 115)

There have been detailed criticisms of Smith's account.[4] However, in this paper I focus on some difficulties arising from applying this formulation to

---

[2] Smith discusses this in his earlier article; see Smith (1987, 36-61). In this paper I focus on his later discussion – see Smith (1994).

[3] The circularity problem has been pointed out by Schueler (1991) and Zangwill (1998). In order to avoid the circularity problem, Humberston (1992) has appealed to second-order background intentions.

[4] The criticism comes from Schueler (1991); Humberston (1992); Zangwill (1998); Sobel – Copp (2001); Dunn (2004). Schueler's and Humberston's target of criticism has been Smith's former discussion in Smith (1987), whereas recent writers focus on his later version – see Smith (1994). I too focus on his later discussion.

emotions. Although Platts glosses Anscombe's view as that only desire is 'a prime exemplar' of those states having world-to-mind direction of fit, Smith is invested in thinking that the account does cover emotions, since he says that states that include having a goal will count as desires on his broad construal (cf. Smith 1994, 117).

2. Emotions and direction of fit

Having established a view of emotion, namely, a Humean functionalism, according to which emotion can be reducible to desire, in what follows, I shall demonstrate that such view is implausible in explaining the intentionality of emotion by applying the concept of direction of fit to emotion.

It might be thought that *hoping* that *p* counts as a desire for Smith and is an emotional state. If so, one might say, hope will have to have, with desire, a world-to-mind direction of fit. The reason why hope, like desire, could have world-to-mind direction of fit is that 'he hopes the team wins but he doesn't want them to' is contradictory. Another reason that hope could have the same direction of fit as desire, one might argue, is because hopes aim at satisfaction rather than at truth. If my hope fails to 'fit' the world, that is not any defect in my hope but in the world. However, the trouble is that if we take hope as having world-to-mind direction of fit like desires, it cannot cover some kinds of hope close to cognitive states. For example, if I know that it is raining, then I can't hope that the road is not wet, though I can wish it were not. When we hope, we also believe that it is possible that something we can do will 'make it more likely that this hope (or other pro-attitude) gets realized'. We cannot have a hope or desire that cannot possibly be satisfied, unless we are committed to irrationality. If this is right, then according to Smith's functional account of direction of fit, 'all hopes will have, like beliefs, a mind-to-world direction of fit,' since 'I can't hope that *p* once I discover that *not p*' (as argued by Schueler 1991, 279-280). If this is the case, we can say that hope that *p*, like beliefs, tends to go out of existence when we perceive that not *p*.

But the problem with seeing hope as having a mind-to-world direction of fit like belief is shown by cases of 'recalcitrant' belief. We can compare the recalcitrant mental state with optic illusions of the Mueller-Lyre lines: the Mueller-Lyre lines continue to appear to be of different lengths while they are known to be equal lengths. In this respect, many people try to develop

this idea to show how, at times, our beliefs, feelings and actions seem to behave with a mind of their own despite our best efforts in trying to control them. This is because, S. Döring argues, the emotional content 'resembles the content of sense-perception in that both kinds of representational content need not be revised in the light of belief and better knowledge' (Döring 2003, 223). She attempts to show this using an analogy emotion with optic illusion of the Mueller-Lyre lines as follows: just as our perception that the two lines differ in length persists in the face of our belief that the lines are the same length, so emotions may persist even though we have relevant and countervailing knowledge. Thus, we can say that a person can have a sense-perception or can feel emotion that *p*, and at the same time believe that *–p*. However, this is not the case in believing that *p*. If this is true, it follows that an emotion, like a sense perception, is not an attitude that can be considered as true or false as such. The point that Döring makes here is that in such cases of sense-perceptions and emotions we are able to be in a contradictory or ambivalent state, while in the case of belief, this is impossible.

Having elaborated on the analogy between emotions and sense-perceptions, now I want to argue against a cognitive conception of emotion which attempts to view it essentially in terms of belief. K. Walton says that 'it is impossible to have the emotions without accompanying appropriate belief'. He writes:

> It seems a principle of common sense, one which ought not to be abandoned if there is any reasonable alternative, that fear must be accompanied by, or must involve, a belief that one is in danger. Charles does not believe that he is in danger; so he is not afraid. (Walton 1979, 6-7)

Given this, cognitivist, for example Kendall Walton, might require that if emotion is accompanied by a belief, and the belief has a direction of fit, namely, a mind-to-world direction of fit, then in some sense the emotion has that as well.[5] Now if we apply this idea to Smith's above formulation, we can

[5] There is a strong version of cognitivism that attempts to view emotion as essentially thought or belief. The question whether the theory of emotion should count emotion as cognitive or not is the question whether cognitive elements, such as belief and judgment, are conceptually necessary or not for having emotion. Robert Solomon has a preference for 'judgment' rather than 'thought' as the label for that cognitive element. Jerome Neu has argued that 'thoughts' rather than judgments are appropriate

gloss it in functional terms as follows: an emotion that *p* is a mental state that tends to go out of existence in the presence of a perception with the content that not *p*. If I believe that there is nothing to be afraid of in the presence of a perception with content that not *p* (e.g., seeing a rabbit is not harmful), then I should not fear it. If I do, it follows that it is a mistake to fear *p* when not *p*, and the fear should be abandoned. But it is possible that fear persists, even when I consciously believe that the rabbit is harmless. If emotions imply, like Solomon and Nussbaum say, evaluative belief, we would have to be said to hold inconsistent beliefs as follows: an evaluative belief (fear) tends to *endure* in the presence of perception that not *p* (the rabbit is harmless). If this is true, is it a mistake, and should it be abandoned? But in actual life we have those emotions; consider the case of phobia. The recalcitrant emotion of groups, as Susan James points out, are even harder to dismiss. For example, as Susan James puts it, a man as a member of the American Republican party may fear Islamic fundamentalism and holding the unchanging belief that fundamentalists are dangerous, despite evidence to the contrary (see James 2003, 228). If a strong view which reduces emotion to belief were right, it could be said that the man or woman, who does not believe that the rabbit is particularly dangerous, cannot be afraid of it, since according to the view, the relevant belief is a necessary element of the emotion. However, the strong cognitivism cannot explain the above cases which are examples of emotional recalcitrance. Hence it follows that emotional recalcitrance gets strong cognitivism into trouble.

Let's apply recalcitrance to the case of hope which is supposed to have mind-to-world direction of fit, like belief. I might still hope, for example, that my team will win even though all the evidence is to the contrary. The Labours genuinely hope they will win the next election. Should the Labours give up their hope that they will win? I'm afraid, the answer is no, because if they did, they would be sure not to win. This raises difficult issues about ra-

---

for capturing what is the cognitive core of emotion. With regard to 'cognition', there has been a lively debate within the 'cognitivist' camp whether the type of cognition in question is better thought of as belief, thought, judgment, or something else. Some writers, for example, Walton (1979) favour beliefs. Neu (1977) suggests that the cognitive elements are thoughts. Solomon (1976; 1980) and Nussbaum (2001; 2004) defend the view that emotions are evaluative judgments. Roberts (1988) suggests 'construal' as an alternative.

tionality, which are beyond the scope of this paper.[6] The point of this objection is to ask whether the cognitive element involved in the emotions is necessarily a state of *belief*. Indeed, cognitive theorists do not always claim that the cognitive element, which is essential to the emotions, should be 'belief.' Belief, among the many cognitive attitudes, is somehow a 'very strong' kind of attitude, for belief itself is necessarily connected with evidential rationality or justification. I do not have a belief when I have no or insufficient relevant evidence. In other words, if cognitivists claim that the cognitive element, which is essential to the emotions, is belief, there is a risk of ruling out a number of 'irrational emotions.' In this respect, cognitivists substitute 'judgment', 'evaluation', 'appraisal,' 'apprehension', or simply 'construal' for the belief. In rejecting the crudely cognitivist approach by assessing whether emotions have the kinds of directions of fit found in belief and desire respectively, I hold the cognitive aspect of emotion to be construal rather than belief – the latter is 'cognitively penetrable'[7] unlike the former and unlike emotion.

Another difficulty with Smith's functionalist account of the 'direction of fit' concerns the latter part of the formulation of desire: 'a desire that *p* tends to endure, *disposing the subject in that state to bring it about that p*.' Suppose that you read about what Caligula did and became angry at the injustice of it all. The propositional content of your emotion, in this case, might be that Caligula is an awful person. But seeing Caligula as an awful person does not by itself provide an end for action, and it won't dispose you to move to act, since in this case there is nothing you can do in order to change the world in such a way that Caligula no longer appears awful to you. I think what he did, ought not to have happened, but realise that I cannot do anything about it. Let's apply this idea to another example; my hoping my team will win. If, as Smith argues, an emotion can be reduced to a merely functional state [that is

---

[6] By the rationality of emotion, I mean the appropriateness or fittingness, while I take the rationality of belief/desire to be the truth/satisfaction, respectively. In this paper I shall only focus on the problem of the rationality of emotion, that is, the appropriateness.

[7] I owe this terminology to Peter Goldie. I think 'cognitive penetrability' and 'cognitive impenetrability' are helpful to explain our emotional experience. Our emotional experiences are cognitively penetrable only if they can be affected by our beliefs. On the other hand, our emotions are cognitively impenetrable when they are not affected by our persistent beliefs. Goldie uses the terminology 'cognitive impenetrability' as the same meaning as 'recalcitrance.' See Goldie (2000, 74-78).

a desire] that disposes the subject toward action, this might involve a desire to put it right. How then does my emotion – hope – move me towards action? I may not do anything to put it right. Rather, my desire involves an attitude toward (a pro-attitude) the winning: I want my team ought to win, but realise that I cannot do anything about it. If this is right, Smith's view, according to which emotion can be reducible to desire, gets into trouble.

Now one might argue that the latter part of Smith's formulation – 'disposing the subject in that state to bring it about that p' – has another difficulty when we apply it to the case of hope about the past. Imagine a case of hope about the past: I hope that I did not behave idiotically at last night's party. In this case, although my hope tends to endure when I know that I behaved idiotically, it won't dispose me to bring it about that I behaved charmingly. For I can do nothing *now* to make it the case. If I regret my behaviour, in this case, it may dispose me to bring it about that *p*; I may try to apologize to the host and to avoid the people who I met at the party. But in the case of my hope that I behaved charmingly, I won't be disposed to make that which I hope for the case, even though my hope does not tend to go out of existence. Hence in this case we cannot say that hope has the same direction of fit as desire. In response to this kind of objection, Smith might say that these can be looked upon as (future directed) hopes about what will turn out about the past. If this is so, then Smith's test still seems to work in this case.

I have demonstrated so far that a direction of fit account cannot deal with emotion, for example, hope. The best way to handle hopes on a direction of fit view would be to say that they involve having a desire for *p* plus the belief that it is still possible that *p*. In general, I have attempted to show, using Smith's analysis that it is implausible that all emotions have the same direction of fit. A fear is inappropriate if the mind doesn't fit the world, and a hope is wrong if the world doesn't fit the mind. For some emotions, we strive to bring about their objects and, for others, we try to avoid their objects. In what follows, I shall use the idea of recalcitrant emotions to show why emotions have neither mind-to-world, nor have world-to-mind direction of fit. I shall discuss cases of pride and jealousy which will provide vivid illustration of this.[8]

---

[8] I am not trying to discuss jealousy and pride in general. I have picked out instances of a kind that are likely to be rated as counter examples to the view that emotions have either mind-to-world (belief) or world-to-mind (desire) direction of fit.

3. Emotional recalcitrance and personal perspective

One might say that emotions may have either, mind-to-world or world-to-mind direction of fit, due to their involving either belief or desire. There are cases in which emotions sometimes have belief-like directions of fit. For example, 'pride' sometimes does not involve desire but belief. My pride at my daughter's achievements depends on the belief, evaluable as true or false, that she has had them. When I am proud of my beautiful house it is because of my belief that the object is mine. This is 'cognitively penetrable.' In being proud of my beautiful house, I first of all must believe that it is valuable; secondly, in order for the feeling to play a role I must believe the house to be in some way connected with me. G. Taylor calls those two beliefs 'explanatory' and 'indentificatory' belief, respectively (cf. Taylor 1985, 27). The 'explanatory' beliefs just explain the relation between the valuable things and the person, whereas 'identificatory' belief refers to something 'closely' related to the person who feels pride. Thus, according to Taylor, 'a person may hold the requisite explanatory beliefs and yet not feel proud.' 'She may regard her beautiful house as a most desirable possession but may not regard this as reflecting on her own worth' (Taylor 1985, 34). Thus in order to feel pride there must be indentificatory belief that 'the agent regards the desirable as something she herself has brought about.' That is, she must regard the information given by explanatory beliefs as her worth.[9] But if we accept this view we cannot explain the following case: in the case of the triumph of the team which I support, pride may involve 'explanatory belief,' but not involve 'identificatory belief,' since I cannot regard the team's victory as one that I myself brought about. Thus in my view, the pride in the triumph of the team does not derive from belief but from my trying to think of the team's victory as mine. In trying to think of the teams' victory as mine, there is no 'fit' or 'directionality', since trying to thinking of X as Y is subject to one's will. This is one of the difference between 'thinking of X as Y' and 'belief', since as Goldie notes, 'believing at will is, as is generally accepted, impossible; one cannot directly try to believe something; at best one can indirectly (albeit irrationally) try to come to believe that thing by, for example, partly oneself in an environment where one is likely to do so' (Goldie 2000, 72). If this is so, we can say that the pride in the triumph of the team is derived from my attitude of thinking

[9] This is, according to Taylor, a sufficient condition for pride.

X as if Y. If this is true, in this case, we can say that pride does not have belief-like direction of fit.

My pride, one might argue, presupposes a desire that my daughter has achievements, a desire with satisfaction-conditions and the world-to-mind direction of fit. But that pride does not itself have satisfaction conditions, for it does not itself set goals for action. We could say that pride sometimes involves mind-to-world (belief), sometimes world-to-mind (desire), or sometimes neither directions of fit. It follows that it is difficult to say that emotions have either direction of fit. The reason why emotions sometimes have neither direction of fit is because we have the personal point of view when we experience emotion. We have different standards of fittingness when we experience emotion. Hence the appropriateness of the emotions can be stated from the agent's perspective.

If we take jealousy as having the same direction of fit as desire, that is, world-to-mind direction of fit, then it could be said that jealousy aims at satisfaction. What then is satisfaction of jealousy? For example, I might just wish my rival ill or dead. I may choose the way in which I confront my rival: a fight or to work hard. If I have low-esteem, I may be depressed or self-destructive. But this emotional reaction cannot give me ultimate satisfaction. One might say that the ultimate satisfaction of one's desire involved in jealousy extinguishes that desire. One way to terminate the desire might be to get rid of a belief that the rival is superior to the agent. If this were true, then one might argue that jealousy has the same direction of fit as belief, for the desire involved in jealousy rests on belief.[10] In what follows I want to show why we cannot say that jealousy has belief-like direction of fit.

4. The appropriateness of emotion

If we take emotions as having a belief-like direction of fit, we cannot explain kinds of malicious envy, or jealousy, for these cases are sometimes 'cognitively impenetrable'. Suppose that Kate and Lucy have been friends and they assess themselves by their success in some competition. They apply

[10] Taylor, for example, argues that jealousy is unjustified, since this emotion must include a false or unjustified belief. See Taylor (1975, 401-402); see also Ben-Ze'ev (1990, 506).

for a job at the same place. But Lucy is accepted and Kate fails in the final interview. So Kate is jealous of Lucy's success. In being jealous of Lucy, Kate thinks of Lucy as an enemy to her self-esteem. So despite their friendship, Kate wishes Lucy is ill. Kate wants Lucy to be absent from the interview due to the illness.[11] This is because, Taylor would say, Kate's jealousy involves false belief, and is morally wrong belief that rests on irrationality and cannot be justified. Now the presumption of this kind of strong cognitivist's inference, it seems to me, is that there is an analogy between our emotional response and the world, and a true belief and the world. According to them, the former relation can be described as 'fittingness', whereas the latter one is maintained by truth. Then the cognitivists argue that these attitudes are both rational. Since they take both to be rational, they argue that emotion can be justified or criticized. Furthermore they claim that justification or criticism of the emotions is always committed to morality or prudence. Taylor claims that 'an emotional reaction is unjustified if it rests on irrationally mistaken beliefs or when it is disproportionate to a given situation' (Taylor 1975, 393). But this is, we can say, what some call a 'moralistic fallacy' (see D'Arms – Jacobson 2000a, 73-74). 'The moralistic fallacy,' according to D'Arms and Jacobson, is 'simply to infer, from the claim that it would be morally objectionable to feel F toward X, that therefore F is not a fitting response to X.' The reason why this inference is fallacious, as they say, is that 'an emotion can be fitting despite being wrong to feel'. If this is right, the emotions' appropriateness or inappropriateness can be said to depend not on the belief which is true or false, but on person's narrative: the situations, his or her other emotions, his moods and his character. When you laugh at an offensive joke, the funniness of the joke might be understood by you in the light of your delighted mood after a delightful day, or in the light of your general disposition to be cheerful. In order to understand the standard fittingness of emotion, let us consider the norm of emotion.

According to de Sousa (1987, 116), an emotion is directed at a 'target'. For example, being afraid of a snake implies that you are thinking of it as dangerous. If this is right, it can be said that an emotion has an intentional content that is evaluative. In order to lend support to this idea that emotions imply evaluations of their targets, de Sousa goes on to argue that each emotion-type has a so-called 'formal object' (corresponding to truth as the formal

[11] R. Roberts presents a similar case to mine. See Roberts (1991).

object of belief). Now in order for the formal object to be the possible target of an emotion, the subject must see the object as having a certain property; otherwise the emotion would not be intelligible. For example, your fear of the snake is only intelligible if there is some feature of it – looking fearful, in this case – which explains why you see it as fearsome. If you were saying that you are afraid of the snake while at the same time denying that there is anything fearsome about it, it would appear to be nonsense. Hence in order for your fear to be intelligible in this case, it is said that the property which has to be ascribed to an emotion's target is the emotion's formal object.

The term 'formal object' is derived from medieval philosophy and has been applied to emotions by Anthony Kenny. According to Kenny, the formal object of a state is the object under that description which must apply to it if it is possible to be in this state with respect to it (cf. Kenny 1963, 189). According to him, when we describe the formal object of emotion, the description refers to belief, that is, one has to believe that something is dangerous in order to feel fear. Recently, however, it is commonly believed that the formal object of an emotion is a property. Hence, de Sousa holds that 'the formal object of fear – the norm defined for its own appropriateness – is the Dangerous' (de Sousa 2002, 251).

The formal object of emotion view in much recent times has been defended by those endorsing a perceptual account of emotions (see de Sousa 2002; 2004; Prinz 2004a; 2004b; Döring 2003; Tappolet 2005). According to this view, an emotion like fear has correctness conditions that are similar to the correctness conditions of perceptual states. Fear is correct or appropriate in so far as its object is really dangerous, in that this is what makes fear correct or appropriate. In his recent discussion, de Sousa claims that emotional truth is concerned with the correctness of the emotional evaluation, and holds that it refers not to semantic satisfaction but to success (cf. de Sousa 2004, 72). In order to lend support to this idea, de Sousa makes a further distinction between emotion with a propositional object and emotion with a direct object. He formulates the former and the latter as follows: 'E (p) is *satisfied* iff p is true, [while] E (p) is *successful* if p actually fits E's formal object' (de Sousa 2004, 72). For example, the formal object of fear is the property of being dangerous. A rabbit-phobic person's fear of rabbits is satisfied if rabbits exist but it is not successful if rabbits are not dangerous. On the other hand, if someone is afraid of aliens, his or her emotion is not satisfied if there are no such aliens. However, it may be successful, since aliens could be dangerous

if they existed. If this is so, emotional truth can be said to be whether the particular emotional object fits the relevant formal object.[12]

Now if de Sousa's argument so far were right, one might raise a question, namely, how do we evaluate the correctness of emotional evaluations? Faced with this question, de Sousa introduces his axiological hypothesis of emotions as perceptions of value. According to him, we can understand what values are only in virtue of our emotional responses. De Sousa calls his view 'axiological holism' for 'it stipulates that we do not apprehend value in discrete units but only in the light of a complex of factors that transcend individual experience' (se Sousa 2002, 255). Among these factors are biological facts, social norms, and 'paradigm scenarios' of individual biography. Yet none of them alone constitutes the norm for emotional truth. '[I]nstead it is the totality of all these factors – biological facts, social, personal, and more – that may properly be confronted with one another in the hope of arriving at something like reflective equilibrium' (de Sousa 2004, 74). Hence, when we apply this holism to our case of jealousy, the evaluation of a person's disposition to be jealous can be made relative to that person's narrative: the situations, his or her other emotions, his moods and his character.

Let's return to our jealousy case: Kate is jealous of Lucy's achievement. I may say that Kate's jealousy is inappropriate, since Lucy made more of an effort, and deserved her success, while Kate did nothing. If my evaluation were right, I could say that it is morally wrong for Kate to be jealous of Lucy. Although my judgment is right, it cannot be right to assess the appropriateness of Kate's jealousy morally, since jealousy does not represent its object as bad or unjust. In my view, jealousy is comparable to the following cases: the snake is not dangerous and at the same time I fear it; we believe that the Mueller-Lyre lines continue to appear to be of different lengths while they are known to be of equal lengths. As I have said before, this implies that emotion, like sense-perception can present in the face of relevant and countervailing knowledge. In this respect, I argue that emotion which do not van-

[12] Recently Mikko Salmela presents an argument for de Sousa's account of emotional truth by arguing that emotions have cognitive content as digitalized evaluative perceptions of the particular object of emotion, in terms of the relevant formal property. Salmela develops this idea by arguing that an emotion is an actual fit between the particular and the formal objects of emotion, and the emotion's propositional content is semantically satisfied, or the target of the emotion exists. See Salmela (2006).

ish in the face of better knowledge is recalcitrant emotion. This is because to some extent our emotions and emotional responses are passive, and cannot be controlled. There are several senses of being unable to control an emotion: 1) being unable to helping feeling it. (e.g., feeling sorry for a burglar); 2) being unable to help expressing it (e.g., the witness in court bursting into tears); 3) being afraid of a grass snake, despite believing that it poses no danger. The third case is one of recalcitrant emotion, which is the emotion over which the agents have no control, even though they believe that *p*. In a similar vein, we can call jealousy recalcitrant emotion.

Although one might doubt that the jealousy is always recalcitrant, we can say that Kate's case is recalcitrant in the sense that her emotion and emotional response are passive and cannot be controlled in the face of relevant and countervailing knowledge.[13] This is because the subject's attitude toward those involved in jealousy is very complex and the desire in which jealousy is involved depends on all kinds of things: the circumstances, character-traits and moods that the person has.

We have seen so far that we cannot say that jealousy has either a belief-like or a desire-like direction of fit. When we say that X is jealous of Y over Z, it may involve desire, belief and perhaps affective elements (feeling). If jealousy has an affective element (feeling), what is it? Suppose that when Kate feels jealousy of Lucy, she may have feeling of weight in the stomach, becomes very nervous, or severely depressed. In my view, all these reductions are inadequate, since emotions are too complex to be identified with one of them. I call such crude views into question by applying the concept of direction of fit to emotions and by presenting recalcitrant emotion – for example, pride or jealousy – that illuminate the phenomenon of emotional recalcitrance. It can be said that this kind of emotional recalcitrance is derived from forcefulness of feeling. Those emotions tend to get out of control, due to the passivity of feeling. We can say that it is appropriate to call jealousy a passion. In this respect, we can say that the felt emotion is appropriate in the sense in which it is experienced in one way or another.

[13] In support of my view that emotions are not easily reducible to belief, desire or feeling and we cannot say that they have either directions of fit, I choose pride and jealousy which are tricky of saying whether they have one of directions of fit.

5. Conclusion

In this paper I have rejected the idea that there is a correspondence between our emotions and the world. If we accept the view that emotions correspond to the world, then we are conceding to the view that emotions are capable of having epistemic warrant. Many contemporary writers regard this kind of warrant as the concept of appropriateness or fittingness that is taken to be analogous to truth in the emotional realm (see, e.g., Goldie 2000; D'Arms – Jacobson 2000a; Nussbaum 2001). Yet, if we allow an analogy between appropriateness and truth, it would seem to allow that emotions are capable of being true or false.[14] However, many philosophers have circumvented the concept of truth in the emotional realm, for there are some emotions that cannot be reduced to propositional attitudes which are eligible for being truth-apt, unlike beliefs, thoughts, and judgments. I have demonstrated so far these cases in terms of, for example, recalcitrant emotions, phobia, pride and jealousy. Especially, I have shown these cases by utilizing the notion of 'direction of fit'. I have demonstrated the relation between direction of fit or correspondence and appropriateness. I have argued that correspondence would not be an asymmetrical whereas direction of fit is an asymmetrical relation between mind and world. I have shown that direction of fit means something like one is appropriateness given the other, which is asymmetrical. Hence it is more than correspondence.

I have presented a series of arguments against a crude view of emotions, namely, a Humean functionalism. My reason for rejecting the crude view rests on a kind of holism, since as Goldie (2000, 235) notes, 'our emotions, moods, and character traits, broadly conceived, can interweave, overlap, and mutually affect each other.' Hence, when we apply this holism to our case of jealousy, the evaluation of a person's disposition to be jealous can be made relative to that person's narrative: the situations, his or her other emotions, his moods and his character. In evaluating a person's emotion, the person's character and mood may play an important role. When you laugh at an offensive joke, the funniness of the joke might be understood by you in the light of

[14] In this respect many contemporary emotion theorists tend to be committed themselves to the claim that emotions are able to be true or false, when they say such things as the appropriate emotions 'enable us to get things right' (see Goldie 2004, 99).

your delighted mood after a delightful day, or in the light of your general disposition to be cheerful. If this is right, the emotions' appropriateness or inappropriateness can be said to depend not on the belief which is true or false, but on a construal depends on personal perspective.

We have specifically seen that Smith's descriptive account of direction of fit fails to explain mental states other than belief and desire, especially, emotions. Furthermore, Smith's descriptive account cannot explain whether an emotion is fitting, for this is a normative question. When we believe p, our attitude is regulated by epistemic norms. On the other hand, I have argued that emotions' fittingness or unfittingness depends neither on the belief which is true or false, nor on the desire which is satisfied or frustrated. Instead, I have argued that emotions' fittingness or unfittingness depends on the construal which depends on personal perspective. Hence, I argue, it is a mistake to give an account of all these, namely, belief, desire and emotion in terms of rationality.

Acknowledgment

I am very grateful to Andy Hamilton and Nick Zangwill for helpful comments on earlier drafts. I owe a special debt to Peter Goldie for invaluable suggestions.
This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2014S1A5B8063466).

References

ANSCOMBE, G. E. M. (1957): *Intention.* Oxford: Blackwell.

BEN-ZE'EV, A. (1990): Envy and Jealousy. *Canadian Journal of Philosophy* 20, No. 4, 487-516.

BEN-ZE'EV, A. (2004): Emotions Are Not Mere Judgments. *Philosophy and Phenomenological Research* 68, No. 2, 450-457.

D'ARMS, J. – JACOBSON, D. (2000a): The Moralistic Fallacy on the 'Appropriateness' of Emotion. *Philosophy and Phenomenological Research* 61, No. 1, 65-90.

D'ARMS, J. – JACOBSON, D. (2000b): Sentiment and Value. *Ethics* 110, No. 4, 722-748.

D'ARMS, J. – JACOBSON, D. (2003): The Significance of Recalcitrant Emotion. In: Hatzimoysis, A. (ed.): *Philosophy and the Emotion*. Cambridge University Press, 127-146.

DANTO, A. C. (1999): Basic Actions and Basic Concepts. In: *The Body/Body Problem.* Berkeley: University of California.

DEONNA, J. A. (2006): Emotion, Perception and Perspective. *Dialectica* 60, No. 1, 29-46.

DE SOUSA, R. (1987): *The Rationality of Emotion.* Cambridge (Mass.) – London: The MIT Press.

DE SOUSA, R. (2002): Emotional Truth. *Proceedings of the Aristotelian Society.* Suppl. Vol. 76, 247-764.

DE SOUSA, R. (2004): Emotion – What I Know, What I'd Like to Think I Know, and What I'd Like to Think. In: Solomon, R. (ed.): *Thinking about Feeling*. Oxford – New York: Oxford University Press, 61-75.

DÖRING, S. (2003): Explaining Action by Emotion. *The Philosophical Quarterly* 53, No. 211, 214-230.

DRETSKE, F. (1981): *Knowledge and the Flow of Information.* Cambridge (Mass.): The MIT Press.

GOLDIE, P. (2000): *The Emotions: A Philosophical Exploration.* Oxford University Press.

GOLDIE, P. (2002): Emotions, Feelings and Intentionality. *Phenomenology and the Cognitive Sciences* 1, No. 3, 235-254.

GOLDIE, P. (2004): Emotion, Feeling and Knowledge of the World. In: Solomon, R (ed.): *Thinking about Feeling*. Oxford – New York: Oxford University Press, 91-106.

HUMBERSTON, L. L. (1992): Direction of Fit. *Mind* 101, No. 401, 59-83.

JAMES, S. (2003): Passion and Politics. In: Hatzimoysis, A. (ed.): *Philosophy and the Emotion*. Cambridge University Press, 221-234.

JAMES, W. (1897/1956): *The Will to Believe and Other Essays in Popular Philosophy*. New York: Dover Publications.

JAMES, W. (1902): *The Varieties of Religious Experience: A Study in Human Nature*. New York – Bombay: Longmans, Green and Co.

KENNY, A. (1963): *Action, Emotion and Will.* Routledge & Kegan Paul.

NEU, J. (1977): *Emotion, Thought, and Therapy: A Study of Hume and Spinoza and the Relationship of Philosophical Theories of the Emotions to Psychological Theories of Therapy*. London: Routledge & Kegan Paul – Berkeley: University of California Press.

NUSSBAUM, M. C. (2001): *Upheavals of Thought.* Cambridge University Press.

PRINZ, J. (2003): Emotion, Psychosemantics, and Embodied Appraisals. In: Hatzimoysis, A. (ed.): *Philosophy and the Emotion*. Cambridge University Press, 69-86.

PRINZ, J. (2004a): Embodied Emotions. In: *Thinking about Feeling.* Oxford University Press.

PRINZ, J. (2004b): *Gut Reactions: A Perceptual Theory of Emotion*. Oxford University Press.

ROBERTS, R. (1988): What an Emotion Is: A Sketch. *Philosophical Review* 97, No. 2, 183-209.

ROBERTS, R. (1991): What is Wrong with Wicked Feeling? *American Philosophical Quarterly* 28, No. 1, 13-24.

ROBERTS, R. (2002): *Emotions: An Essay in Aid of Moral Psychology*. Cambridge University Press.
SALMELA, M. (2006): True Emotions. *The Philosophical Quarterly* 56, No. 224, 382-405.
SCHUELER, G. F. (1991): Pro-Attitudes and Direction of Fit. *Mind* 100, No. 400, 277-281.
SCHUELER, G. F. (1995): *Desire.* Cambridge (Mass.): The MIT.
SEARLE, J. (1983): *Intentionality: An Essay in the Philosophy of Mind.* Cambridge University Press.
SEARLE, J. (2001): *Rationality in Action.* The MIT Press.
SMITH, M. (1994): *The Moral Problem.* Oxford: Basil Blackwell.
SOBEL, D. – COPP, D. (2003): Against Direction of Fit Accounts of Belief and Desire. *Analysis* 61, No. 269, 44-53.
SOLOMON, R. C. (1976): *The Passions.* New York: Doubleday.
SOLOMON, R. C. (1980): Emotion and Choices. In: Rorty, A. (ed.): *Explaining Emotions.* Berkeley: University of California Press.
SOLOMON, R. C. (1993): *The Passions: Emotions and the Meaning of Life.* 2nd ed. Indianapolis: Hackett.
TAPPOLET, C. (2005): Ambivalent Emotions and the Perceptual Account of Emotions. *Analysis* 65, No. 287, 229-233.
TAYLOR, C. C. W. (1986): Emotions and Wants. In: Marks, J. (ed.): *The Ways of Desire.* Chicago: Precedent Publishing.
TAYLOR, G. (1975): Justifying the Emotions. *Mind* 84, 390-402.
TAYLOR, G. (1985): *Pride, Shame, and Guilt: Emotions of Self-Assessment*. Oxford: Oxford University Press.
TAYLOR, G. (1988): Envy and Jealousy: Emotions and Vices. *Midwest Studies in Philosophy* 13, No. 1, 233-249.
VELLEMAN, D. (1992): The Guise of the Good. *Noûs* 26, No. 1, 3-26.
WALTON, K. (1979): Fearing Fictions. *The Journal of Philosophy* 75, No. 1, 5-27.
YANG, S. (2009): Emotion, Intentionality and Appropriateness of Emotion. *Organon F* 16, No. 1, 82-104.
ZANGWILL, N. (1998): Direction of Fit and Normative Functionalism. *Philosophical Studies* 91, No. 2, 173-203.

Organon F 23 (1) 2016: 50-71

Is Harry Frankfurt's "Doctrine of Sufficiency" Sufficient?

HUN CHUNG

Department of Philosophy. University of Arizona
213 Social Sciences Building. 1145 E. South Campus Drive. Tuscon, AZ 85721-0027. USA
hunchung1980@gmail.com

RECEIVED: 15-07-2015 • ACCEPTED: 05-11-2015

ABSTRACT: In his article, "Equality as a Moral Ideal", Harry Frankfurt argues against economic egalitarianism and presents what he calls the "doctrine of sufficiency." According to the doctrine of sufficiency, what is morally important is not relative economic equality, but rather, whether somebody has enough, where "having enough" is a non-comparative standard of reasonable contentment that may differ from person to person given his/her aims and circumstances. The purpose of this paper is to show that Frankfurt's original arguments in support for his doctrine of sufficiency have critical problems that Frankfurt himself does not properly recognize. In the end, I will argue that in order to solve these problems the doctrine of sufficiency cannot help but to incorporate certain prioritarian commitments – commitments which many would view as implying economic egalitarianism. This is embarrassing for a doctrine whose raison d'être was mainly to defeat economic egalitarianism.

KEYWORDS: Egalitarianism – Frankfurt – prioritarianism – sufficiency – sufficientarianism.

1. Introduction

In "Equality as a Moral Ideal", Harry Frankfurt gives a very incisive criticism against "economic egalitarianism" understood as a doctrine that claims that there is moral value in *equality itself* (cf. Frankfurt 1987). A major part

of the article tries to show why it is a deep mistake for anybody to hold economic egalitarianism. Along the way, Frankfurt presents his own alternative doctrine to economic egalitarianism, which he calls "the doctrine of sufficiency". The doctrine of sufficiency claims that what is morally important is for people to have enough; where "having enough" is a non-comparative standard of reasonable contentment that may differ from person to person given his/her aims and circumstances. In his original paper, Frankfurt presents two main arguments in support for his doctrine of sufficiency. The purpose of this paper is to show that both of these arguments have critical problems that Frankfurt himself does not properly recognize.[1]

2. Utility thresholds and the condition of scarcity

We start with Frankfurt's own example: the size of population is *ten*, a person needs at least *five* units of resources in order to survive, and there are *forty* units of resources. If we are intending to save anybody from this situation, then it is necessary that some must receive *more than others*; in other words, if we are intending to save at least one person, then, *inequality is necessary*.

We can see that, in this situation, an equal distribution of resources based on economic egalitarianism results in the worst possible outcome; namely, that everybody dies. Frankfurt (1987, 30) claims that "Surely in this case it would be morally grotesque to insist upon equality!".

The reason why an equal distribution of resources results in an undesirable outcome in the above situation is mainly because of the existence of what Frankfurt calls "utility thresholds". Frankfurt's main use of the notion of "utility thresholds" was to object to "the principle of diminishing marginal utility". According to the principle of diminishing marginal utility, when somebody consumes a certain type of good, the marginal utility that that good brings to that person tends to diminish. This is mainly because people

[1] After the publication of Frankfurt's seminar paper, the doctrine of sufficiency has attracted both proponents (see Anderson 1999; Benbaji 2005; Crisp 2003; Frankfurt 1987; Frankfurt 1997; Huseby 2010) as well as opponents (see Casal 2007; Roemer 2004), and there exists a vast literature that has been developed afterwards. However, the main focus of this paper will be on Frankfurt's original arguments against economic egalitarianism contained in his original paper.

generally tend to get satiated and derive less satisfaction towards the same good when they consume it over and over again.

However, Frankfurt points out that not all goods follow this rule. This is because there are certain goods that actually bring *more utility* to the person *after sustained consumption* than at first. This is when proper appreciation of the good gradually develops only after being exposed to a series of repeated trials and experiences. Here, the repeated trials and experiences serve as what Frankfurt calls a "warming up" process (see Frankfurt 1987, 26). It is easy to find numerous examples of this sort of good in our ordinary life; classical music, fine art, art house movies, certain types of gourmet food all require constant effort and a development of a certain level of maturity in order for the person to give proper appreciation to it.

Frankfurt explains that, when the good in question is money, what corresponds to this "warming up process" is "saving". This usually happens when one is trying to buy a certain good that gives unrivaled satisfaction compared to other goods, but which is too expensive for one to purchase unless one saves up for it. In this case, the last dollar saved that completes the full price of that good would give a utility that is far greater than any other dollar saved up to that very point. Here, the last saved dollar permits what Frankfurt (1987, 27, 30) calls a *crossing of a "utility threshold"*. Generally speaking, whenever there is a utility threshold for a certain good, a non-continuous jump in utility gain occurs at that very crossing threshold point. This makes the total utility achieved by crossing the utility threshold much greater than the individual sum of the utility gained by each individual unit of the good taken separately. In this sense, there is an additional value that is attached to *the completion* of a utility threshold itself when a good happens to have one.[2]

We can see that, in the previous example, each individual has a utility threshold; the utility threshold is five units of resources. Five units of resources is the borderline which demarcates life and death. So, anybody having five units of resources acquires a gain in utility that is significantly larger than the utility a person achieves when the person has only four or less than four units of resources.

---

[2] So, Frankfurt plausibly explains that a complete collection of 20 different items has a greater utility than an incomplete collection of 20 items that include duplicates. Here, *completeness* of the collection *itself possesses utility*. See Frankfurt (1987, 27-28, footnote 12).

We can also see that the situation is depicted in a way that the total amount of available resources is *scarce*; there are forty units of resources and there are also ten people which each need five units of resources in order to survive. This means that there are not enough resources to put everybody above the utility threshold; more specifically, in this case, there are not enough resources to save everybody's life.

The main reason why an equal distribution of resources is problematic is because an equal distribution would generally tend to generate a *much fewer* number of individuals who are *above* the utility threshold when resources are scarce; in this case, an equal distribution would put *everybody below* the threshold. This consequence is hard to accept especially when being below the threshold means something grave, such as death, as it is the case in the above situation.

So, in order to save at least one person in the above situation, an unequal distribution of resources is necessary. Presumably, the most reasonable distribution in this case would be to save eight persons with the forty remaining units of resources by giving each of them the five requisite units for survival. Since it is practically feasible to save a total of eight persons in the situation, any alternative distribution that saves less than eight persons seems hardly morally defensible.[3] If a certain distribution principle saves less than eight people, then this is a strong reason to think that the distribution principle in question is not the correct one for this particular situation.

This is why economic egalitarianism seems morally problematic in this particular example; it kills everybody when it is perfectly possible to save eight. So, it seems, at first sight, that economic egalitarianism cannot be a reasonable solution to this example.[4] However, the crucial question is not whether or not economic egalitarianism is a reasonable solution, but whether or nor *Frankfurt's doctrine of sufficiency itself* can be a reasonable solution to this situation. Then, the crucial question is: Can the doctrine of sufficiency save eight persons?

---

[3] Anybody who proposes an alternative distribution that saves less than eight persons would have to justify why his/her preferred distribution would be better even though it kills more people than what is necessary. It seems very unlikely for anybody to be able to offer any cogent justification of this sort.

[4] The reason why I say "first sight" is that I will later suggest a possible way for economic egalitarianism to solve this problem.

Frankfurt almost seems to take it for granted that his doctrine of sufficiency will be able to save eight persons; in fact, his main purpose of presenting the above example in the first place was to show that his doctrine of sufficiency is superior to economic egalitarianism *in precisely this respect*. Frankfurt writes:

> Under conditions of scarcity, then, an egalitarian distribution may be morally unacceptable. Another response to scarcity is to distribute the available resources in such a way that as many people as possible have *enough*, or in other words, to *maximize the incidence of sufficiency*. This alternative is especially compelling when the amount of a scarce resource that constitutes enough coincides with the amount that is indispensable for avoiding some catastrophic harm – as in the example just considered, where falling below the threshold of enough food or enough medicine means death. (Frankfurt 1987, 31 emphasis added)

Here, Frankfurt implies that the distribution principle of 'distributing the available resources in such a way that as many people as possible have enough' or, in other words, the distribution principle of 'maximizing the incidence of sufficiency' would be able to save the maximum number of persons (in this case, eight) and that this would be the distribution principle that the "doctrine of sufficiency" would mandate. However, as it becomes evident in the later part of the article, this doesn't actually turn out to be the case. This is because the doctrine of sufficiency does *not* construe "having enough" as "having the amount of resources that is necessary to avoid some catastrophic harm (i.e. death)".

In section VII of the article, Frankfurt explicates quite clearly what the doctrine of sufficiency regards as "having enough". There, Frankfurt distinguishes between two senses of "having enough"; one is that a *limit* has been reached which implies that having more would be undesirable, and the other is that *a certain requirement or standard has been met* without any implication that a larger quantity would be bad. According to Frankfurt,

> In the doctrine of sufficiency the use of the notion of "enough" pertains to *meeting a standard* rather than to *reaching a limit*. To say that a person has enough money means that he is content, or that it is reasonable for him to be content, with having no more money than he has. And to say this is, in turn, to say something like the following: the person does not (or cannot reasonably) regard whatever (if anything) is unsatisfying or dis-

> tressing about his life as due to his having too little money. (Frankfurt 1987, 37)

So, to put it in a simplistic way, "having enough", according to the doctrine of sufficiency, means one has enough to be *reasonably content* with one's life while leaving open the possibility that one would be willing to have more if this were an option that could be achieved without paying significant costs. In other words, having enough is *compatible* with having a mild preference towards having more.

What "having enough" rules out, according to Frankfurt, is having *an active interest* to seek more money than what one already has (cf. Frankfurt 1987, 39). So, when one has enough, one might still be aware that one's situation could be made better than the way it is now; but one does not really care about whether or not one's situation could be improved since one is already quite content with the way things are right now. In other words, when one has enough, say, money, we can say that one finds his/her current monetary situation *reasonably satisfying*.

However, when can we say that somebody's contentment of his/her current state of affairs is *reasonable*? If "having enough" means "being reasonably content", then what does "being reasonably content" imply? Frankfurt provides a quite explicit answer:

> It is essential to understand that having enough money differs from merely having enough to get along or enough to make life marginally tolerable. People are not generally content with living on the brink. The point of the doctrine of sufficiency is not that the only morally important distributional consideration with respect to money is whether people have enough to avoid economic misery. *A person who might naturally and appropriately be said to have just barely enough does not, by the standard invoked in the doctrine of sufficiency, have enough at all.* (Frankfurt 1987, 38, emphasis added)

So, according to the doctrine of sufficiency, "having enough" doesn't mean "having barely enough to survive" since one *cannot be reasonably content* if one only has barely enough; as Frankfurt himself claims "people are not generally content with living on the brink".

However, we can see in the previous example that giving somebody five units of resources is *equivalent to* giving that person just *barely enough resources to survive*, since it is assumed that five units of resources is the

minimum amount for anybody to sustain life. This means that the doctrine of sufficiency *cannot* regard giving five units of resources to somebody in the previous example as giving that person "enough".

It seems, then, reasonable to suppose that each individual consists of *two separate utility thresholds*; (a) the threshold of basic survival and (b) the threshold of reasonable contentment. What the doctrine of sufficiency regards as "enough" or "sufficient" is when an individual has enough resources to cross the latter threshold *not* the former threshold. When one is merely provided with resources that enable one to barely cross the first utility threshold, then, according to Frankfurt, the person "does not, by the standard invoked in the doctrine of sufficiency, have enough at all."

This means that saving eight individuals by giving each of them five units of minimum resources in the previous example cannot be an instance of 'distributing the available resources in such a way that as many people as possible have enough' that is mandated by the doctrine of sufficiency. In fact, as long as the two utility thresholds of basic survival and reasonable contentment do not coincide (as it is the case in most practical situations), distributing five units of resources to eight people would actually be a way to *minimize* the incidence of sufficiency; the incidence of sufficiency, in this case, would be *zero*.

Let's illustrate this from a more concrete example. Consider the following situation

Frankfurt's Example of Scarcity Modified

1. Available Resources: 40 units
2. Individuals: Andy, Bob, Chad, Derk, Erin, Fred, Gil, Hun, Ion, Jay (total: 10 persons)
3. The Utility Thresholds of Each Individual

Individual's Name	Andy	Bob	Chad	Derk	Erin	Fred	Gil	Hun	Ion	Jay
Utility Thresholds										
a) Threshold of Survival	(a) 5	(a) 5	(a) 5	(a) 5	(a) 5	(a) 5	(a) 5	(a) 5	(a) 5	(a) 5
b) Threshold of Reasonable Contentment (= Threshold of Sufficiency)	(b) 6	(b) 6	(b) 6	(b) 7	(b) 7	(b) 8	(b) 9	(b)10	(b)11	(b)12

This is a modified version of Frankfurt's example of the situation of scarce resources. In this modified version, the utility thresholds of each of the ten individuals are revealed.

We can see that each individual has *two* separate utility thresholds; (a) the threshold of survival and (b) the threshold of reasonable contentment. The threshold for survival is the same for every individual; everybody needs at least five units of resources in order to maintain life, and this is so regardless of each individual's personal characteristics or temperaments.

However, the threshold for reasonable contentment differs from person to person. This is because the amount to which one may feel reasonably satisfied is partly a function of the individual's personality, as well as his/her own way of valuing things in life, which may have developed throughout the individual's course of life. Note that different people may have different thresholds of what they regard as "enough" satisfaction; some people might simply have what are known as "expensive tastes" which require a lot of resources to satisfy (e.g. *Jay* in the above example); others might have very modest tastes which can be quite easily satisfied even by a meager amount of resources (e.g. *Andy* in the above example.)

It is important to understand that recognizing that different people have different thresholds of reasonable contentment (or sufficiency) is an *integral part* of the doctrine of sufficiency. We can see this from the fact that one of the major criticisms that Frankfurt raises against economic egalitarianism is that economic egalitarianism is essentially *alienating* in the sense that it focuses primarily on the sheer size of how economic benefits are distributed without taking people's specific interests and needs into account. Frankfurt writes:

> A concern for economic equality, construed as desirable in itself, tends to divert a person's attention away from endeavoring to discover – within his experience of himself and of his life – what he himself really cares about and what will actually satisfy him, although this is the most basic and the most decisive task upon which an intelligent selection of economic goals depends. Exaggerating the moral importance of economic equality is harmful, in other words, because it is alienating. (Frankfurt 1987, 23)

Frankfurt emphasizes essentially the same point in a later paper, when he writes:

> Egalitarianism is harmful because it tends to distract those who are beguiled by it from their real interests. (…) The essential thing is not that he compares his situation with theirs, but that he understands his own needs. (…) What one person will require in order to serve his own most authentic interests effectively does not depend upon what another person has. His requirements may differ very considerably (…) from the requirements of individuals who are devoted to attaining goals that differ from his. (…) The erroneous assumption that equality is worth having for its own sake distracts people, in other words, from what is most essential. It leads them to become alienated from themselves. (Frankfurt 2000, 91-92)

So, we can say that one of the major merits that the doctrine of sufficiency has in relation to economic egalitarianism is that it respects, at the level of economic distribution, what is truly important to individuals by trying to provide what is sufficient for each specific person *given his/her specific circumstance, aims, and needs*. In other words, it is important to understand that it is an essential part of Frankfurt's doctrine of sufficiency that it recognizes that different people may have different thresholds of reasonable contentment assessed from his/her own specific circumstances, aims, and needs, and that it is this threshold of reasonable contentment – which differs from person to person – that the doctrine of sufficiency so emphatically urges to satisfy.

This means that when the doctrine of sufficiency tries to maximize the incidence of sufficiency, what it is trying to maximize is, *not* the number of people who has barely crossed the minimum level of subsistence, *but rather*, the number of people who is quite satisfied with his/her life; in other words, based on the example that I have provided above, *what the doctrine of sufficiency tries to maximize is the number of people who cross threshold (b) (= the threshold of reasonable contentment) not threshold (a) (= the threshold of survival)*.

So, when we apply the doctrine of sufficiency to the above example, the distribution that would maximize the incidence of sufficiency would be saving *Andy, Bob, Chad, Derk, Erin, Fred* by giving them each 6, 6, 6, 7, 7, 8 units of the 40 units of available resources. This is the best way to maximize the incidence of sufficiency given the available resources.

Here, we can see that we have managed to make 6 individuals quite content with their own lives while letting 4 individuals die of starvation. Since it was possible to save 8 individuals by distributing 5 units of resources to each

of them with the 40 units of available resources, what this shows is that the distribution required by the doctrine of sufficiency actually *kills 2 additional individuals* when it was perfectly possible to save these two people from dying.

Moreover, the reason why these two individuals had to die is mainly because, the doctrine of sufficiency, in an attempt to maximize the incidences of sufficiency (at the level of reasonable contentment), used up the remaining resources, which were left after saving the lives of 6 individuals, to further satisfy these individuals' *non-basic needs* in order to make them content with their lives to the extent that they *no longer have any active interest* to seek more, when it was perfectly possible to use these remaining resources to satisfy other people's *basic needs* and save two more lives! In other words, by following the doctrine of sufficiency, 2 additional lives had to be sacrificed in order to fully satisfy the non-basic needs of 6 individuals. In any case, this seems hardly morally justifiable.

Although it is true that the satisfaction of certain non-basic needs is important for somebody to lead a sufficiently satisfying life given his/her specific aims and interests, the importance of leading a sufficiently satisfying life cannot be compared to the importance of saving a human life itself. Therefore, whenever the satisfaction of non-basic needs and saving a human life conflict, it seems *prima facie* that the saving of a human life should always take precedence over the satisfaction of non-basic needs. The doctrine of sufficiency is flawed to the extent that it takes the precedence between the satisfaction of non-basic needs and saving a human life backwards; and we can see that this is the case in the above example.

It should be further noted that this flaw of the doctrine of sufficiency is not confined to such farfetched thought experiments. Whenever there is a scarcity of resources – where not everybody's basic needs can be fully met or where everybody's basic needs can be fully met but only barely – the doctrine of sufficiency would always recommend a distribution policy that would leave the basic needs of much more people left unsatisfied than what was practically achievable; in other words, whenever we apply the doctrine of sufficiency in circumstances of scarcity, there would always be cases where some people's basic needs are sacrificed for the sake of satisfying other people's non-basic needs. And this is a reason to think the doctrine of sufficiency is flawed in some morally important way.

Then, how might Frankfurt reply to this objection? There seems to be a number of moves that he can make. One move (as it is motivated by the

above example) is to allow the existence of *multiple* sufficiency thresholds that range from minimum basic subsistence to reasonable contentment to full contentment, and claim that the satisfaction of somebody's lower sufficiency threshold takes *lexicographic priority* over the satisfaction of another person's higher threshold. This is actually a move that Robert Huseby makes when he defends a dual-threshold version of the doctrine of sufficiency – a version that assumes a maximal (i.e. the threshold of reasonable contentment) and a minimal (i.e. the threshold of critical survival) sufficiency threshold – and claims,

> First, individuals below the maximal sufficiency threshold should have absolute priority over individuals above this threshold. (…) Between the minimal and maximal sufficiency thresholds, I propose that we should apply a constrained and inverse form of prioritarianism. (…) Second, strong priority should be given to those below the minimal sufficiency threshold. (Huseby 2010, 184-185)

This modification of the doctrine of sufficiency will indeed solve the problem presented in "Frankfurt's Example of Scarcity Modified" above and successfully save 8 people. However, Huseby's dual-threshold version of the doctrine of sufficiency faces a problem from which Frankfurt's original version of the doctrine of sufficiency is actually free: *the problem of wasted resources*.

Consider two individuals: call them individual 1 and individual 2. Suppose both individuals' minimal sufficiency thresholds (i.e. the threshold of critical survival) and maximal sufficiency thresholds (i.e. the threshold of reasonable contentment) are respectively 5 and 7 units of resources. Suppose that individual 1 already owns 7 units of resources (i.e. he/she meets his/her maximal sufficiency threshold) while individual 2 owns nothing. Suppose that we have 1 additional unit of resource to distribute. Note that there is no way to redistribute the total amount (i.e. 8 units) of resources that could save both individuals.

Huseby's dual-threshold version of the doctrine of sufficiency, by giving *absolute priority* to individuals below the maximal threshold over individuals above it, *requires* us to distribute the additional unit of resource to individual 2. As a result, individual 2 dies and the resource is wasted. In other words, Huesby's dual-threshold version of the doctrine of sufficiency *morally requires us to waste resources* in this situation. Note that Frankfurt's original

doctrine of sufficiency does not suffer from the same problem, as it gives no priority to distributing the remaining resources to the worse-off unless doing so raises that person above the threshold of reasonable contentment.[5] So, modifying Frankfurt's doctrine of sufficiency in a way that takes account of multiple sufficiency thresholds will not solve the problem without cost.

Another move that Frankfurt may make to solve our imminent problem is to restrict the doctrine of sufficiency with what may be called the "Scanlonian Proviso".[6] According to Scanlon, an action X in circumstances C is wrong if and only if "any principle that permitted one to do X in those circumstances could, for that reason, reasonably be rejected" (Scanlon 1998, 95). In light of this, one may restrict Frankfurt's doctrine of sufficiency as follows:

> Frankfurt's Doctrine of Sufficiency with Scanlonian Proviso
> Maximize the incidence of sufficiency – interpreted as reasonable contentment – unless doing so would be disallowed by a principle that one may not reasonably reject.

A general principle that states, "Do not kill more human lives than what is absolutely necessary", seems to be one such principle that one cannot reasonably reject. If we go back to "Frankfurt's Example of Scarcity Modified", now, with the Scanlonian proviso operating, Frankfurt's doctrine of sufficiency so restricted will no longer distribute the available resources in a way that saves only 6 rather than 8 people. This is because doing so will not be allowed by the general principle, "Do not kill more human lives than what is absolutely necessary", a principle that one may not reasonably reject.

This solves the problem. But, note that exactly the same move is available to the economic egalitarian as well. That is, the economic egalitarian may avoid the same problem by restricting the egalitarian principle with the Scanlonian proviso in exactly the same way. Consider:

> Economic Egalitarianism with Scanlonian Proviso
> Distribute economic resources equally across individuals unless doing so would be disallowed by a principle that one may not reasonably reject.

---

5 This will become more apparent a little bit later.

6 I thank an anonymous referee for suggesting this for me.

Restricted by the Scalonian proviso, economic egalitarianism will no longer distribute the available material resources equally (and, thereby, kill everybody) as doing so will be disallowed by the same general principle as above, which states "Do not kill more human lives than what is absolutely necessary", a principle that nobody can reasonably reject.

What all this shows is that Frankfurt's doctrine of sufficiency may solve the problem in "Frankfurt's Example of Scarcity Modified" only by either (a) facing a new problem (i.e. the problem of wasted resources) or (b) by making a move that is also readily available for the economic egalitarian to use. In short, unlike what Frankfurt thinks, examples of scarcity do not give any particularly strong reasons to favor doctrine of sufficiency over economic egalitarianism.

3. The doctrine of sufficiency and urgent needs

Frankfurt emphasizes that economic egalitarianism and the doctrine of sufficiency are *logically independent*; an equal distribution can entirely lack sufficiency, and a distribution that satisfies sufficiency can be quite unequal. So, considerations that support one stance cannot be presumed to support the other. However, Frankfurt argues that many proponents of economic egalitarianism provide grounds that actually support only the doctrine of sufficiency and mistakenly think that they have provided grounds for economic egalitarianism.

One common way to argue for economic egalitarianism is to contrast the abject situation of the absolute poor and the situation of the rich. Frankfurt agrees that it is true that the fact that there are people who are suffering from abysmal poverty is itself a situation that is morally undesirable which calls for rectification. And in order to rectify the situation of the absolute poor it might even be necessary to redistribute the surplus resources of the rich and give it to the poor. Undoubtedly, this would make the resulting distribution more equal. However, according to Frankfurt, the mere fact that abysmal poverty is morally undesirable does not entail that there is something wrong with inequality itself, nor does the fact that improving the situation of the poor calls for a more equal distribution entail that equality itself was what we were aiming for when we tried to improve the situation of the poor. Rather, Frankfurt claims that the main reason why absolute poverty is morally objec-

tionable is not because people suffering from absolute poverty have *less*, but because they do not have *enough*.

So, according to Frankfurt, when a proponent of economic egalitarianism argues that a more equal distribution is needed in order to improve the situation of the absolute poor, what he/she is really arguing for is not economic egalitarianism but rather the doctrine of sufficiency (see Frankfurt 1987, 33-34). This conclusion is reinforced when we see that economic egalitarians are not usually troubled by the significant inequality that exists between the rich and the upper middle class.

Another typical confusion of the economic egalitarian, according to Frankfurt, is their ungrounded assumption that the worse-off person always has more *urgent needs* that are unmet than the better-off person. However, according to Frankfurt, this is not true; not only can the relatively worse-off person not have any urgent needs that are unsatisfied, but it might even be the case that the situation of the worse-off person is actually quite good. And if this were to be the case, Frankfurt claims that, a worse-off person could reasonably accept his/her current situation without presuming that any other distributive situation would make him or her worse (see Frankfurt 1987, 36).[7]

Frankfurt presents Nagel as a typical economic egalitarian who relies on the mistaken assumption that the worse-off person always has more urgent needs that are unmet than the better-off person in order to argue for the moral appeal of equality.

> Nagel illustrates his thesis concerning the moral appeal of equality by considering a family with two children, one of whom is "normal and quite happy" while the other "suffers from a painful handicap." If this family were to move to the city the handicapped child would benefit from medical and educational opportunities that are unavailable in the suburbs, but the healthy child would have less fun. If the family were to move to the suburbs, on the other hand, the handicapped child would be deprived but

[7] Here, we can see that Frankfurt is implicitly attacking the acceptability condition that Rawls' "difference principle" ultimately relies on; according to Rawls, any unequal distribution must be able to be reasonably accepted even by the worst-off person of that specific distribution, and we cannot reasonably expect the worst-off person to accept an unequal distribution unless it can be shown that any other alternative distribution would make his or situation even worse.

> the healthy child would enjoy himself more. Nagel stipulates that the gain to the healthy child in moving to the suburbs would be greater than the gain to the handicapped child in moving to the city; in the city the healthy child would find life positively disagreeable, while the handicapped child would not become happy "but only less miserable." (Frankfurt 1987, 36)

According to Frankfurt, Nagel claims that the egalitarian decision in this situation would be to move to the city. And the reason that Nagel provides in order to support this egalitarian decision is that, although it is true that the healthy child would benefit much more by moving to the suburb, the handicapped child has a much *more urgent need* that needs to be satisfied due to his being in a *worse-off position*.

However, according to Frankfurt, this is a mistaken analysis. Frankfurt agrees that the handicapped child has a much more urgent need that needs to be satisfied. This is so, even if the benefit that the healthy child would receive by the family moving to the suburb would be much greater. However, the main reason why the handicapped child has a much more urgent need that must take priority is not simply because the handicapped child is *worse-off* than the healthy child; it is rather because the condition that the handicapped child is suffering is *significantly bad* viewed from an absolute scale. Therefore, Frankfurt claims,

> …the most cogent basis for Nagel's judgment in favor of the handicapped child has nothing to do with the alleged urgency of providing people with as much as others. It pertains rather to the urgency of the needs of people who do not have enough. (Frankfurt 1987, 37)

In other words, according to Frankfurt, our moral intuition that tells us that the family should move to the city rather than the suburbs *supports the doctrine of sufficiency rather than economic egalitarianism*.

However, again I am not sure whether the doctrine of sufficiency can really explain why the family should move to the city in order to satisfy the more urgent need of the handicapped child. Suppose for the sake of argument that the morally right answer to Nagel's example is for the family to move to the city. Remember that the doctrine of sufficiency aims to maximize the incidences of sufficiency where an incidence of sufficiency is achieved when one is provided with enough resources that enable him or her to cross the threshold of "reasonable contentment".

In the above example, the healthy child is described to be "normal and quite happy". From this we can infer that the healthy child is already above the utility threshold of "reasonable contentment". Again, the fact that the healthy child is above the threshold of reasonable contentment doesn't imply that giving more to the healthy child would not give him more satisfaction.

Then, what about the handicapped child? The handicapped child is described as "suffering from a painful handicap". From this it is reasonable to infer that the handicapped child is below the utility threshold of reasonable contentment. However, we can say that the handicapped child is still above the threshold of survival; this is because the child's handicap, although severe, is not something that threatens the child's life. This means that the handicapped child's utility level would be somewhere *in between* the threshold of survival and the threshold of reasonable contentment; presumably, the handicapped child's utility level is much closer to the threshold of survival than it is to the threshold of reasonable contentment considering that the child is suffering from a, not just ordinary, but a "painful" handicap.

Now, in order to see which decision (between moving to the city and moving to the suburbs) that the doctrine of sufficiency would support, we would need to see which decision would actually maximize the incidences of sufficiency, where sufficiency is measured by whether or not somebody's utility level is above the threshold of reasonable contentment.

If the family moves to the city, then the utility level of the healthy child drops below the threshold of reasonable contentment. This can be inferred from the fact that, by assumption, the healthy child would find life *positively disagreeable* in the city. If a "normal and quite happy" child starts to find his life "positively disagreeable", then this suggests that the child is no longer reasonably content with his/her life. So, by moving to the city, the healthy child drops below the threshold of reasonable contentment.

Then, what happens to the handicapped child if the family moved to the city? The utility level of the handicapped child would undoubtedly go up; but, based on the assumption of the story, *only marginally*. This can be inferred from the fact that if the family moved to the city, then the handicapped child would "not become happy but only less miserable." In other words, although the utility level of the handicapped child would go up, it would still be below the threshold of reasonable contentment. In short, even if the family moved to the city, the handicapped child would not achieve "sufficiency".

Now, suppose that the family moved to the suburbs. How would this make the situation turn out? In terms of the healthy child, moving to the suburbs would make the already quite happy child much happier. So, the utility level of the healthy child would still be above the threshold of reasonable contentment but by a wider margin. In terms of the handicapped child, moving to the suburbs would definitely not increase the child's utility level, but it would, nonetheless, not drop the utility below the threshold of survival; the handicapped child would not die even if the family moved to the suburbs.

So, here is the summary of the situation: if the family moves to the city, the total incidence of sufficiency drops from 1 to 0. If the family moves to the suburbs, then the total incidence of sufficiency remains the same, which is 1. Therefore, in order to maximize the incidence of sufficiency, the family should move to the suburbs. In other words, it turns out that the doctrine of sufficiency supports the (non-egalitarian) decision to move to the suburbs!

Let's modify the situation in certain ways so that the doctrine of sufficiency would generate the desired answer for the situation at hand; which is to support the family's decision to move to the city.

First, what if we assume that the utility level of the healthy child would not drop below the threshold of reasonable contentment even if the family moved to the city; the healthy child would unquestionable find less enjoyment than he/she would have if the family moved to the suburbs, but suppose that the child would still find his/her life quite satisfying in the city as well. In this case, the incidences of sufficiency would be the same (i.e. 1) regardless of whether the family moved to the city or whether the family moved to the suburbs. And if this were the case, wouldn't the doctrine of sufficiency support the family's decision to move to the city by considering the desperate situation of the handicapped child as the "tie breaker"?

Not necessarily. In order to see that the doctrine of sufficiency would not necessarily support the family's decision to move to the city even in this situation, we would need to see what Frankfurt claims about the distribution of additional resources that would not contribute to increasing the incidences of sufficiency.

Suppose, in Frankfurt's original example of scarce resources, the available resources were 41 units instead of 40 units. In this case, one unit of resource would still be left after saving eight people by giving them each five

units of resources which is required for minimum survival. How should this one extra unit distributed? According to Frankfurt, it doesn't automatically follow from such situation that the remaining one extra unit should be given to one of the two persons who haven't been allocated with any resources. According to Frankfurt, this is because,

> …one additional unit of the resource in question will not improve the condition of a person who has none. By hypothesis, that person will die even with the additional unit. What he needs is not one unit but five. It cannot be taken for granted that a person who has a certain amount of a vital resource is necessarily better off than a person who has a lesser amount, for the larger amount may still be too small to serve any useful purpose. (Frankfurt 1987, 31)

Frankfurt adds,

> Those below a utility threshold are not necessarily benefited by additional resources that move them closer to the threshold. What is crucial for them is to attain the threshold. Merely moving closer to it either may fail to help them or may be disadvantageous. (Frankfurt 1987, 32)

From this Frankfurt concludes,

> It may be morally quite acceptable, accordingly, for some to have more than enough of a certain resource even while others have less than enough. (Frankfurt 1987, 32)

This has very important implications for our current purposes, since we can think of the family's decision on whether they should move to the city or whether they should move to the suburbs as *essentially the same* as the decision on how we should distribute the one extra unit of resource in our first example.

Regardless of which decision the family makes, the healthy child would always be above the threshold of reasonable contentment by assumption. However, the family's moving to the suburb would make the healthy child's life much more satisfying. In this way, the family's moving to the suburb would be analogous to the decision of giving the extra unit of resource to one of the eight persons who already received enough resources to live; the family's moving to the suburb would be giving the healthy child *more than enough*, when his sibling, the handicapped child, has *less than enough.*

By contrast, the family's decision of moving to the city would be analogous to the decision of giving the extra unit of resource to one of the two persons who have not received any resources. Just as giving the extra unit of resource to one of these two persons would only make the person move slightly closer to the threshold of survival without crossing it, the family's moving to the city would only let the handicapped child move slightly closer to the threshold of reasonable contentment without crossing that threshold; as it is assumed, the handicapped child would not become happy but only *less miserable* by moving to the city.

If this is the case, then we can see that the doctrine of sufficiency would not necessarily recommend the family to move to the city. In fact, we can see from Frankfurt previous remarks that the doctrine of sufficiency would actually regard the family's decision to move to the suburbs as *morally acceptable*.

This is because, although it is true that the family's moving to the city would improve the situation of the handicapped child, the improvement that the handicapped child would be too small for the doctrine of sufficiency to regard as morally significant. According to the doctrine of sufficiency, improvements are morally significant only when they make people cross thresholds of reasonable contentment. Since moving to the city does not make the handicapped child cross the threshold of reasonable contentment, it is perfectly permissible (or it might even be preferable), according to the doctrine of sufficiency, for the family to move to the suburbs. In short, even in this modified situation, the doctrine of sufficiency does not recommend the family to move to the city.

If we really want the doctrine of sufficiency to recommend the family to move to the city, it is necessary for us to modify *both* the situation as well as the doctrine of sufficiency itself.

First, the situation would have to be modified so that the handicapped child would *die* if the family moved to the suburbs and *live* if the family moved to the city. Second, we would need to make a move similar to that of Huseby and allow the existence of *multiple* sufficiency thresholds for which the satisfaction of lower sufficiency thresholds takes lexicographic priority over the satisfaction of higher sufficiency thresholds. That is, in order to allow the family to make the morally right decision to move to the city on the basis of the doctrine of sufficiency, the doctrine of sufficiency *itself* would have to be modified, perhaps, in the following way:

The Lexicographic Doctrine of Sufficiency
(1) Maximize the incidence of crossing sufficiency thresholds.
(2) Whenever the satisfaction of two or more utility thresholds conflict, the satisfaction of the lower sufficiency threshold takes absolute priority over the satisfaction of higher sufficiency threshold.

The lexicographic version of the doctrine of sufficiency would be able to recommend the family to move to the city in the modified scenario in which the handicapped child would *die* if the family moved otherwise.

This is because even if the healthy child's welfare will drop below the threshold of reasonable contentment by moving to the city, the modified version of the doctrine of sufficiency would still recommend the family to move to the city as such sacrifice is necessary to move the handicapped child above the critical threshold of survival – a threshold which takes lexicographic priority over the threshold of reasonable contentment of the healthy child. Also, unlike Frankfurt's original doctrine of sufficiency, the lexicographic version would be able to save eight individuals in the first example of scarce resources. However, this is all because, unlike the doctrine of sufficiency, there is a certain *prioritarian element* embedded in the lexicographic version of the doctrine of sufficiency; namely, that, in adjudicating which sufficiency threshold to satisfy, the threshold of the worse off person takes priority over that of the better off person.[8]

However, it should be noted that even the lexicographic version of the doctrine of sufficiency would not be able to recommend the family to move to the city in Nagel's *original example,* in which the handicapped child would not die regardless of where the family moved; this is because, in such case, there will be no conflict between satisfying either the lower sufficiency threshold of the handicapped child or the higher sufficiency threshold of the normal child, as the handicapped child will, in either option, remain *in between* the threshold of survival and the threshold of reasonable contentment. As the healthy child will be above the threshold of reasonable contentment in the suburbs while he/she will fall below such threshold in the city, the fact that the handicapped child will remain in between the thresholds of survival and reasonable contentment in either option mandates that even the lexicographic version of the doctrine of sufficiency will morally require the family to move to the suburbs.

---

[8] For a defense of the priority view, see Parfit (1997; 2002).

In order for the lexicographic doctrine of sufficiency to recommend the family to move to the city in Nagel's original example, it must give up Frankfurt's basic stance concerning the distribution of additional resources that does not contribute to achieving any additional sufficiency thresholds; in other words, the lexicographic doctrine of sufficiency would have to claim that after maximizing the incidences of crossing sufficiency thresholds (in order of lowest to highest), any additional resources that are left should be distributed to the worse-off person even if this does not contribute to making that person cross any additional sufficiency thresholds.

However, by doing so, we can see that, now, not only does the doctrine of sufficiency have some prioritarian elements, but it has virtually collapsed into prioritarianism, which claims that "benefits to the worse off should be given more weight" even when doing so does not result in crossing of any additional sufficiency thresholds (cf. Parfit 1997, 213). Of course, prioritarianism is not exactly the same as economic egalitarianism; unlike economic egalitarianism, prioritarianism does not *aim at* achieving economic equality *per se*. However, it is clear that, by giving more moral weight to the worse off (independent of whether such measure results increases the number of crossing sufficiency thresholds), in any given problem of economic distribution, prioritarianism has a *built-in bias towards equality*. This is why some authors have characterized prioritarianism as a doctrine that is "derivatively (if not, directly) egalitarian" (cf. Benbaji 2005, 312). In this sense, prioritarianism is a member of a broadly egalitarian family of views. This means that if we do think that it is morally preferable for the family to move to the city for the sake of the handicapped child in Nagel's original example, then it seems that this cannot be properly explained without invoking some basic intuitions – more specifically, prioritarian intuitions – that naturally support economic egalitarianism. This is, indeed, a very embarrassing result for the doctrine of sufficiency whose raison d'être was mainly to defeat economic egalitarianism.

References

ANDERSON, E. (1999): What is the Point of Equality? *Ethics* 109, No. 2, 287-337.

ARNESON, R. (2013): Egalitarianism. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy.* Summer 2013 Edition, available at: http://plato.stanford.edu/archives/sum2013/entries/egalitarianism/.

BENBAJI, Y. (2005): The Doctrine of Sufficiency: A Defense. *Utilitas* 17, 310-332.
CASAL, P. (2007): Why Sufficiency is Not Enough. *Ethics* 117, 296-326.
CRISP, R. (2003): Equality, Priority, and Compassion. *Ethics* 113, No. 14, 745-763.
FRANKFURT, H. (1987): Equality as a Moral Ideal. *Ethics* 98, No. 1, 1-13.
FRANKFURT, H. (1988): Necessity and Desire. *Philosophy and Phenomenological Research* 45, No. 1, 21-43.
FRANKFURT, H. (1997): Equality and Respect. *Social Research* 64, 3-15.
FRANKFURT, H. (2000): The Moral Irrelevance of Equality. *Public Affairs Quarterly*, Vol. 14, No. 2, 87-103.
HUSEBY, R. (2010): Sufficiency: Restated and Defended. *The Journal of Political Philosophy* 18, No. 2, 178-197.
PARFIT, D. (1997): Equality and Priority. *Ratio* 10, No. 3, 202-221.
PARFIT, D. (2012): Another Defense of the Priority View. *Utilitas* 24, Special Issue 3, 399-440.
RAWLS, J. (1971): *A Theory of Justice.* Cambridge: Belknap Press.
SCANLON, T. M. (1998): *What We Owe to Each Other.* Harvard: Belknap.

Organon F 23 (1) 2016: 72-82

Evidence vs. Virtues in Epistemic Disagreement

MARINA BAKALOVA

Institute for the Study of Societies and Knowledge. Bulgarian Academy of Sciences
Moskovska Str., 13A. 1000 Sofia. Bulgaria
marina.bakalova@gmail.com

RECEIVED: 21-10-2014 • ACCEPTED: 16-07-2015

ABSTRACT: In situations of peer disagreement there are two kinds of factors that matter. These are the factors internal to the discussion, such as evidence exposed and arguments presented by both sides and there are also factors external to the discussion, also called "independent factors". The external factors include mainly virtues and competences of the participants. There are two main theories about epistemic disagreement, "the steadfast view" and "the conciliationism", and each of them stresses the importance of one group of these factors over the other. This paper is a defense of the greater epistemic significance of independent factors over internal factors. However, it is not a defense of the conciliationism which takes independent factors to be systematically the ultimate arbiter in situations of peer disagreement. The argument in the paper goes like this. Although the steadfast view receives strong intuitive support from two cases presented by Thomas Kelly: "Right and Wrong" and "Wrong and Wronger", I argue that the view is undermined by Timothy Williamson's recent "Very Improbable Knowing" argument. This argument shows that for some basic type of evidence E when S uses it in favor of p, it is very improbable that S knows that S knows that p. Therefore, in situations of peer disagreement, S is unjustified to push her evidence in support of her side. There are arguably some exceptions, e.g. when one claims to have knowledge based on a priori evidence and on holistic evidence, but these are not sufficient to save the day for the steadfast view. In contrast to that, the reflective knowledge of one's first order competences and virtues (i.e. external factors) is not vulnerable by Williamson's argument. One reason for that is because we know about independent factors on the basis of holistic evidence. I claim that our epistemic goal in the face of peer disagreement is to end up on the side that is non-accidentally closer to truth. In accordance with achieving this goal, it is safer to stick to independent factors in resolving peer disagreement situations than to

follow one's nose concerning first-order evidence disclosed by the opponents. This might seem a counterintuitive result, which makes it worthy of further discussion.

KEYWORDS: Disagreement – evidence – knowledge attribution – social epistemology – virtue epistemology.

Disagreements, overt or hidden, are widely spread. I may disagree with you about lifestyle, scientific presuppositions, marketing strategies, moral considerations and religious beliefs, whether we should take the right or the left path on our mountain trip, etc. However, some possible disagreements are strange and inapt. For instance, normally it would be inadequate to disagree whether Ana is here with us in the room or not, or whether 1+1=2, or for that matter to disagree about the directions that a sober and competent passerby gives you on the street when you are lost and need them.

This article is about the epistemological problem of disagreement. The problem arises from the following sort of situations. Imagine that two people, A and B, who are equally competent in *p*-relevant domain and share nearly the same evidence for the issue at hand, overtly disagree about *p*. In the most extreme case A would believe that *p* and B would believe that non-*p*. Is rational disagreement in such cases possible and how an awareness of disagreement should affect the beliefs of each side?

Much of the contemporary debate on disagreement, I think, draws from the insight that a peer's opinion could boost or lower our confidence in believing something. Take the famous example by David Christensen:

> *Mental Math*: After a nice restaurant meal, my friend and I decide to tip 20% and split the check, rounding up to the nearest dollar. As we have done many times, we do the math in our heads. We have long and equally good track records at this (in the cases where we've disagreed, checking with a calculator has shown us right equally frequently); and I have no reason (such as those involving alertness or tiredness, or differential consumption of coffee or wine) for suspecting one of us to be especially good, or bad, at the current reasoning task. I come up with $43; but then my friend announces that she got $45. (Christensen 2011, 2)

Mental Math shows that if both I and my reliable friend got $43, this should make me more convinced in my answer. But if I got $43 and my

friend got $45 – this should make me less confident in my answer. This intuitively shows that psychological reports about others (assuming that they sincerely express their minds) serve as a kind of epistemic resource or evidence. On the other hand, psychological reports about me almost never serve as such a resource. I do not take into account the fact that I believe that *p* as further evidence in favor of *p*.

One explanation of this asymmetry is that, since every person believes for reason, I assume that under certain conditions my opponent's belief expresses reasons bearing upon the question at hand which I do not possess. This suspicion is based on my knowledge of my opponent's competences or epistemic virtues. Some people *aim* at forming their beliefs upon correct reasons on regular basis. When a person is epistemically virtuous, that person in normal circumstances is trustworthy. When I know that my opponent is reliable and trustworthy, especially on the question at hand, it is reasonable for me to assume that her belief expresses reasons. Since my reasons are not always reflectively accessible, I do not have to expect that my opponent will be able to formulate and utter her precise reasons outright. I just take her belief to be a hallmark of reasons that could, in case of disagreement, defeat my side.

This paper will focus on factors that should be taken into account when we make a decision of how to behave in the face of disagreement. There are two groups of relevant factors in disagreement situations: factors internal to the discussion, such as disclosed evidence and arguments of the two opponents, and factors external to the discussion, such as their general reliability and virtues. The two groups of factors are weakly related to each other. What is going on in the concrete case is only weakly determined by our competences, but is not entailed by them. This is so, because one can be very competent but wrong in the concrete case if one's claimed opinion is based on misleading evidence.

My aim here is to show that in relation to achieving our cognitive goal in situations of peer disagreement external factors are surprisingly more reliable to take into account than internal ones.

1. Epistemic peerhood

The traditionally used notion of epistemic peerhood is the first thing that matches the distinction between internal and external factors and shows their

importance in the debate of epistemic disagreement. It is generally agreed that an epistemic peer of S is somebody who is roughly symmetrical with S in certain relevant aspects. Here is a more precise definition.

> *Epistemic Peerhood:* A and B are **epistemic peers** regarding *p* if they are roughly symmetrical with each other in certain aspects related to *p*.

Which are the relevant aspects of the required symmetry? Two main aspects are most often taken into account. These are cognitive and evidential equality:

> *Cognitive equality:* A and B are **cognitive equals** regarding *p* if A and B have equal competence, expertise, or virtues in the *p*-relevant domain.
>
> *Evidential equality:* A and B are **evidential equals** regarding p if A and B have equal grasp of p-relevant evidence.

We acquire knowledge about each of these two aspects in different ways. On the one hand, we get to know cognitive equality through second-order considerations concerning track record of our opponent, her reputation, her behavior in the debate etc. These are what we initially called "external factors" or "independent reasons". On the other hand, for judging the relevance of our opponent's exposed evidence and the quality of her arguments based on that evidence (what we initially called "internal factors"), we use our abilities to gather, select and interpret evidence for building up arguments. We can sometimes reason from cognitive to evidential equality. For example, we can reason from virtues to quality of one's arguments, but we cannot reason from virtue to correctness of her evidence. In other words, if my opponent is very competent, I can safely conclude that she cannot end up with a clumsy argument, but I cannot conclude that she has not been using misleading evidence in forming her argument, so I cannot ultimately judge about the quality of her evidence from her virtues. So, the two components are only partially related.

2. Theories of epistemic disagreement

There are two main theories that provide answers to the question what ought one to do, epistemically speaking, in the face of disagreement. They

are standardly called: "consiliationism" and "the steadfast view". Each of them stresses on either internal or external factors in dispense of the opposite group of factors as normatively determining our epistemic behavior in the face of peer disagreement.

Conciliationism is the view that peer disagreement has a significant epistemic bearing, and we should revise our beliefs in every case of peer disagreement. The strongest version of concliationism is the so called *equal weight view*, according to which when two peers find themselves disagreeing with each other they should split the difference in half, absent reasons that are independent of the debate to do otherwise. The trigger for this view is avoiding begging the question against your opponent. More precisely, the view does so by being committed to the principle of independence:

> *Principle of Independence*: In evaluating the epistemic credentials of another's expressed belief about *p*, in order to determine how (or whether) to modify my own belief about *p*, I should do so in a way that doesn't rely on the reasoning behind my initial belief about *p*, which is independent of the very disagreement between the parties. (Christensen 2009, 758)

What these independent reasons amount to is not always explicitly mentioned in the literature. It seems that these are higher-order considerations such as the fact that my opponent believes that *p*, her epistemic virtues (competence, expertise), and no indications that she is drunk or under the influence of drugs, that she is joking, or that she is not sincere etc. Note that these independent reasons do not concern an assessment of the evidence or arguments of my opponent since they are a part of the disagreement. I take the central one of them to be competence or virtue and I will hereafter refer to virtues or competences only when I talk about external or independent factors.

Although conciliationism accounts for our warning behavior in the case of Mental Math, there is one very intuitive problem with the view. This is the problem that in most cases one side of the disagreement is closer to truth than the other.[1] Hence in the default case and therefore, in general, it is highly irra-

[1] Of course, we can imagine a version of the Mental Math example in which one side of the disagreement comes to the result $43 and the other side comes to the result $47, while the true result is $45. In such case none of the sides is closer to truth than the other. But such cases would be surprisingly accidental and we are justified in tak-

tional for that side to revise her belief by splitting the difference. The problem is well formulated by Christensen in two examples taken from Thomas Kelly:

> *Right and Wrong*: Right and Wrong are mutually acknowledged peers considering whether P. At t_0, Right forms 0.2 credence in P, and Wrong forms a 0.8 credence in P. The evidence available to both of them actually supports a 0.2 credence in P. Right and Wrong then compare notes, and realize they disagree. (Christensen 2011, 2)

If Right and Wrong are to split the difference, each must end up with credence 0.5. But this is counterintuitive.

The other example given by Kelly is the following:

> *Wrong and Wronger*: Wrong and Wronger are mutually acknowledged peers considering whether P. At t_0, Wrong forms 0.7 credence in P, and Wronger forms a 0.9 credence in P. The evidence available to both of them actually supports a 0.2 credence in P. Wrong and Wronger then compare notes, and realize they disagree. They follow the dictates of Equal Weight, and at t_1 they compromise at 0.8. (Kelly 2010, 3)

The problem in this second example is that by compromising according to the dictates of equal weight view Wronger made his belief more rational. But as Kelly rightly points out, "it is dubious that rational belief is so easy to come by" (Kelly 2010, 126).

Another argument against conciliationism in general is that in certain cases it does not make sense to revise our belief(s) in the face of disagreement because of the compelling justification that favors our side. Ernest Sosa provides four kinds of reasons that, when involved in justification of our beliefs, could make them resistant to opposition (see Sosa 2010). First, these are *a priori* reasons; second, these are phenomenal reasons concerning my own mental states such as "I have a headache"; third, these are holistic reasons such as Moore's total evidence against the skeptic; and finally reliance on one's epistemic community can serve as such a compelling reason. Arguably some of these reasons entail knowledge such as phenomenal reasons, and a priori reasons. This suggests that there are exceptions from the conciliationist rule in which it would be clearly irrational to follow it.

---

ing them as exceptional. The default cases are not like that because the actual probability of equal distance from truth is very low.

On the other camp, the steadfast theorists suggest that external factors are not of much epistemic significance and that we should rather focus on improving our own beliefs when faced with disagreement. For this we should focus on factors internal to the discussion such as evidence and arguments. Kelly writes: "Once I thoroughly scrutinize the evidence about *p*, the fact of someone else's disagreement cannot undermine the rationality of my belief" (Kelly 2005, 192).

The steadfast view is supported by the frequent epistemic asymmetry of the opponents revealed by the examples of Right and Wrong and Wrong and Wronger, and also by the privileged access to one's own mental states. Lackey calls the referent of privileged access "personal information". On the one hand, I have direct access to the grounds of my belief, like for instance vivid phenomenological experience, and on the other hand, I have personal information about the normal functioning of my own cognitive faculties, while lacking such information about my opponent. In certain cases, as Lackey argues (for instance in Sosa's cases mentioned above), personal information can provide a symmetry breaker in favor of my side (see Lackey 2010).

Besides, there is no problem with the steadfast view that mirrors that of conciliationism. While conciliationism loses plausibility by certain cases of disagreement in which it is more rational to stick to our guns, the fact of occasional belief-revisions does not pose a problem for the steadfast theory. The view does not explicitly forbid belief-revisions, at least when I realize an imperfection in my argument as a result of peer disagreement. The steadfast view allows for belief revision in the face of opponent's arguments, since these are not independent of the discussion. The Mental Math example is of this sort – it provokes the thought that my own argument could be improved as a result of double checking.

At this point it seems that the steadfast view has some advantages. I am now going to start with my original discussion which is supposed to reveal a strong objection to the view.

3. Our goal in the face of disagreement

So far, we saw some pluses and minuses of the theories favoring either internal or external factors. Now I want to focus more closely on the epis-

temic bearing that each of these factors has in resolving cases of peer disagreement.

As a starting point, though, we need to have a clear grasp of what our goal is in the face of peer disagreement. What do we want to achieve, epistemically speaking, in the face of disagreement? Some philosophers think that our natural goal in such situations is to acquire knowledge. John Hawthorne and Amia Srinivasan propose and discuss a knowledge based norm for disagreement, which they call "Knowledge Disagreement Norm" (KDN). According to KDN, in a case of disagreement about whether *p*, where S believes that *p* and H believes that not-*p*:

> KDN: (i) S ought to trust H and believe that not-*p* iff were S to trust H, this would result in S's knowing not-*p*
> (ii) S ought to dismiss H and continue to believe that *p* iff were S to stick to her guns this would result in S's knowing *p*, and (iii) in all other cases, S ought to suspend judgment about whether *p*. (Hawthorne – Srinivasan 2013, 11-12)

It is certainly desirable to be on the knowing side. However this demand seems to be too strong. We can acquire knowledge in the way described by KDN only if we stick to the side which is correct. But there are cases of disagreement in which A has got it wrong and B even wronger. Sometimes it is epistemically better to end up on the side which is closer to truth and yet in absence of knowledge than sticking to the wronger side. It is epistemically better if one's progress towards the truth is non-accidental. But KDN cannot grasp the kind of progress in such occasions; it works only for cases in which one side has got it right and the other wrong. For the general case though, I think, the goal of disagreement should be defined differently from KDN, namely as getting closer to truth in a non-accidental way. Call this "Closer to Truth Disagreement Norm" or CTDN for short.

CTDN: (i) S ought to trust H and believe that *p iff* were S to trust H, this would result in S's getting closer to truth in a non-accidental way;
(ii) S ought to dismiss H and continue to believe that *q* iff were S to stick to her guns this would result in S's staying non-accidentally closer to truth, and;
(iii) in all other cases, S ought to suspend judgment about whether *p* or *q*.

By "getting closer to the truth non-accidentally" one can mean different things, but for the sake of clarity I will mean here "getting closer to truth in a reliable way", reliability condition of course includes also reliable reasoning. Having this formulation in mind, we can return to our main question: to what extent my ending up non-accidentally on the side which is closer to the truth depends on factors internal to the discussion, and to what extent it depends on independent factors? I want to suggest, based on probabilities, that looking at our virtues or competences relevant to the field of the debate could be epistemically more profitable than looking at our first-order evidence internal to the discussion. In what follows, I will use an argument by Timothy Williamson to show this. My aim is to throw a glove for further discussion along this track.

4. Evidence vs. virtues

In his paper "Very Improbable Knowing", Timothy Williamson presents the following case, a simplified and more ordinary version of which is this (see Williamson 2011). Imagine a minimalist clock with two pointers showing approximately that it is 8:00. When S looks at the clock from certain distance in normal perceptual conditions, S can easily take it to be 8:00 o'clock, when the minute's pointer is in position 7:58, or 7:59, or 8:01, or 8:02. These positions form a field H, which we can call: "margin of error field" or "margin of error set of positions". There are three basic elements in H as well as along the whole clock: endpoints (when the pointer is in positions 7:58, 7:59, etc.), intervals between endpoints, and midpoints – when the pointer is between two endpoints. Assume that the pointer is in position 0 when it is exactly 8:00. S knows that it is 8:00 *iff* it is in position 0, but not in a position from the margin of error field H. By dividing the space-time intervals within H as much as we want, we can reduce the probability of S knowing that *p* to 0.

The idea of the example is to show that given our margins of error, the probability of S knowing that she knows that *p* (that it is 8:00), based on the evidence that she possesses, is very low. The same holds for S's knowing that she is non-accidentally closer to the truth of *p*. The implication of this case for our discussion so far is the following. In majority of cases in which S thinks that she possesses good evidence in favor of *p*, she cannot be sure that

she knows that *p*. This is bad for the steadfast view, because one is not justified to remain steadfast in the absence of good evidence supporting her certainty. Note that the more sophisticated the debate is in terms of evidence, the greater the risk of error given our margins of error, and the more destructive Williamson's argument would be for the steadfast view.

Williamson's case is not equally applicable to any kind of evidence or evidence gathering. When we refer to *a priori* and *holistic* evidence the situation would be slightly different. It is hard to imagine how our *a priori* intuition would be a subject to margins or error argument. The same holds for holistic evidence, more precisely holistic evidence about the reliability of our faculties. If this evidence were leading to improbable knowledge, then Williamson's argument would not go through because he is using this kind of evidence to build up his case. More importantly, we use evidence of this sort when it comes to assessing external factors relevant to peer disagreement situations. We keep track of the cognitive success of our cognitive faculties as well as of the competences and virtues of other people by using holistic evidence based on coherence. If holistic evidence is not vulnerable to the very improbable knowing argument, this naturally puts external factor in better position than internal factors.

To what extend this argument favors conciliationism? It shows, though only in a negative way, that our evidence of external factors is more truth-conducive than the kind of evidence usually associated with internal factors. However, it does not follow that in all cases reconciliation is called for. There are some exceptions in which our first-order evidence is more than reliable. Possible such cases relate to *a priori*, holistic, and phenomenal evidence. So, although the argument presented here is a defense of independent factors, it does not favor conciliationism. The acquired result that independent factors could be a better guide to truth in peer disagreement situations than the internal first-order evidence heating up the discussion might seem counterintuitive, and I think it is worthy of further discussion.

References

CHRISTENSEN, D. (2009): Disagreement as Evidence: The Epistemology of Controversy. *Philosophy Compass* 4, No. 5, 756-767.

CHRISTENSEN, D. (2011): Disagreement, Question-Begging and Epistemic Self-Criticism. *Philosophers' Imprint* 11, No. 6, 1-22.

HAWTHORNE, J. – SRINIVASAN, A. (2013): Disagreement without Transparency: Some Bleak Thoughts. In: Lackey, J. – Christensen, D. (eds.): *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press, 9-31.

KELLY, T. (2005): The Epistemic Significance of Disagreement. In: Hawthorne, J. – Gendler, T. S. (eds.): *Oxford Studies in Epistemology. Vol. 1*. Oxford: Oxford University Press, 167-197.

KELLY, T. (2010): Peer Disagreement and Higher-Order Evidence. In: Feldman, R. – Warfield, T. (eds.): *Disagreement*. New York: Oxford University Press, 111-174.

LACKEY, J. (2010): A Justificationist View of Disagreement's Epistemic Significance. In: Haddock, A. – Millar, A. – Pritchard, D. (eds.): *Social Epistemology*. Oxford: Oxford University Press, 298-325.

SOSA, E. (2010): The Epistemology of Disagreement. In: Haddock, A. – Millar, A. – Pritchard, D. (eds.): *Social Epistemology*. Oxford: Oxford University Press, 286-294.

WILLIAMSON, T. (2011): Improbable Knowing. In: Dougherty, T. (ed.): *Evidentialism and its Discontents*. Oxford: Oxford University Press, 147-165.

Organon F 23 (1) 2016: 83-101

Analytic Method

MILOŠ KOSTEREC

Department of Logic and Methodology of Sciences. Faculty of Arts
Comenius University in Bratislava. Gondova 2. 814 99 Bratislava. Slovak Republic
milos.kosterec@gmail.com

RECEIVED: 05-09-2015 • ACCEPTED: 04-12-2015

ABSTRACT: This paper proposes a non-trivial definition of the notion of analytic method. Working within the so-called instructional model of method, I distinguish three kinds of instructions which occur in methods: *selective, executive*, and *declarative* instructions. I discuss the relation between each of these and the analyticity of a method. Then I define the notions of an *analytic use* of an instruction and of an *analytic instruction*, which are at the basis of the proposed definition of an analytic method. Finally, I discuss the issue of circularity in the presented model which arises if we consider a finite agent testing a method for analyticity.

KEYWORDS: Analytic method – analytic proposition – closure – instruction – knowledge base.

0. Introduction

The notion of an analytic method is widely used but rarely characterized.[1] Methods such as defining, explication or conceptual analysis are often considered analytic, but it is not clear which specific features lead to analyticity. The aim of this article is to provide a definition of *analytic method*. The defi-

[1] This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0149-12. I would like to thank my colleagues for helpful discussions. I am also grateful for comments by the reviewers.

nition should be non-trivial, i.e., not all methods should turn out to be analytic but at least one should.

I assume that an analytic method can be used without undertaking empirical research as a necessary component of any of the steps prescribed by the method.[2] Thus the use of an analytic method enlarges the researcher's knowledge without traversing the logical closure of such knowledge. In other words, one uses an analytic method to obtain, decode or make explicit information which is hidden, encoded or entailed by the information in a preexisting knowledge base.

The paper is divided into the following sections. Section 1 contains a brief specification of my theoretical framework. In Section 2, three types of instructions are introduced. In Sections 3 and 4, I further deal with selective instructions and the problems of information gain they present for the instructional model of methods. The roots of the problem are described in Section 5 and the role of information access is discussed there as well. In Section 6, I deal with the role of information access by distinguishing three different kinds of knowledge bases. Section 7 proposes definitions of analytic instruction and analytic method. Section 8 discusses the role of declarative instructions. It is followed by a case study about the analyticity of the method of explication in Section 9. In Section 10, I discuss the problems of a finite agent testing a method for analyticity. The paper ends with a brief concluding summary.

1. The framework

I presuppose some features of the instructional model of method presented in Bielik – Kosterec – Zouhar (2014a; 2014b; 2014c; 2014d) and exemplified, e.g., in Halas (2015a; 2015b). This model views a method as a systematic guide for reaching some epistemic goal. More precisely, any method is an ordered set of instructions which, when followed, lead one to a given goal. The use of a method is usually driven by a problem that cannot be resolved within one's knowledge base, i.e., the set of knowledge the agent uses. Hence, a method can be used to change that knowledge base in

[2] Analytic methods can, of course, be used to analyze the results of a *previous* empirical research.

order to resolve the problem. For example, one might not know whether a formula of propositional logic is a tautology. She can use the truth table method for solving this problem.

A method usually prescribes a series of steps. When following any of the steps, we are involved in a process specified by the method. Methods and processes are therefore closely related. There are a variety of systems used to model processes in general, like procedural models (see, e.g., Duží – Číhalová – Menšík 2014) or Petri nets (cf. Murata 1989). The instructional model views methods as composed of instructions, usually expressed by imperatives. I take imperatives to have semantic content *sui generis*. In general, they denote a relation between input states and output states. Below, I examine instructions that denote relations among states of knowledge. The features of a method are studied using the compositionality principle in the following form: The relevant features of the method are determined by its instructions and their composition. Thus the definition of *analytic method* shall specify some constraints on instructions and/or their composition. In the next section, I present a typology of instructions which will be later used to specify these constraints.

2. A typology of instructions

According to the instructional model, a method is an ordered set of instructions. In the context of the AMESH research project,[3] several methods have been studied within this framework – modeling, defining, explication, idealization, conceptual analysis, etc. Instructional models were developed for the methods of explication and sampling (Bielik – Kosterec – Zouhar 2014d), as well as abstraction and idealization (Halas 2015a) or definition (Zouhar 2015). The methods were specified in an idealized form (i.e., without context of use)[4] and their use was described in some case studies. For the purposes of the present paper, the main result of this previous work is that

[3] See http://www.amesh.sk/english.

[4] The papers cited also abstract from the explicit notion of an agent. Although here, when speaking informally, I sometimes mention an agent, I do not presuppose one explicitly in the instructions. There are different ways of viewing knowledge base: information in the set can be considered in relation to an agent, but also without regard to an agent, simply as a problem base.

one may roughly distinguish three types of instructions which occur in methods: *selective*, *executive* and *declarative*.[5]

Selective instructions order an agent to pick one of the possible ways to proceed further in a process. For example, the agent may be instructed to choose the number of samples which will be studied later. Selective instructions can also task an agent with picking an arbitrary value which is required for subsequent steps of the method – e.g., with picking a natural number from a range as an initial guess which is later improved upon.

This selection is seldom completely arbitrary. Usually some filtering features are specified. For example, consider the following two selective instructions:

A) Pick a natural number!
B) Pick an even natural number!

Instruction B is more specific than instruction A. Selective instructions are the main source of indeterminacy of methods. Hence even though a method is stated as a set of instructions, its result may nevertheless be undetermined.

The second type of instructions is *executive instructions*. As providers of the actual computation steps of the procedure formulated in the method, they form the core of a method. While selective instructions order an agent to acquire input parameters from her environment, execution instructions specify a computation using these parameters. Usually, selective instructions provide the material, so to speak, and executive instructions build further on that material.[6] Some examples of executive instructions are:

Find the greatest common divisor of numbers *a* and *b*!
Solve the equation *E*!
Compute the median value of the measurement data!

---

[5] These are types of *simple* instructions. *Complex* instructions contain other instructions.

[6] Of course, selective instructions are not the only providers of material for executive instructions – one executive instruction can provide material for another one as well.

Similarly to a practice common in programming, where existing code is often reused as part of a larger task, or in logic, where proofs are used within other proofs, a method can itself serve as an executive instruction within another method.

The third type of instructions is *declarative instructions*. They prompt an agent to publicly declare her results. Some examples of declarative instructions are:

Declare that there is no largest prime!
Declare the number of inhabitants of the capital city!

The term "public" as used here is very general. It covers cases such as a classroom full of students or the set of readers of an academic journal. It may also simply be the agent herself. From a general point of view, the nature of the audience can be disregarded. The important feature of these instructions is that the agent assumes the declared results as true in further work.

3. The information value of selection

Assuming the instructional model of method, a specific kind of instructions often occurs (among others) in the methods thus reconstructed. These instructions have at least one common feature: they instruct the agent to pick an object. The object may be a material or an abstract entity or even a set of such entities. Usually, the agent chooses from various possibilities provided either by her knowledge or by her actions using that knowledge. In general, these instructions are a source of indeterminacy in the method. One cannot determine, at least not solely on the basis of the instructions themselves, which object will be selected.[7]

[7] One way of looking at this is as a branching of the ways in which the knowledge of an agent can be enhanced, if she follows the method. A selective instruction prescribes a choice, not the object to be chosen. Different choices may thus provide different ways of enhancing knowledge of the researcher. This can be modeled, for example, in a tree. Another kind of branching occurs when a method includes complex instructions, i.e., instructions composed of sub-instructions (e.g., "Stand up and shut the door!"). In Bielik – Kosterec – Zouhar (2014b; 2014c) we modeled only hypothetical instructions of the form "*if … then … else*" as leading to branching.

The notion of selection from many possibilities is closely related to the notion of probability. The theory of information assigns information value to selections (cf. Shannon 1948). Roughly stated, the lower the probability, the higher the information gain, if there is some at all. The notion of information is also closely related to the notion of knowledge (see Abramsky 2008). My aim here is to argue that, contrary to intuitions, selective instructions do not present any relevant information gain. From this point of view, not all of the information about an agent's progress is relevant. What matters are the changes of the information states of the knowledge base.

4. The problem of selective instructions

If we include information provided by selective instructions into our model, we face indeterminate changes in the knowledge base.[8] To avoid this, we could exclude that piece of information, but this should not be a simple *ad hoc* adaptation. I shall now show that the problem of indeterminacy due to selective instructions arises from neglecting the role of information access in the instructional model.

Bielik – Kosterec – Zouhar (2014a; 2014b; 2014c; 2014d) did not state any conditions for the analyticity of method. One of the reasons was the very problem of selective instructions. On the one hand, it was clear that they do not present any relevant information or knowledge gain. On the other hand, all instructions were modeled as possible knowledge changers. The original model did not deal with the indeterminacy brought about by selective instructions. Therefore, it faces the problem that the use of a selective instruction surpases the logical closure of the initial explicit knowledge base. Hence, any method containing a selective instruction wouldn't fit our intuitions about an analytic method.

One way to solve this problem is to simply leave out the selective instructions when testing the method for analyticity. The main problem of this approach is that of justifying it in a non-*ad hoc* way. Another solution would be to pick out some relevant part of the knowledge base and test the analyticity of a method only with respect to the logical closure of the selected part of the

[8] These changes are indeterminate in the way that a) they are not constant and b) they are not entailed and cannot be exactly predicted.

base. But again, the problem is how to justify such an approach in a non-*ad hoc* way.

In testing methods for analyticity, the analyticity of instructions should be the key indicator.[9] The first approach disregards selective instructions, but for no good reason. The second approach views the information gain provided by selective instructions as irrelevant, but again, for no good reason. Both approaches assume that selective instructions ought to be left out. But is that so?

5. Information access

The main piece missing in the instructional model of method is a model of the flow of information between the agent, her explicit knowledge base and her operational knowledge base. The difference between the explicit and the operational base can be described as follows. Imagine yourself using a notebook computer. Its hard drive contains some explicit data. It certainly does not contain all the logical consequences of those data. After you sign into your account, not even all of your explicit data are accessed all at once. They are only *accessible*, not *accessed*. You first pick some of the information stored on your hard drive to access it, and then proceed to make changes to it. Afterwards you may save the changes or cancel your work. The hard drive here is analogous to the explicit knowledge base. The information you accessed – e.g., documents opened, music played, etc. – is analogous to the operational base.

The instructional model of method does not presuppose this difference between information that is accessible and information that is actually accessed and worked with.[10] However, there occurs an important flow of information between them. A model that is able to capture this would not be *ad hoc*.

---

[9] Further analysis would be required as to whether the analyticity of instructions is similar or related in some way to the analyticity of propositions.

[10] The dividing line between accessible information and information actually accessed requires a more detailed examination which must be postponed to a different occasion. Of course, at a closer look, the dividing line could appear much less strict than it is supposed here.

Below, I shall model information access connected with instructions. The main idea is as follows: selective instructions provide access to information from the explicit knowledge base. They therefore do not change the information state of this base. One does not modify information merely by accessing it. Selective instructions simply retrieve some of the information already included in the knowledge base and provide it to the operational base. Consider, again, the analogy with one's work on a computer: one does not change a document by opening it. Similarly, in the context of a method, the change of information may only be due to executive instructions.

6. Executive instructions vs. selective instructions

A common feature of all three types of instructions is that they describe steps between the states of some knowledge base. It is important, however, to distinguish between three different knowledge bases.[11]

The *explicit knowledge base* of an agent can be viewed as a base of all information accessible by selection. In other words, an agent need not provide any thoughtful step to obtain information from this base.[12] The *operational knowledge base* represents the information already accessed. This base contains the propositions the agent is currently working with (e.g., a finite set of axioms when proving a theorem). The *implicit knowledge base* represents the bounds of analyticity of the initial explicit knowledge base. We can think of it as a logical closure of the explicit knowledge base.[13] A selective instruction

[11] For the purposes of this paper, knowledge base can be viewed simply as a pair of a universe (a set of objects) and a set of propositions.

[12] Of course, selection may require *some* work in advance (e.g. an agent will have to pick some information within a specific range).

[13] The need to differentiate between explicit information and its closure is well established in epistemology (see, e.g., Dretske 2005) and computer science (cf. Vardi 1989). The consensus is that an agent need not know all of the logical consequences of her knowledge (see also Jago 2014, Ch. 6). The difference between logical entailment and knowledge closure is discussed in the approach of relevant alternatives (see Holliday 2012; 2015). From a technical point of view, the set is closed on an operation if it contains the results of the application of that operation on all its members. For example, the set of natural numbers is closed on the operation of addition. The closure of explicit initial knowledge base presupposed here is the union of the *validity*

provides a step from a state of the explicit knowledge base to a state of the operational knowledge base. It may also provide a step between states of the operational base (e.g., the agent selects a possible way of proceeding further). A declarative instruction provides a step from a state of the operational base to a state of the explicit base.[14] Finally, an executive instruction provides a step between states of the operational base.

I shall now clarify the difference between selective and executive instructions. It seems that the intersection of these two sets is not empty: selective instructions may apparently provide steps between states of the operational base, such as when one chooses from several possibilities during computation. However, I propose to model selective instructions exclusively as steps from states of the implicit knowledge base to states of the operational base.[15] In the previous section, I argued why we need not consider the information provided by a selective instruction as relevant. When a selective instruction serves as a step between two states of the operational base, this means that we are at a point of the process where previous executive instructions have already provided us with some results. The following steps of the method require us to select from these preliminary results. This selection is not predetermined. But as long as the possibilities were obtained by analytic instructions,[16] they should be included in the implicit knowledge base of the agent, i.e., the logical closure of the explicit knowledge base. Thus, in the case of

---

closure, the *closure on logical entailment* and the *closure on semantic analysis*. I also presuppose that the explicit initial knowledge base contains all of the relevant mathematical theories. The validity closure of the explicit base contains all sentences which are true whenever all sentences of the explicit base are true. The logical closure of the explicit base contains all the logical consequences of sentences included in the initial explicit knowledge base. The closure on semantic analysis of the explicit base contains all relevant semantic parts of the sentences in the explicit base.

[14] The result of a declarative instruction (i.e. declaration) possibly enriches the explicit base. Therefore, it possibly changes its state.

[15] In some methods, perhaps, an executive instruction has to be executed to make a step from the implicit base to the explicit base possible. I do not consider such cases in this paper.

[16] I define this term in Section 7. Simply put, the execution of an analytic instruction (which possibly results in a change of the operational knowledge by some proposition) never results in a proposition which is not included in the implicit knowledge base.

analytic methods, we can model selective instructions as steps from states of the implicit knowledge base to states of the operational base. Therefore, the set of selective instructions and that of executive instructions are disjoint.

But how can we know whether this is in fact possible? The possibility hinges on the assumption that all of the executive instructions preceding the given selective instruction are analytic. If the method is not analytic, it is because it contains at least one non-analytic executive instruction (i.e., an instruction which provides us with a proposition that is not in the implicit knowledge base and by doing so broadens our initial knowledge). In this latter case, the proposed transformation of selective instructions may not work. Thus, when testing a method for analyticity, we need to look no further than at the executive instructions which can be clearly differentiated from the selective instructions.

7. Analytic instruction and analytic method

Until now, I have distinguished three types of instructions, arguing that only executive instructions should play a role when checking a method's analyticity. The difference between selective and executive instructions has been based on the notion of *analytic instruction*. This section aims to specify this notion.

A useful distinction that will be important later is one between an instruction and its manifold uses. The same instruction can accept different inputs and it can also be used more than once within the same method. Generally, the use of instruction is the application of the instruction to some input.

Now, let me introduce the notion of the *descriptive result of the executive instruction* (DRE):

> The DRE of the executive instruction *I*, which provides the output *b* for the input *a*, is *a*-*I*-*b*.

a-*I*-*b* is a structured description of instruction *I* accepting input *a* and leading to the result *b*. The structure is very simple. Its only use is in distinguishing the input of the use of instruction (*a*), the instruction itself (*I*), and the result of the use of instruction (*b*). For every use of the instruction there is a corresponding description. However, the DRE is not a proposition – it is just a structure. In methods used in science, the input is an object from the state of

knowledge base (the reader is reminded that the base is a pair of a set of objects and a set of propositions). The use of an instruction (i.e., its execution accepting a given input) provides us with an object (*b*) which enriches the universe of the knowledge base.[17] The use of an instruction also enriches the set of propositions of the knowledge base by a descriptive proposition.[18] Therefore, the use of an instruction possibly changes the state of the operational knowledge base of the scientist.

There is one specific proposition that can be assigned to each DRE. The proposition is obtained in the following way. The semantic analysis of an instruction reveals a semantic operator which is central to the instruction. For example, the main operation of the instruction

I) Add numbers *a* and *b*!

is the operation of addition (*add*). My task here is not to specify the method for identifying main operation of an instruction. I simply presuppose that there is such an operation for every instruction. Of course, this strong assumption may appear incorrect due to the vast number of kinds of imperatives. A more detailed justification of this assumption would require a thorough investigation into the semantics of imperatives. The main reason for making this assumption here is the semantic principle of compositionality. According to this principle, for every complex semantic unit[19] there is a tree structure which represents how simpler semantic units are composed to form the complex unit. Any tree has a single root node which is the result of the application of the last operation to combine the simpler semantic parts into the final, complex one.

Further, I presuppose that for every main operation in an imperative there is another closely related operation. In my example, the main operation was *add*, and the corresponding other operation is +. This needs closer examination. Let us compare the two structures:

a) Add 2 and 3!

---

[17] That is, operational knowledge base.

[18] I shall return to this shortly. Here it suffices to state that the descriptive proposition conveys information about the result of an instruction accepting some input.

[19] From the structuralist point of view, a proposition is a structured semantic unit. Imperatives are structured semantic units as well.

b) 2 + 3

I argue that *add* and + have different semantic contents. This is readily seen from the fact that in structure a), *add* is unary (it applies to the complex *2 and 3*), while in b), + is binary. Another difference is that by combining the operation *add* with its argument we compose an imperative, whereas by combining + and its arguments we compose a number (or a construction of a number), but certainly not a proposition. Therefore it is reasonable to consider the semantic content of *add* and + as different.

Nevertheless, there is a close connection between these two operations. The main semantic operation of the imperative (*imperation*) prescribes some action on its input. The result of the prescribed imperation is exactly the same object which is denoted by the corresponding complex structure b). I use the DRE to explain the connection:

The DRE of imperative a) is: 2, 3-Add x and y!-5.

We can describe the result of b) as follows:

c) 2 + 3 = 5

We obtain this proposition from the DRE as follows. We first obtain the input of the DRE (2, 3). Then we obtain the imperation of the instruction of the DRE (*add*). This imperation is closely connected to the operation (+). We apply the operation to the input arguments and obtain the result (5). This is described by the proposition c). In a similar way, *any* DRE can be assigned a *descriptive proposition*.[20]

Finally, I can turn to formulating the definitions. First, let me define *the analytic use of an instruction*:

Def 1: The use of instruction *I* is *analytic* iff its descriptive proposition is analytic.[21]

---

[20] Of course, this needs further investigation. At the moment, if the assumption of a close connection between imperations and operations is correct, then the acquisition of a descriptive proposition from any DRE is a straightforward process.

[21] This paper does not specify the notion of *analytic proposition* and simply presupposes it. Some models of analytic propositions are provided, e.g., by Duží (2010 and 2013).

The descriptive proposition is assigned to a particular input-output pair provided by the use of the instruction. The instruction usually generates a whole relation and not only one such pair. A more general definition is thus required:

Def 2: The instruction *I* is *analytic* iff all its uses are analytic.

Here, the distinction between the analytic use of an instruction and the analyticity of the instruction itself is crucial, simply because for some instructions only some of their uses are analytic.[22]

Testing an instruction for analyticity is by no means an easy task. I discuss some features and possible problems of such testing later. With the two previous definitions at hand, I can now specify the conditions of analyticity of methods viewed as ordered sets of instructions:

Def 3: The method *M* is *analytic* iff all uses of its executive instructions are analytic.

How does one go about testing a method for analyticity? One has to analyze uses of its executive instructions. Sometimes it is possible to generalize about them. One then has to check whether all executive instructions included in the method lead to analytic descriptive propositions. DREs can be used for this purpose.

8. Declarative instructions

Selective instructions provide access to accessible information, while executive instructions prescribe actual computations of the method. The last

[22] Consider the following instruction inspired by Cmorej (1996): Test whether *x* is as old as Peter! The result of this test for some object is either *True* or *False*. Consider an object, *A*, different from Peter. Suppose we made the testing and the result was *False* (e.g., *A* is younger than Peter *at the time of testing*.). The DRE is: "A-Test whether *x* is as old as Peter!-*False*". The descriptive proposition is: "Test whether (*A*) is as old as Peter = *False*". But this proposition is not analytic. It is possible (perhaps in some other world) that *A* is as old as Peter. On the other hand, if we use this instruction with respect to Peter, then the result will always be *True*. Hence, for some instructions, only some of their uses are analytic.

type of instructions is declarative instructions. Should they be tested for analyticity?

The general scheme for a declarative instruction is:

Declare *A* to be in relation *R* with *B*!

A semantic analysis reveals that the main operator (besides exclamation), is *to declare*. Therefore, the main operation provided by a declarative instruction is declaration or public announcement. The field of public announcement logics is already well established (see Baltag – Moss – Solecki 1998; Wang 2013). A common view is that the public announcement of a proposition changes the situation of an agent in that after the announcement, she takes the proposition to be true. What happens is an update of the agent's knowledge base. Above, I discussed the closures on initial explicit knowledge base of the agent. Consider now the operator of declaration. It does not seem to be analytic, necessary or determined – one cannot predict what will be declared.

If we include declarative instructions among instructions tested for analyticity, the results will be practically trivial. All methods containing a declarative instruction will not be analytic. This situation is similar to the one we have seen above, with selective instructions. On the one hand, it seems that declarative instructions are part and parcel of methods. On the other hand, their inclusion seems to leads to surpassing the limits of the closures of the initial explicit knowledge base.

The remedy could be seen, again, in considering the role of the information access. The difference between selective and executive instruction was that executive instructions do not operate on the explicit knowledge base. They do not access information – what they do is computation (in a very general sense) using information already accessed. Now, consider declarative instructions. Rather than providing computation using accessed information, they provide updating of the explicit knowledge base with respect to the results obtained by executive instructions. In our computer analogy, what they do is saving changes. Declarative instructions usually provide a statement which explicitly names or labels some element. Could we leave these instructions out of testing for the analyticity of methods? Again, a positive answer based on non-*ad hoc* reasons is required.

The role of declarative instruction is to declare a result obtained by previous instructions. The declaration of result is not a result in itself. If a result of

some executive instruction is not analytic, the announcement of the result can do nothing to change that. Thus the analyticity of a result is independent of the announcement. In other words, it would be strange to consider a proof as non-analytic just because it was published or announced. Therefore, the testing a declarative instruction for analyticity is redundant.

9. Case study

Bielik – Kosterec – Zouhar (2014d) proposed a model of the method of explication. Here is its simplified version:

1. Select the content, A, which does not provide the required theoretical functions!
2. Declare the content A to be the explicandum!
3. Select the contents, B, C, D, which are to be used!
4. If the contents are not clear enough, then clarify them, else construct the content, E, out of B, C, D!
5. Test whether the constructed content is theoretically valuable!
6. Declare the content E to be the explicans of the content A!

The first step in testing this method for analyticity is dividing the instructions involved into selective, executive and declarative. Obviously, instructions 2 and 6 are declarative. These will not be tested for analyticity. Instructions 1 and 3 are selective. They presuppose availability and accessibility of the contents A, B, C and D, and they merely provide us with access to these contents. Instructions 4 and 5 are executive. Let me now describe their testing.

Instruction 4 is a complex instruction of the form "*if … then … else …*". The instructions in each of the clauses need to be tested individually – the entire instruction is analytic only if all the instructions in all clauses are analytic. The first clause includes the method of clarification. This is an example of a method being used as an executive instruction in another method. Thus, the analyticity of the method of explication depends on the analyticity of the method of clarification.[23] The second clause contains an instruction that leads us to construct content out of some specified building blocks. The main op-

[23] The method of explication is therefore a complex method which has another method as its subpart.

erator of this instruction is *to construct* and the associated operation is construction. The construction of content out of specified parts seems unproblematic. Therefore, the corresponding descriptive proposition can be considered analytic. Instruction 4 is therefore analytic if the method of clarification used is analytic.

Instruction 5 is a testing instruction. It assumes that some criteria of theoretical value are specified. A constructed content enters as input and the output is a positive or negative answer. Let BCD be the content constructed in the previous instruction. The DREs are

BCD-(Test for value)-is valuable.

or

BCD-(Test for value)-is not valuable.

The descriptive propositions are:

BCD is valuable according to the test.

or

BCD is not valuable according to the test.

The analyticity of this instruction obviously depends on the analyticity of the method of testing.

To summarize, we can see that the method of explication as stated here contains other methods. The analyticity of the former therefore depends on analyticity of the latter. If all the methods contained are analytic, then the whole method is analytic. In this section, I described using the proposed definitions. First we have to select the executive instructions. Then we have to check their uses. If an executive instruction is method in itself, then we must check this method for analyticity.

10. The problem of analytic test

So far, I have discussed the kinds of instructions involved in methods and their influence on analyticity of methods. I argued that only executive in-

structions should be taken into consideration. Let me now turn to one more possible source of problems.

The definition of analytic instruction is ultimately based on the notion of analytic proposition. Thus, the ability to distinguish analytic and non-analytic propositions enables one to distinguish analytic and non-analytic instructions (at least within the model proposed here). The crucial question here is *how we test propositions for analyticity*. The simple, off the cuff answer seems to be that we apply some testing methods. This, however, raises another question. Should we use analytic methods in order to tell analytic propositions from non-analytic? A negative answer seems strange. If non-analytic tests were used in testing for analyticity of propositions, this would make our knowledge of analyticity dependent on empirical factors. It seems that the tests must be analytic. But this makes our knowledge of analyticity of propositions dependent on analyticity of some method. Have we come full circle?

The issue is related to the cardinalities of the set of analytic methods and the set of analytic propositions. It is clear that if there were only one analytic method and only one analytic proposition, our model of analytic methods would be circular. The method would be analytic because the proposition is analytic, and we would only ever know that the proposition is analytic because the method is considered analytic. But what if the number of methods is not finite? Once again, after a while, we get circularity. After using up all analytic propositions from the finite set we would need another to support the analyticity of some testing method. So it seems that we should presuppose that both sets, the set of analytic methods and the set of analytic propositions, are infinite. This then need not lead to the circularity mentioned. The drawback here is that we are finite entities which can only provide finite tests. We therefore have to consider some method or proposition to be analytic without justification prescribed by the presented model of analyticity of the methods.

The reason for the circularity is that the testing methods for analyticity of propositions should be analytic. I call this *the problem of analytic test*, which is due to our notion of analytic method presupposing that we, as finite agents, know at least some analytic propositions beforehand without any test.

11. Conclusion

The main aim of this paper was to provide a definition of the notion of analytic method. Using the instructional model of methods, I distinguished three different types of instruction. Selective instructions are used to access information or to provide steps undetermined by the method itself. Executive instructions provide the actual computation steps of the method. Declarative instructions serve to declare the results obtained by the method.

I conditioned the analyticity of method by the analyticity of executive instructions involved. An executive instruction is analytic if descriptive propositions obtained from the descriptive result are analytic. I therefore conditioned the analyticity of methods by the analyticity of propositions. However, as soon as we consider the role of finite agents in testing of methods for analyticity, the problem of analytic test arises.

References

ABRAMSKY, S. (2008): Information, Processes and Games. In: Adriaans, P. – van Benthem, J. (eds.): *Philosophy of Information: Handbook of the Philosophy of Science.* Amsterdam: Elsevier Science, 483-550.

BALTAG, A. – MOSS, L. S. – SOLECKI, S. (1998): The Logic of Public Announcements, Common Knowledge, and Private Suspicions. In: Gilboa, I. (ed.): *TARK '98: Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge*. San Francisco: Morgan Kaufmann, 43-56.

BIELIK, L. – KOSTEREC, M. – ZOUHAR, M. (2014a): Model metódy (1): Metóda a problém. [The Model of Method (1): Method and Problem.] *Filozofia* 69, No. 2, 105-118.

BIELIK, L. – KOSTEREC, M. – ZOUHAR, M. (2014b): Model metódy (2): Inštrukcia a imperatív. [The Model of Method (2): Instruction and Imperative.] *Filozofia* 69, No. 3, 197-211.

BIELIK, L. – KOSTEREC, M. – ZOUHAR, M. (2014c): Model metódy (3): Inštrukcia a metóda. [The Model of Method (3): Instruction and Method.] *Filozofia* 69, No. 8, 637-652.

BIELIK, L. – KOSTEREC, M. – ZOUHAR, M. (2014d): Model metódy (4): Aplikácia a klasifikácia. [The Model of Method (4): Application and Classification.] *Filozofia* 69, No. 9, 737-751.

CMOREJ, P. (1996): Empirical Essential Properties. *Organon F* 3, No. 3, 239-261.

DRETSKE, F. (2005): Is Knowledge Closed under Known Entailment? In: Steup, M. - Sosa, E. (eds.): *Contemporary Debates in Epistemology.* Boston: Blackwell, 13-26.

DUŽÍ, M. (2010): The Paradox of Inference and the Non-Triviality of Analytic Information. *Journal of Philosophical Logic* 39, No. 5, 473-510.

DUŽÍ, M. – ČÍHALOVÁ, M. – MENŠÍK, M. (2014): Logical Specification of Processes. In: Tokuda, T. et al. (eds.): *Information Modelling and Knowledge Bases*. Amsterdam: IOS Press, 45-63.

DUŽÍ, M. – JESPERSEN, B. (2013): Procedural Isomorphism, Analytic Information, and β-Conversion by Value. *Logic Journal of the IGPL* 21, No. 2, 291-308.

HALAS, J. (2015a): Abstrakcia a idealizácia ako metódy spoločensko-humanitných disciplín. [Abstraction and Idealization as Methods of Social Sciences and Humanities.] *Organon F* 22, No. 1, 71-89.

HALAS, J. (2015b): Pojem *racionálneho aktéra* a metóda idealizácie. [The Notion of *Rational Agent* and the Method of Idealization.] In: Sucharek, P. – Marchevský, O. (eds.): *Racionálne – iracionálne: Zborník vedeckých príspevkov*. Prešov: Prešovská univerzita, 263-267.

HOLLIDAY, W. H. (2012): Epistemic Logic, Relevant Alternatives, and the Dynamics of Context. In: Lassiter, D. – Slavkovik, M. (eds.): *ESSLLI Student Sessions, LNCS 7145*. Berlin: Springer, 109-129.

HOLLIDAY, W. H. (2015): Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism. *Journal of Philosophical Logic* 44, No. 1, 1-62.

JAGO, M. (2014): *The Impossible*. Oxford: Oxford University Press.

MURATA, T. (1989): Petri Nets: Properties, Analysis and Applications. *Proceedings of the IEEE* 77, No. 4, 541-580.

SHANNON, C. E. (1948): A Mathematical Theory of Communication. *Bell System Technical Journal* 27, 379-423; 623-656.

VARDI, M. Y. (1989): On the Complexity of Epistemic Reasoning. In: Parikh, R. (ed.): *Proceedings of the Fourth Annual Symposium on Logic in Computer Science*. Piscataway: IEEE Press, 243-252.

WANG, Y. – CAO, Q. (2013): On Axiomatizations of Public Announcement Logic. *Synthese* 190, Suppl. 1, 103-134.

ZOUHAR, M. (2015): Metóda definovania. [The Method of Defining.] *Filozofia* 70, No. 4, 258-271.

Organon F 23 (1) 2016: 102-124

Fictional Discourse. Replies to *Organon F* Papers (Part II)

PETR KOŤÁTKO

Department of Analytic Philosophy. Institute of Philosophy. Czech Academy of Sciences
Jilská 1. 110 00 Prague 1. Czech Republic
kotatko@flu.cas.cz

In what follows I continue in replying to papers which have been collected by Juraj Hvorecký and appeared in the first Supplementary Issue of *Organon F* 2015.[1] I need not repeat (but cannot resist doing so) how exciting it was to read that exceptional bunch of papers, written by brilliant philosophers and my good friends. As usual, in my replies I am voicing agreement and disagreement with the same pleasure (the latter perhaps with greater passion), and take both as a sign of deep respect.

Anders Pettersson[2]

The provocative title of Anders' paper and his radical exposition of the paradox he finds in the widely shared theoretical notion of a text as an abstract object present a good motivation for clarifying certain concepts – and Anders does a considerable work in this field. He works with a strong concept of abstract objects as atemporal entities, with the counterintuitive consequence that texts, fictional stories, characters engaged in them, symphonies,

[1] The first part can be found in Koťátko (2015). In both parts I preserve the order in which the papers originally appeared. I owe my thanks to Juraj Hvorecký for a thorough review of the text and correction of my English.

[2] Reply to Pettersson (2015). All the page references which appear without the author's name refer to this paper.

laws etc., if conceived as abstract entities, cannot be created. This generates an interesting tension with Zsofia Zvolenszky's paper based on the concept of abstract artifacts – abstract entities brought into existence by human creative activities. Nevertheless, even if we accept such a concept (as I think we should), and classify texts as abstract artifacts, the question how or in what sense something abstract can be involved in relations and bear properties we are used to ascribe to physical objects poses an interesting challenge.

Anders is not satisfied with explanations of the kind given by Stephen Davies: symphony is an abstract object but "it can be noisy and triumphant at its close" – which means that "its 'well formed instances' must be noisy and triumphant at its close" (p. 120). To be sure, this involves certain ambiguity in ascribing properties to their bearers. Noisy in straightforward sense (or: noisy "in the last instance" – not parasitic upon any more fundamental case of something having some property) can be only spatiotemporal events: but abstract entities like symphonies can have properties which have normative implications concerning noisiness of certain acoustic events, namely those which are presented as their instances. There is no harm, I would say, if we, for the sake of brevity, describe those properties by using the term "noisy". We normally do so and that's why it would sound counterintuitive to deny that symphonies can be noisy. With the explanation just given we can avoid category mistake (of straightforwardly, literarily ascribing physical properties to abstract entities) without falling into conflict with ordinary usage. To do justice to the way we use language, we simply have to admit that when making an utterance of the form *S is P* we do not have to predicate *P* to the referent of *S*. We can assign it to its instances, if the referent is an abstract entity, saying thereby something rather complex *about that abstract entity*: namely that it is such that all its instances (or "proper", "standard", "well formed" instances) have the property specified by *P*.

Perhaps we can borrow the terminology Pavel Tichý in (1978) has introduced in the context of theory of descriptions. Since he insisted that expressions like "the US president" refer in all their uses to a concept (also: "character" or "office"), rather than to an individual (uniquely satisfying the concept or occupying the office in the time of the utterance), he had to make a special move to account for sentences like "The US president is usually a white rich male". In such a case Tichý speaks about *predication de re*: the property specified in the predicate is not ascribed to the concept referred to by the expression in the grammatical subject but to the individual uniquely

satisfying the concept. On the contrary, in cases like "The US president is electable" the property is ascribed to the concept itself: that is *predication de dicto*.[3] The relation of satisfying (or occupying) which holds between Tichý's concepts (or offices) and individuals satisfying (or occupying) them can be approached as a special case of the relation between abstract entities and their instances and I think we can keep Tichý's terminology even on this more general level. In our case, it makes certainly sense (though not the same sense) to say that noisiness belongs to the symphony as an abstract object and that it belongs to its instances (particular performances): in the latter case the sense is straightforwardly simple, in the former it includes a mediating construction specified above.

With the same right one can say that it makes good (though not the same) sense to say that we read a text as an abstract entity and to say that we read a particular instance of that text. The specific feature of this case is that in doing so we draw attention to two aspects of a rather complex phenomenon in which relations to the text as an abstract entity and to its physical instance are inseparably interconnected. Since reading, in the literal sense of the word „reading", includes:

(1) focusing on some particular physical object (typically string of shapes on the paper, on the wall etc.);
(2) identifying it as an instance of a text as abstract entity;
(3) interpreting the sentence types belonging to the text as bearing certain literal meanings;

[3] In other words, Tichý had to allow for plurality on the level of predication if he wanted to avoid it on the level of referents of descriptions (i.e. if he did not want to admit that "the US president" sometimes refers to a concept and sometimes to an individual uniquely satisfying it in the time of the utterance). Our situation is analogical: we have to join Tichý in postulating two kinds of predication, if we want to insist that the terms like "symphony" or "text" in all their uses refer to an abstract entity. The question is whether we should insist on this. If we do so, it is hard to see how to account for sentences like "Can you pass me that text?" Here we certainly do not ask the addressee to pass us an instance of some text (as a contextually identified abstract entity) but to pass us particular (deictically identified) physical object which is, as we presuppose, an instance of some (unspecified) text as an abstract entity. This should be distinguished from an utterance of "Could you give me the text?" made by an actor during the rehearsal: here he asks to be given some (whatever) instance of particular text as abstract entity (text of particular play).

(4) interpreting the particular occurrences of instances of these sentences with their literal meanings, produced in given context, as used to perform speech acts with particular force and propositional contents;
(5) occasionally, interpreting the fact, that such and such speech acts have been directly (literarily) performed in given context, as a performance of some indirect speech acts (or: as generating some implicatures in Gricean sense).[4]

So, it makes equally good sense to relate reading to the text as an abstract entity and to its physical instance: in fact, both relations are present and interwoven in the act of reading. None of them can in itself constitute reading: neither mere contact with a string of shapes or sounds nor mere intellectual (interpretative) operations with a text as an abstract entity instantiated in our mind, not based on a simultaneous intercourse with its external physical instance, would be called "reading".

Here is another example of ambiguity on the level of predication. If we, as I have suggested in Koťátko (2004), define literary work as a structured complex of the text's literary aspirations, we presuppose that it makes sense to ascribe aspirations to a text (similarly like we ascribe intentions to literary works when using the term "intentio operis"). But the definition also provides space for ascribing them to the literary work. In these two ascriptions we just appeal to two ways in which something can belong to something: particular literary aspirations belong to the text as to their bearer (the text bears and indicates certain aspirations) and they belong to the literary work as to the whole which includes them as its parts.[5] Similarly, when speaking about reading the text of *Madame Bovary*, we can mean both the text as an abstract entity and its particular instances, activating thereby two correlative aspects of the sense of "reading something" (cf. the preceding paragraph).

---

[4] This is not to say that reading is a series of steps proceeding in the order presented here: for instance, the hypothesis about what the author of a text intended to convey in given context (step 5) can help us in identifying the words written (step 2): think e.g. of the case of an illegible manuscript.

[5] This may recall long philosophico-logical tradition beginning with Aristotelian interpretation of the system of categories as representing various ways in which something can belong to something – and, on the level of predication, something can be ascribed to something (cf. Anal. prot. I, c. 37, p. 49a).

In the light of all this, we have no reasons to complain that the common use of the word "text" balances between the physical and non-physical reading,[6] as Anders shows in Sections 1 and 2: this is certainly not a case for philosophical "therapy" (also here I join Anders: cf. p. 124). He also points out (p. 122), that the ordinary use of the word "text" in some contexts refers not only to symbols (as types), but also to their meanings. I believe that even the theoretical usage should allow us to keep these things together as distinct *levels of the text*. Taking into account also the distinction between meanings of expression types and meanings expressed by their use in given context, we should distinguish:

(1) $text_1$ as a syntactically identified series of symbols (types);
(2) $text_2$ as a syntactically identified series of symbols (types) interpreted according to the conventions of particular language;
(3) $text_3$ as a syntactically identified series of symbols (types) interpreted as used to express certain propositional contents with certain illocutionary force.[7]

In Koťátko (2013) I have tried to analyze various ways in which identification of these text-levels within the literary interpretation contributes to (or depends on) identification of the literary work.

To sum up, I agree that the attempts "to press the ordinary concept of a text to perform heavy theoretical duty for which it was never conceived" (p. 131) lead to aporias and appreciate the precision with which Anders shows their nature. On the other side, I still believe that it makes good sense to try to

---

[6] This is what I would prefer to say instead of Anders' classification of the "ordinary conception of a text" as "illogical" or "logically incoherent" (p. 124, 127 ff.). I find the concept manifested in the ordinary use of the word "text" rather dynamic or flexible than incoherent: it enables us to do justice to the text's complex mode of existence, including both a concatenation of expression-types endowed with meanings and physical instances of the former.

[7] "As used" (applying to sentence-types) is to be read "as used within the given text as a type" – so even on his level we do not shift from a text as an abstract entity to its instances. The $text_3$ clearly does not coincide with David Davies' notion of work defined as "a text-type as used as an artistic vehicle in a particular generative context" (cf. Pettersson 2015, 126). The identification of the $text_3$ and the identification of the literary work (in my sense – cf. above) are mutually dependent moves, requiring respect to "particular generative context" (circumstances of the text's origin).

work with the theoretical concept of the text in a way which will be compatible with our ordinary use of the term "text" and do justice to the intuitions manifested in this use. I do not find this aspiration unrealistic since, unlike Anders, I do not find the ordinary concept of text incoherent.

Göran Rossholm[8]

Göran discusses, in the most illuminating way, the problem of completeness/incompleteness of fictional worlds and their components and the way in which we should approach various kinds of gaps we find in texts of narrative fiction. In one moment (cf. p. 136) he quotes from Ingarden: "In a real object such Leerstellen are not possible. At most the material is unknown" and adds: "Why not just say the same about fictional and factual accounts that mention a table without mentioning the material it is made of: we do not know the material and that is all. Koťátko does not discuss this question" (p. 136).

In fact, I have made some comments on this issue and they are, I suppose, in full agreement with Göran's position. Let me try to imitate a possible dispute (opened by Göran's question) with the adherents of the incompleteness thesis (IT). Their most straightforward reply could go as follows:

> (IT) In the case of our talk about real entities in everyday communication – in contrast to the fictional discourse – there is a sphere which determines what our descriptions leave underdetermined, namely the actual world. This world contains, independently of our descriptions, complete individuals, definite even in those respects which are epistemically (currently or in principle) inaccessible to us.

I believe that the adherent of the completeness thesis (CT) should reply as follows:

> (CT) An analogical presumption belongs to our role of interpreters of literary works. The world in which the story of Flaubert's novel takes place is the world we must presuppose if Flaubert's text is to fulfill its lit-

[8] Reply to Rossholm (2015). All the page references which appear without the author's name refer to this paper.

erary functions. And this is a complete world in which complete human beings live in complete settings, find themselves in complete situations and take part in complete events, while the narrator naturally provides us with only an incomplete description of all this.

The dispute could then continue e.g. this way:

(IT) If you insist that the world of the novel is complete, the question arises in which mysterious way the author managed to create it, granted that he can provide us only with incomplete sets of descriptions of characters, objects, settings, events and so forth.
(CT) It is entirely sufficient that the author has written a text whose literary functions require the reader to presume a complete world as the world where the story of the novel takes place, where its protagonists live etc. In this way the author has established this world as the world of his novel.[9] Descriptions contained in the text of the novel have thus acquired the status of incomplete descriptions of complete individuals, settings, events and the like.

So far, the dispute concerned the ontology of fiction. But the adherents of (IT) can convert it to epistemological one, by arguing that the reader can hardly presume that the literary text provides her with incomplete descriptions of complete entities, when she knows that she has no chance of completing these descriptions in any respect that would go beyond the text. I believe that there is no substantial asymmetry between the "ordinary" and fictional talk even in this field – but this part of the defence of (CT) can be found in Koťatko (2010, 98) and Göran refers to it in his paper.

The arguments I have ascribed to the proponents of (IT) fit well together with Göran's "preliminary" answer to the question "Why do many scholars treat the factual and fictional discourses so differently with respect to incompleteness?" (p. 136). I find his explanation sharp and convincing.

---

[9] In fact, I believe that the world we are supposed to relate the text of narrative fiction to is the actual world: the narrative functions of the text require that the reader approaches its sentences (in the *as if* mode) as records of utterances of an inhabitant of the actual world (the narrator) who tells us what happened in this world. Cf. e.g. my reply to Zsofia Zvolenszky bellow.

This dispute with the proponents of (IT) requires certain important qualification – which brings us to the second part of Göran's paper. What I oppose is the claim that it *belongs to the very nature* of fictional entities and the world they inhabit that they are ontologically incomplete. But I do not want to deny that a piece of narrative fiction may present the world the narration is about and its inhabitants as essentially incomplete. If one believes, as I do, that the text of narrative fiction requires from us to assume (in the *as if* mode) that its story takes place in the actual world, the completeness of the world to which the narration refers is guaranteed in advance – but only on the presumption that the actual world is complete. Obviously, the same concerns the issue of contradictions: the coherence of the fictional world (despite the contradictory statements made occasionally by the narrator) can be guaranteed by its identification with the actual world – granted that the actual world does not include contradictions. I have repeatedly referred to Samuel Beckett as an example of an author who does not seem to share these presumptions and, as a writer, does not intend to create the illusion that they are right. The starting point is his account of the world as universal chaos ("I can't see any trace of a system anywhere") and the consequence is his rejection of narrative techniques which, according to his opinion, serve to conceal the chaotic nature of reality and construe an illusionary picture of an ordered, and hence coherent and complete world.[10] The main target of his criticism is (not surprisingly) Balzac, but he distances himself even from Kafka as not consequent enough: the way his characters behave is not compatible with the fact that the chaotic nature of the world does not leave space for personal integrity and continuous purposeful action.[11] Correspondingly, I agree with Göran that Kafka's world is not of the kind I have ascribed to Beckett. It certainly does not lack order:

---

[10] Just one illustration of the incompletness of Beckett's world: in the *Unnamable* it comes out that one cannot decide whether the names "Molloy", "Moran", "Malone" (and several others) refer to the same individual or to different individuals and whether any (or all) of them are identical with the narrator. This does not serve to illustrate the narrator's momentary indisposition but the nature of the world he lives in and the inscrutability of personal identity in such a world. The reader is not expected to presuppose that this world includes some facts, unknown to the narrator, which could decide such issues.

[11] "The Kafka's hero has a coherence of purpose… He's lost but he's not spiritually precarious, he's not falling to bits. My people seem to be falling to bits" (Shenker 1997, 162).

Kafka's characters are permanently confronted with signs of a hostile and impenetrable order and are endowed with a desperate desire to gain insight into it and find an efficient way of behaving within it. Moreover, as Göran's examples show, although their position might be tricky, for instance it may be unclear whether K. in *Trial* is or is not arrested (as a consequence of the introductory episode), this need not mean that the world they live in is incomplete (i.e. that its order is disrupted by gaps). Göran leaves this interpretation available as one of the alternatives (cf. p. 138) but I would opt for another one. According to it, this case shows that our dichotomy "arrested – not arrested" is not quite apt for describing the position in which K. finds himself (perhaps the same can be said about the position of some of our actual contemporaries): but precisely the way in which this dichotomy fails shows a lot about K.'s situation. In general, this part of Göran's paper demonstrates that our inability to answer certain questions concerning the characters or the story may play very different roles in the interpretation of a work of fiction and that our analysis should be sensitive to these distinctions.

I find equally illuminating the last part of the discussion Göran went into. If there is any point of disagreement, it would concern the degree of our reliance on what we are told in a text of narrative fiction and in everyday communication: I do not find the difference as radical as Göran (cf. p. 142). The trust in truthfulness of the assertions we are addressed in the ordinary discourse, is, according to some authors, anchored in the very nature of linguistic communication.[12] One need not share this view and still hold that trust is the default attitude which we abandon only under the pressure of evidence (or at least suspicion) that the speaker is not reliable. But even if this happens we have often a good chance to find out how things really are. For instance we may conclude that the speaker exaggerates her role in the events she is de-

[12] For instance, Lewis (1983) famously approaches linguistic convention as a "convention of truthfulness and trust". It follows that if in some moment the inhabitants of the United Kingdom cease to believe that the vast majority of assertive utterances of English sentences addressed to them in everyday communication are true, in that very moment English will cease to be conventionally fixed as the language spoken by the inhabitants of UK. Similarly, McDowell (1980) has argued that in such a case assertion (as a type of speech act) could not fulfill its basic function adopted from the prelinguistic forms of communication, namely the function of providing the addressee with an "epistemic surrogate" of direct experiencing the relevant state of affairs (specified in the propositional content of assertion).

scribing or that she misidentifies the cause of what has happened. Quite often we are able to recognize this without having any independent source of information about the subject matter in question – we simply register something suspicious in *what* the speaker says or in the *way* she expresses herself and draw on our general knowledge about how things usually go. And from the same resources we are often able to put together quite a reliable picture of how things really are in the case in question, contrary to what the speaker says. Obviously, the construction of a literary work may include reliance on our ability to exploit precisely these skills, acquired in everyday communication. One of the intended effects of a literary text may be a discrepancy between the assertions we are ascribing (in the *as if* mode) to the narrator on one side and our reconstruction (in the *as if* mode) of the actual course of events. In such cases, to use Göran's way of putting it, "the reader sees through what is said" (p. 140): and I fully agree that this (what the reader is supposed to see behind the words) is what we the reader is in the last instance told to be true – not by the narrator but by "the book", as Göran has aptly put it. In the most radical case I can think of, the message we receive "in the last instance" is that the nature of the world is such that it leaves no space for meaningful utterances. With respect to the fictional world of Beckett's texts it certainly holds – for cooperative readers – that this "is true because the book says so" (p. 140). But if we admit that Beckett's aspiration is to let us see the actual nature of the world we live in (to make the universal chaos visible to us), we have a good reason to raise the question of truthfulness again.

Manuel García-Carpintero[13]

Let me start with Manuel's account of the nature of the fiction-maker's creative acts: "…Joyce's utterance is not an assertion but a different speech act of pretending or make-believe which should be understood in terms of norms stating contents that proper appreciators of Joyce's tale ought to *imagine*" (p. 153). According to my view, the specification of the author's creative act is something which should follow from our reply to a more fundamental question: what do the literary functions of the text require from the *inter-*

[13] Reply to García-Carpintero (2015). All the page references which appear without the author's name refer to this paper.

preter? In other words, what are the interpretative moves the reader has to make in order to allow the text to fulfill its literary functions? Then we can characterize the writer's achievement so that she creates a text whose literary functions impose such and such requirements on the readers. As far as I can see, this does not mean, on the author's part, to *pretend* to make assertions, promises etc.,[14] neither to *actually perform* them, but simply to write sentences whose literary functions require that they are interpreted in certain way (namely, in case of narrative fiction, that they are read, in the *as if* mode, as records of the *narrator*'s assertions, promises etc.). And it is part of the author's being aware of what she is doing that she *intends* her sentences to function in this way.

Speaking about "requirements" or "the moves the reader has to make" may invite normative interpretation – but I do not think (apparently unlike Manuel) that the writer's creative acts or their results establish real *obligations* any more than the instructions for operating a washing machine do. According to my view, what these requirements or instructions in effect say is this: "If you want the literary text/the machine to fulfill its intended functions for you, you should do so and so (and avoid doing so and so)". Read in this way the instruction is purely instrumental. The introductory conditional clause in this hypothetical imperative certainly does not specify any wish we are obliged to have: the reply "No, I don't think this is precisely what I want" makes quite a good sense. Of course, it will be found inappropriate or even invite sanctions *in special contexts*, in which our discourse and connected non-linguistic actions fall within the domain of some *extra* norm, like "All students in this class (or: all inhabitants of this hostel) are obliged to interpret classical Czech literature (or: to use the owner's washing machine) in the way specified above!" Obviously, any type of discourse and any kind of activity in general can occasionally appear within the domain of various sorts of norms – but that in itself does not show that it is *intrinsically normative*.[15]

But no matter how much we may differ concerning the nature of the author's creative acts and their results, the main challenge to be found in Manuel's rich and subtle paper concerns the functions of proper names in the

[14] For an opposite view, see e.g. Searle (1979); Kripke (2013).

[15] This corresponds to one of the arguments to be found in Kathrin Glüer's and Åsa Wikforss' polemics with normativists in the theory of meaning – cf. my comments on their paper in the first part of this series of replies.

texts of narrative fiction and should be taken very seriously by all non-descriptivists in this field: "Even if Joyce's act is not an assertion but rather an invitation to his reader's imagination, the purported imaginings should nonetheless have contents; and non-descriptivists must tell us what, on their view, the contribution of names such as "Mr. Leopold Bloom" to such contents is" (p. 153). And on another place: "While the mode of thinking through which we think of Venus when we assert 'Hesperus is smaller than Mars' is intuitively and theoretically irrelevant to what we assert, in that many other modes of thinking about it may do as well, the corresponding modes of thinking 'about' Marlow and Holmes provided by the relevant fictions are essential to their contents: no proper appreciation can ignore them; no proper appreciation can do without building the corresponding files, starting with 'the object picked out by the relevant 'Marlow' naming practice', and stacking into it all the information about the character derived from the fiction" (p. 157).

Let me start with the suggestion (following from the preceding quotation) that the utterance of "Marlow is a clever detective" in Chandler's text has a descriptive propositional content which can be specified as "The object picked out by the relevant 'Marlow' naming practice is a clever detective". Or, as it is put elsewhere (p. 154): "The object called 'Marlow' is a clever detective". I take these two versions as equivalent and fully compatible with the view I will present bellow, even if my own presentation of the descriptive contribution of proper names to propositions (expressed by sentences containing the names) will be a bit different. But I do not share Manuel's conclusion concerning rigidity: "On this view, the semantics of textual uses is descriptive and hence singular terms *both empty and non empty* in them are not rigid" (p. 154; cf. also p. 157).

Everything depends on the function we assign to descriptions like "the object called 'Marlow'" – and the same concerns all the Marlow-descriptions we collect when reading Chandler's text. Let me recall (again – cf. my reply to Marián Zouhar) Kripke's famous example of a case in which the referent of a proper name is fixed by description: the story of Jack the Ripper (cf. Kripke 1980). As everybody knows, this name has been given to the person who committed certain brutal murders in Soho and has never been unveiled, so that "the person who committed those and those murders" is the only available way of identifying the bearer of the name. But as Kripke insists, this does not mean that the name just abbreviates this description: it is not so that with respect to any possible world *w*, the name refers to the person satis-

fying that description in *w*. Instead, the name refers, with respect to any possible world, to the person who satisfies that description in the actual world: hence even if the referent is fixed descriptively, the name is a rigid designator.

Now, to get closer to our original (Marlow) example, let us imagine that the identifying knowledge we can connect with the name is even poorer (and less spectacular) than in Jack's case. I witness a conversation in which the name "Jan Novák" is repeatedly uttered. It is quite natural to interpret the situation so that both speakers use the name "Jan Novák" for one of the hundreds of persons that have been baptized with that name and are continuously referred to by means of utterances of that name. This can be put so that I think about Jan Novák as the person uniquely satisfying the description: "the man who has been assigned the name 'Jan Novák' at the beginning of the chain to which these utterances belong (or: the chain activated in this conversation)". In this way I have descriptively fixed the referent of that name (as used in given conversation) with respect to all possible worlds, including those in which that person is not called 'Jan Novák'.[16] To that person I then assign various descriptions which I collect while following the conversation.

I believe it is the same with the name "Marlow" as used in Chandler's text and with the description "the person called 'Marlow'", or its equivalents. You may object that the mechanism described above cannot work in case of fiction, since here (unlike in Jan Novák's case) there is no way of identifying the world to which we should primarily apply the relevant description in order to pick out the referent of the name (as the person uniquely satisfying the description *there*) and to fix it with respect to all other possible worlds. Since there is no criterion which could select precisely one among all those possible worlds in which everything what is said in Chandler's text is true: all of them

[16] I fully agree with Manuel's insistence (on p. 159) that descriptive identification provides a good basis for thinking or speaking *about* an individual – not only in cases in which descriptions are used "refentially" in Donnellan's sense. The fact that it makes sense to say "the richest man in the world, *whoever he is*", is irrelevant in this respect, since you can equally well say "Jan Novák, whoever he is" or "this gentleman, whoever he is". The "whoever he is" clause just indicates that the mode of identification you have used is the only one available to you, not that it is insufficient for picking out the individual you want to speak about.

have the same right to be called "story-worlds" of Chandler's novel.[17] But it should be clear that we are in the same situation even in case of any "serious" (i.e. non-fictional) text like police report, newspaper reportage, theoretical article etc.: no such text *in itself* provides us with a criterion for selecting one among all those worlds in which all of its sentences are true. But obviously, no such problem arises in reality since the communicative function of these texts require that we automatically relate the assertions made in them and the referential function of the singular terms used in them to the *actual* world. I believe that we do (and are supposed to do) the same with the texts of narrative fiction.

Let me turn from Marlow to Emma Bovary, to refresh our imagination a bit, and consider what I am supposed to do when reading Flaubert's text. If it is to fulfill its narrative functions for me I have to approach its sentences (in the *as if* mode) as records of utterances of a real person (the narrator) who tells us what happened in the actual world. And it belongs to the role of the reader that I evaluate these utterances (in the *as if* mode) as true in the actual world – until the narrator proves to be (in some respect) unreliable.[18] Within this approach to Flaubert's sentences, I interpret (in the *as if* mode) the occurrences of the name "Emma Bovary" in them so that the narrator, in his uttering that name, links himself to the chain of uses of that name in the actual world. Hence I can identify (in the *as if* mode) the person the narrator speaks about as the individual uniquely satisfying (in the actual world) the following description: "the person who has been given the name 'Emma Bovary' in the act of baptism at the beginning of the chain to which the narrator's utterances belong". This description, including reference to the name "Emma Bovary" and to the narrator's utterances in which it occurs, can be labelled as "parasitic" or "derivative" or "nominal" or "metalinguistic" or "formal" in the sense that it is based on the general mechanism of the referential functioning

[17] If we stop at this point we will have to admit with Currie (2003) that "Marlow", as it appears in Chandler's text, is not a name of an individual but rather a name of an "individual role", interpreted as a function from possible worlds to individuals.

[18] If the narrator is construed as unreliable in that radical sense that she (consciously or unconsciously) misrepresents what happens, this requires from the readers to evaluate her utterances (in the *as if* mode) as false and to reconstruct what "actually" happened in contrast to what the narrator presents as having happened. As this and the preceding point show, I do not share Manuel's view that the utterances of sentences within literary texts are "not intuitively truth-evaluable" (p. 147).

of names rather than on factual information regarding the bearer of the name.[19] As the reader I presume (in the *as if* mode) that this formal description is satisfied by precisely one person in the actual world;[20] and it is the person identified in this way to whom I assign all the "non-parasitic" descriptions which I collect while reading Flaubert's text. They, indeed, enter into my picture of Emma and thereby into the way I think about her: hence I fully agree with Manuel's insisting on their relevance (e.g. on pp. 157, 164).

Surely, it would be bizzare to interpret this so that my Emma-thoughts are in fact thoughts about all those individuals which satisfy those descriptions in various possible worlds.[21] Instead, I am thinking (in the *as if* mode) about Emma as that individual which satisfies them in the actual world. And nothing should prevent me from thinking about *this very individual* (picked out in the way just mentioned) also with respect to other possible worlds, for instance to speculate about what would have happened with Emma if she did not meet Rudolph.[22] This shows that despite the crucial importance we have assigned to descriptions (both of the parasitic and non-parasitic kind), "Emma" behaves like a rigid designator. And it should be so, if we are supposed to approach (in the *as if* mode) the sentences contained in the literary text as records of utterances of the narrator inhabiting the actual world.[23] Since then we should assume that he (as well as all the characters) breathes like us, digests like us and also uses proper names in the way we do.

---

[19] This description can be quite naturally reduced to the form "the person called 'Emma Bovary'", if we do not find it necessary to account for the fact that "Emma Bovary" may have more than one bearer.

[20] To borrow Manuel's term (p. 161), this presumption can be called "reference fixing presupposition", with the addition that its acceptance by the reader is a matter of pretence (ibid.) or, as I prefer to put it, that it is made in the *as if* mode.

[21] Or, equivalently, about the Emma-role interpreted as a function from worlds to individuals (cf. Currie 2003).

[22] The alternative would be to think about Emma, when reading Flaubert's text, as about an entity which has its properties like being in love with Rudolph, being called "Emma", etc. necessarily – that means not to think of her as a human being like us. But then it is difficult to imagine how Emma's story could make a good sense to us, invoke empathy in us etc.

[23] Of course, this counts only for those who (like myself) believe that proper names, as they are used in everyday communication, are rigid designators.

Within the account I am advocating here, all this is not to be interpreted as a matter of an *import* from the actual world to the fictional world created by the writer; or, in Marie-Laure Ryan's terms (cf. e.g. Ryan 1991; 2010), from the "actual actual world" to the fictional world to which we move in our imagination, pretending to accept it as actual (within the operation called "re-centering"). According to this account, no such moves are needed: the text directs us to the actual world, and in this world everything remains as it is, except the changes required by the text. This, among other things, enables us to interpret the narrator quite straightforwardly as using English, Spanish etc. (rather than some fictional correlates of them), without having to transport (in our imagination) our linguistic practices or their results to some other world. Similarly, this enables us (and requires from us) to approach the names borrowed from the ordinary discourse as keeping their referents even when used in fiction.[24]

If sentences of a literary text contain a name like "Robespierre", as it is e.g. in Hugo's novel *Ninety Three*, we are supposed to deal with it in the same way as with the name "Emma Bovary". We presume that the narrator is speaking about a person who satisfies the formal (parasitical) description: "the person who has been given the name 'Maximilien de Robespierre' at the beginning of the chain to which the narrator's utterances belong". In addition to this, we assume (in the *as if* mode) that the chain to which the narrator has linked himself when uttering the name "Robespierre" is the same chain which the editors of Hugo's novel joined in their historical notes and the same chain which my history teacher joined when uttering the name "Robespierre" in his exposition of the French revolution, and the same chain that I joined when being examined at school. This assumption enables me to attach, when reading Hugo's text, to the name "Robespierre" not only the descriptions that I find in the text itself, but also the descriptions that I find in the editorial historical notes, as well as those which I manage to put together from my schooldays – in all cases on the condition that they are compatible with the descriptions provided by the literary text.

I believe that all this belongs to the way in which the text of narrative fiction is anchored in the actual world and relates our thought, imagination and sensitivity to this world. But I do not claim that the approach I am advocating

[24] So, my present claim about the fictional use of names like "Napoleon" or "Paris" is not based on the idea of "importation", as it is presented by Manuel (cf. p. 162).

here represents the only possible way of accounting for the fact that our historical knowledge about Robespierre, including those parts of it which are not mentioned in the text of the novel, can be exploited in our interpretation in a productive way. When discussing this point, Manuel points to the difference between things which belong to the content of fiction and things we have just to assume or imagine in order to make sense of the fiction:[25] let me put it (for short) so that the things of the latter kind belong to the interpretative "scaffolding" surrounding the fiction. I find this distinction productive: needless to add, the question always arises where to draw the dividing line in particular respects. For instance, I would say that the construct of the *implied author*, if it has any relevant application at all, belongs to the scaffolding. On the contrary, I would say that the referential function of the name "Napoleon", as I have described it, belongs directly to the way in which the narrator of *War and Peace* is supposed to tell us the story. This, if I am right, introduces the actual emperor into the content of the story (for Manuel's opposite suggestion see p. 161). But I am rather uncertain about the rest: as to our knowledge about the historical Napoleon, I would hesitate where to draw the line. Shall we say that everything that is not explicitly said or implicated by the narrator but is needed to make sense of what he says or what the characters say and do, or what is needed to work out the implicatures, etc., belongs to the content of the literary work – or does it belong to the scaffolding? And a considerable part of our historical knowledge certainly need not but may be exploited in our reading the text and appreciating the story we are told: that would perhaps require introduction of some third category.

Zsofia Zvolenszky[26]

Zsofia is right that what I have said about the role of fictional names remained on the level of their use within the texts of narrative fiction. Let me briefly resume my view on this and then to proceed to its implications con-

[25] "There are imaginative acts required to understand the text that are merely ancillary to the determination of the contents that the text invites proper appreciators to imagine" (p. 161).

[26] Reply to Zvolenszky (2015). All the page references which appear without the author's name refer to this paper.

cerning other kinds of use, discussed by Zsofia. The question is whether on some of these levels it comes out that I can, or even should, combine my approach with (some sort of) artifactualism, as Zsofia suggests (p. 176).

Textual use

(1) Andrei Bolkonski entered the room.

The question is what the reader is supposed to do with this sentence in order to allow Tolstoy's text to fulfill its literary functions for her. The reply I have suggested in my reaction to Manuel García-Carpintero's paper (and in some earlier texts) goes as follows: The reader is supposed to interpret, in the *as if* mode, the occurrence of this sentence in the text as a record of an utterance made by a real person (the narrator) who tells us what happened in the actual world. This requires from us to assume (in the *as if* mode) that the narrator uses the word "Andrei Bolkonski" in the same way in which we use proper names in ordinary communication, which means that he joins certain chain of uses of the name "Andrei Bolkonski", at the beginning of which that name has been assigned to particular person.[27] About that man the narrator claims (as we assume in the *as if* mode) that he entered the room – and on other places he makes other statements, including counterfactual ones, about that very man.

Notice that once we accept this approach, there is no space left for the problem of the identification of the person the narrator is speaking about (Zsofia is addressing this problem on p. 175). We are supposed to simply assume (in the *as if* mode) that it is the person satisfying (in the actual world) the description "the person who has been assigned the name 'Andrei Bolkonski' at the beginning of the chain to which the narrator's utterances belong". If somebody wonders how could Tolstoy succeed to identify precisely one person as Andrei Bolkonski, if there are, in various possible worlds, countless persons satisfying all the Bolkonski-descriptions to be found in Tolstoy's text, the reply should be very simple. The author succeeded to do so simply by producing a text the literary functions of which require us to make

[27] Obviously, this assumption does not exclude the possibility that there may be other chains at the beginning of which another person has been baptised with a phonologically identical name.

the move (to adopt the assumption related to the *actual world*) described above.

Paratextual use

(2) Bolkonski is a prince.

This can be quite naturally paraphrased in one of the familiar ways using some fiction-operator, e.g.

(2i) In Tolstoy's novel, Bolkonski is a prince.

But the real point at issue is how to interpret this paraphrase. I suggest to unpack it in the following way:

(2ii) The literary functions of Tolstoy's novel require us to assume (in the *as if* mode) that there exists a person referred to by the narrator as "Bolkonski" and that that person is a prince.

That certainly does not commit us to the existence of an entity called "Bolkonski", whatever its ontological status is supposed to be. Of course, one can say: Bolkonski is one of the persons we have to assume (in the *as if* mode) as existing, in order to make sense of Tolstoy's novel, but this does not commit us to any hypostasis – it amounts to saying that the literary functions of Tolstoy's text impose such and such demands on us.

Intertextual use

(3) Bolkonski is a less passionate man than Fabrizio.

Utterances of this kind can be, quite naturally, classified as a special case of paratextual uses of fictional names. I suggest to unpack our present case in the following way:

(3i) The man we are required (as readers of *War and Peace*), to assume as the bearer of the name "Andrei Bolkonski" exhibits less passion than the man we are required (as readers of *La chartreuse de Parm*) to assume as the bearer of the name "Fabrizio del Dongo". ("Exhibits" is here a shortcut for "exhibits in his behaviour described in the relevant novel".)

Nonexistence claims

(4) Bolkonski does not exist.

Here I opt for the following metalinguistic paraphrase:

(4i) The word "Bolkonski", *as it appears in Tolstoy's novel*, does not have any referent in the actual world (i.e. it is not a proper name of any real person).

This paraphrase is, due to the clause in italics, perfectly compatible with the possibility (a) that there is (in the actual world) a person called "Bolkonski", (b) that there is a person satisfying all the Bolkonski-descriptions we find in Tolstoy's text and (c) that there exists another text of narrative fiction in which the name "Bolkonski" is (equally like "Napoleon" in Tolstoy's novel) used to refer to a real person. I believe that no sound interpretation of (4) should exclude such possibilities.[28]

Metatextual use

(5) Bolkonski is a fictional entity.

The interpretation I am suggesting combines elements of the paraphrases of (1) and (4):

(5i) The word "Bolkonski", as it appears in Tolstoy's novel, does not have any referent in the actual world but the literary functions of the novel require that we assume (in the *as if* mode) the opposite.[29]

---

[28] One may object that (4) does not include any explicit reference to Tolstoy's novel. Generally speaking, we have two options. Either we approach (4) as used to straightforwardly claim that there exists no person called "Bolkonski" (and claims of such a kind have no relation to our present considerations). Or we understand (4) as implicitly related to the use of the word "Bolkonski" in certain context (in particular conversation, newspaper article, police report, novel etc.). Obviously, in our discussion the latter case is relevant and the context of the use is fixed in advance by our interest in the status of Tolstoy's characters.

[29] This is, of course, an interpretation in theoretical terms and is not supposed to show how the typical users of sentences like (3) would explain what they meant. Per-

As the suggested interpretation of (5) shows, I do not share Zsofia's view that "metatextual uses" (of words like "Bolkonski") "require us to include fictional characters as abstract artifacts in our ontology" (p. 173) and the same concerns all other kinds of use we have considered. So, I believe that we can make sense of all kinds of discourse considered so far without committing ourselves to any ontology including literary characters as abstract artifacts. What we assume is just the existence of texts the literary functions of which require from the interpreters certain moves – assuming or accepting or imagining certain things, including the existence of certain human beings (obviously, not abstract artifacts, but creatures of flesh and bone).

As far as the writer's achievement is concerned, it consists simply in creating a text with these functions. If they require that the readers imagine or assume (in the *as if* mode) a human being *X* with certain name, outlook, temperament, personal history etc., we can indeed say that the author has *created* this being – but this should be understood as a paraphrase of: the author created a text with such and such literary functions (raising such and such demands on the readers).

Let me stress that there is no animosity to abstract entities behind these suggestions. I do not have any problem with classifying the literary text (in opposition to its instances) as an abstract entity and the same concerns the literary work, understood as a structured complex of the text's literary functions. I just fail to see how the presumption of literary characters as abstract entities can help us to understand the way in which texts of narrative fiction work or to properly interpret the kinds of uses of fictional names we have been discussing above or to avoid problems we would otherwise inevitably fall into.

It should be clear that if understood in the way suggested above, neither of the statements (1) – (5) are made in the *as if* (or *pretense*) mode. Correlatively, if we make them sincerely, we *believe* (rather than *make-believe*) that they are true. Another thing is that these statements (except (1)) include some kind of reference to the literary functions of the text and *these functions*, as I understand them, require from the readers certain moves in the *as if* mode. So, I agree with Zsofia (p. 173) that there is no pretense inevitably involved

haps they would say something like: "'Bolkonski' is not a name of a real person but we are supposed to pretend that it is."

in uttering (5),[30] and unlike her I also believe that the same concerns (2). As construed above (in 2ii), an utterance of (2) is a full blooded claim about the literary functions of Tolstoy's novel, which does not involve any pretense of a commitment to the existence of Andrei Bolkonski. So, although I ascribe to pretense (or to the *as if* attitudes) an essential role in interpreting fiction and I do so within an approach which does not assume literary characters as abstract entities, I do not seem to be a "pure pretense theorist" of the kind criticised by Zsofia (with reference to Thomasson's arguments, p. 173; cf. Thomasson 2003).

I can only admire the delicate discussions Zsofia went into concerning the phenomenon of inadvertent creation of abstract artifacts, I agree that the abstract artifactualists concerning fiction should admit that this phenomenon is quite widespread and I believe that they should not be worried about this. So, I do not feel temptation to misuse the fact that Zsofia's subtle observations and arguments concerning this issue can be, as she notices, "easily turned upside down and construed as a new set of reasons for resisting artifactualism about fictional characters" (p. 183). Neither do I feel motivated to look for any other arguments against interpreting fictional characters as abstract artifacts. I am just unable to offer to them an appropriately decent and comfortable place within the approach I am proposing.

References

ARISTOTLE (1989): *Prior Analytics*. (Translated by R. Smith). Indianopolis: Hackett Publishing Company.

CURRIE, G. (2003): Characters and Contingency. *Dialectica* 57, No. 2, 137-148.

GARCÍA-CARPINTERO, M. (2015): Is Fictional Reference Rigid? *Organon F* 22, Suppl. Issue, 145-168.

KOŤÁTKO, P. (2004): Truth-Conditions and Literary Aspirations: Basic Correlations. In: Rossholm, G. (ed.): *Fiction and Perspective*. Frankfurt a. M. – New York – Paris: Peter Lang Verlag, 155-170.

KOŤÁTKO, P. (2010): Who Is Who in the Fictional World. In: Koťátko, P. – Pokorný, M. – Sabatés, M. (eds.): *Fictionality-Possibility-Reality*. Bratislava: aleph, 89-100.

[30] I say "inevitably involved" to account for the platitude that each sentence can serve to make pretended claims, for instance on the theatre stage or as part of some conversational play.

KOŤÁTKO, P. (2012): Radical Narration. In: Currie, G. – Koťátko, P. – Pokorný, M. (eds.): *Mimesis: Metaphysics, Pragmatics, Cognition*. London: College Publications, 178-193.

KOŤÁTKO, P. (2013): Text, Work, Interpretation: Some Implications of the Menard Case. In: Koblížek, T. – Koťáko, P. – Pokorný, M. (eds.): *Text and Work*. Praha: Literaria Pragensia, 54-67.

KOŤÁTKO, P. (2015): Sense, Reference, Speech Acts, Norms and Other Issues. Replies to *Organon F* Papers (Part I). *Organon F* 22, No. 4, 546-576.

KRIPKE, S. (1980): *Naming and Necessity*. Oxford: Blackwell.

KRIPKE, S. (2013): *Reference and Existence*. Oxford: Oxford University Press.

LEWIS, D. (1983): Languages and Language. In: *Philosophical Papers. Vol. 1.* Oxford: Oxford University Press, 163-188.

MCDOWELL, J. (1980): Meaning, Communication and Knowledge. In: Van Straaten, Z. (ed.): *Philosophical Subjects: Essays Presented to P. F. Strawson.* Oxford: Oxford University Press.

PETTERSSON, A. (2015): Can Texts Be Read? The Anatomy of a Paradox. *Organon F* 22, Suppl. Issue, 118-132.

ROSSHOLM, G. (2015): Contribution to Fictional Epistemology. *Organon F* 22, Suppl. Issue, 133-144.

RYAN, M.-L. (1991): *Possible Worlds, Artifficial Intelligence and Narrative Theory.* Bloomington – Indianapolis: Indiana University Press.

RYAN, M.-L. (2010): Fiction, Cognition and Non-Verbal Media. In: Grishakova, M. – Ryan, M.-L. (eds.): *Intermediality and Storytelling*. Berlin – New York: De Gruyter, 8-26.

SEARLE, J. (1979): The Logical Status of Fictional Discourse. In: *Expression and Meaning*. Cambridge: Cambridge University Press.

SHENKER, I. (1997): An Interview with Beckett. In: Federman, R. – Graver, L. (eds.): *Samuel Beckett. The Critical Heritage*. London: Routledge.

THOMASSON, A. (2003): Speaking of Fictional Characters. *Dialectica* 57, No. 2, 205-223.

TICHÝ, P. (1978): *De Dicto and De Re*. Philosophia 8, No. 1, 1-16.

ZVOLENSZKY, Z. (2015): Inadvertent Creation and Fictional Characters. *Organon F* 22, Suppl. Issue, 169-184.

Organon F 23 (1) 2016: 125-127

Response to Asger Kirkeby-Hinrup

FRITS PER WILLIOTH GÅVERTSSON

Department of Philosophy. Lund University
Helgonavägen 3. 221 00 Lund. Sweden
frits.gavertsson@fil.lu.se

In his "How Choice Blindness Vindicates Wholeheartedness" (2015) Asger Kirkeby-Hinrup, despite previous qualms (see Kirkeby-Hinrup 2014), argues that the account of free will proposed by Harry Frankfurt (1971; 1988) via its reliance on the notion of wholeheartedness is vindicated by recent empirical data from choice blindness experiments (cf. e.g. Hall – Johansson – Strandberg 2012; Hall – Johansson – Tärning – Sikström – Deutgen 2010; Hall et al. 2013; Johansson – Hall – Sikström – Olsson 2005; Johansson – Hall – Sikström – Tärning – Lind 2006).

In choice blindness experiments a test-subject is presented with a choice and asked to justify her preference for the choice made over the alternatives, but in the experimental manipulations the test-subject is presented with an alternative she did in fact not choose *as if* she had in fact chosen it. The *choice blindness effect* is that test-subjects rarely detect the manipulation but rather confabulate reasons preferring the option they did not in fact choose. The choice blindness effect has been demonstrated in different sensory modalities and across a myriad of social domains thus making the phenomenon seem rather pervasive. Such experiments, it has been argued, present a problem for free will since if we are, given choice blindness, blind to the outcome of our decisions and provide *post hoc* reasons for choices this might very well generalize to every decision we make so that conscious deliberation and reason responsiveness prior to a choice might have no impact upon the choices we end up making. Thus, choice blindness *prima facie* provides support for accounts favouring substantive revision, or complete abandonment, of the notion of free will while telling against ac-

counts that rely (heavily) on explicit cognitive processes such as deliberation and introspection.

Kirkeby-Hinrup's response is interesting in that it, unlike other attempts, does not rely on alternative interpretations of the data in question. Instead Kirkeby-Hinrup argues that Frankfurt's account of free will can sidestep objections based on choice blindness since it takes (potentially unconscious, fallible, not fully determined) wholehearted identification – understood as desires endorsed by higher-order desires followed by a higher order volition the effect of which is that the individual wishes her particular desire (the target of the higher-order desire) to be her *effective* desire (potentially by an automatic process) – as sufficient for the exercise of free will (Kirkeby-Hinrup 2015, 204-205). In addition, Kirkeby-Hinrup argues, the notion of wholeheartedness, while difficult to operationalize empirically, provide a useful metatheoretical concept for delineating the limits of choice blindness that both he and the choice blindness experimenters agree is intuitively there (see, e.g., Kirkeby-Hinrup 2015, 205; Hall et al. 2010).

While sympathetic to Kirkeby-Hinrup's general mode of argument I do believe that two comments are in order. Firstly, the line of argument pursued by Kirkeby-Hinrup generalises and is available to any account of free will or authenticity (i.e. the person's deepest and most genuine commitments and desires) as opposed to authority (i.e. psychological elements that represent the person as the author of his own life, of choices about what to do (see Lippert-Rasmussen 2003, 368)) that incorporates a mechanism for delineating authentic desires or volitions and that does not demand excessive reliance on conscious deliberative processes at the time of choice (but may well require or allow these to be active in the formation of said authentic states).

Secondly, it would appear that non-hierarchical – i.e. accounts that do not treat higher-order motivational states as embodying one's real self (see Arpaly – Schroeder 1999) – coherentistic accounts fare even better than hierarchical accounts of the kind proposed by Frankfurt. Non-hierarchical accounts completely sidestep the threat of regress pertaining to higher-order desires that often figure as an objection against accounts like Frankfurt's (see, e.g., Watson 1975, 217-218; Wadell Ekstrom 2005, 49; Lippert-Rasmussen 2003, 354). A coherentistic account, such as the one proposed by Laura Waddell Ekstrom (2005), that takes authenticity to be a matter of how well a given desire coheres with one's character (understood as e.g. a central nexus of beliefs, desires, volitions or preferences) can provide a mode of explanation as

to why *post hoc* justification occurs in cases of manipulation. Given that the strive for authenticity is understood in terms of coherence, attempts at bringing choices one has just made – or believe oneself to have made – into coherence with the rest of one's desires and preferences trough *post hoc* rationalisation is exactly what we would expect under the circumstances given such an account. This constitutes an explanatory route not obviously open to Frankfurt, or by extension Kirkeby-Hinrup.

References

ARPALY, N. – SCHROEDER, T. (1999): Praise, Blame, and the Whole Self. *Philosophical Studies* 93, No. 2, 161-188.

FRANKFURT, H. (1971): Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68, No. 1, 5-20.

FRANKFURT, H. (1988): Identification and Wholeheartedness. In: *The Importance of What We Care About*. Cambridge University Press, 159-176.

HALL, L. – JOHANSSON, P. – STRANDBERG, T. (2012): Lifting the Veil of Morality: Choice Blindness and Attitude Reversals on a Self-transforming Survey. *PloS one* 7, No. 9, e45457.

HALL, L. – JOHANSSON, P. – TÄRNING, B. – SIKSTRÖM, S. – DEUTGEN, T. (2010): Magic at the Marketplace: Choice Blindness for the Taste of Jam and the Smell of Tea. *Cognition* 117, No. 1, 54-61.

HALL, L. – STRANDBERG, T. – PÄRNAMETS, P. – LIND, A. – TÄRNING, B. – JOHANSSON, P. (2013): How the Polls Can Be Both Spot on and Dead Wrong: Using Choice Blindness to Shift Political Attitudes and Voter Intentions. *PloS one* 8, No. 4, e60554.

JOHANSSON, P. – HALL, L. – SIKSTRÖM, S. – OLSSON, A. (2005): Failure to Detect Mismatches between Intention and Outcome in a Simple Decision Task. *Science* 310, No. 5745, 116-119.

JOHANSSON, P. – HALL, L. – SIKSTRÖM, S. – TÄRNING, B. – LIND, A. (2006): How Something Can Be Said about Telling More than We Can Know: On Choice Blindness and Introspection. *Consciousness and Cognition* 15, No. 4, 673-692.

KIRKEBY-HINRUP, A. (2014): How to Get Free Will from Positive Reinforcement. *SATS, Nothern European Journal of Philosophy* 15, No. 1, 20-38.

KIRKEBY-HINRUP, A. (2015): How Choice Blindness Vindicates Wholeheartedness. *Organon F* 22, No. 2, 199-210.

LIPPERT-RASMUSSEN, K. (2003): Identification and Responsibility. *Ethical Theory and Moral Practice* 6, No. 4, 349-376.

WADDELL EKSTROM, L. (2005): Alienation, Autonomy, and the Self. *Midwest Studies in Philosophy,* 29, No.1, 45-67.

WATSON, G. (1975): Free Agency. *Journal of Philosophy*, 72, No. 8, 205-220.

Organon F 23 (1) 2016: 128-138

Jaroslav Peregrin: *Inferentialism: Why Rules Matter*
Palgrave Macmillan 2014, 272 pages

Peregrin begins his *Inferentialism* by making clear that its topic will be first and foremost the enigma of meaning, i.e. probably the most central one of analytical philosophy. With possibly different emphases, he is going to approach it from the point of view of Brandomian inferentialism. The author offers a look at inferentialism which is fresh and strikes particularly by its clarity. Peregrin profits from the debates about inferentialism which have taken place during the last years, especially as he chooses some of the most common and natural objections to it and attempts to show, by and large with success, that they are not fatal for the doctrine.

The book is divided into eleven chapters, each terminated by a short summary of the main points. The chapters are divided into two main groups. In the first one some more general philosophical issues related to inferentialism are discussed, the second one is dedicated to philosophy of logic.

1. General tenets of Peregrin's inferentialism

In the first chapter Peregrin points to the fact that other doctrines regarding meaning which come to mind more naturally, most prominently various forms of representationalism, face serious difficulties. We get a brief rehearsal of the problems. Typically, some important vocabularies, such as the logical one, are hard to accommodate into representationalism, as it is not clear what their referents are supposed to be. Even in the case of empirical vocabularies, the issue linked to the Quinean inscrutability of inference arise. Yet the main focus of this book is not so much to criticize the rival views. Peregrin points at the problems with other approaches mainly to convince us that it is legitimate to give a try to inferentialism, counterintuitive as it might be. The comparison with different approaches to the problem of meaning appears mostly implicitly, when some misconceptions regarding inferentialism are clarified. The doctrine thus presented then speaks rather for itself.

Inferentialism in general is based on the contention that meaning of a given sentence is given by its inferential relationships with the other sentences. The meanings of words consist in their contribution to the meanings of the sentences

they can occur in. A simple example would be the conjunction sign, the meaning of which is given by the obvious introduction and elimination rules. The inferentialism advocated here is fundamentally normative, as opposed to descriptive varieties. It is the inferential rules, that is what is correct and what is incorrect to infer from a given (set of) statement(s) which determines meaning, not the actual inferential moves we make. Thus an objection that inferentialism is mistaken, as we have to understand the meanings of linguistic units (that is primarily sentences) before we can make inferential moves between them, "loses its bite" when aimed at normative inferentialism. There still might be a story about the relation between the actual moves and their normative statuses to be told, yet Peregrin in fact adresses that later in the book, discussing the necessity of implicit rules.

To contrast inferentialism with the more intuitive representationalist accounts of meaning, an interesting example from law due to Alf Ross (see Ross 1957) is adduced, which exemplifies the typical anti-inferentialist convictions. This author shows that some words from his discipline, such as "ownership" are in a way no real words, as they do not refer to anything, but merely express the link between the conditions of their application and the consequences thereof (e.g. between buying something and having the right to bestow it on one's children).

There is simply nothing more to this word. Yet according to inferentialism the situation is the same with all words. How can such an account get off the ground in case of empirical vocabulary? We have to specify some important features of the inferentialism presented here.

2. Pragmatism, holism and the empirical vocabulary

First of all, the inferences countenanced cannot be only the ones sanctioned by (some) logic. They have to include also the rules which are called, in the Sellarsian tradition, material (and these are then not seen as enthymemes). And furthermore there is no principled distinction of analytical and synthetic inferences. Sometimes quite empirically looking inferences have to be accepted as, in fact, indispensable for meaning of certain words. The conceptual framework and the contents we fill it with cannot be clearly demarcated. As every statement can be made immune to revision, so every statement can be sacrificed.

And even the most general features of our conceptual schemes are not completely free from the influence of empirical world, though such influences can be described only indirectly.

To accommodate the empirical vocabulary, though, we have to countenance also the rules which somehow connect the language with the non-linguistic real-

ity. The worry is that the two realms are just too heterogeneous to make something like that possible. The realm of causality and the realm of reasons are radically different. Yet this dilemma, which is shown with particular vivacity by McDowell (in his famous 1994), can be relieved by abandoning the supposition that the difference between the linguistic and the non-linguistic is a difference between what is inside and outside. Language is essentially public and moreover it is a system of embodied rules. This feature makes it more analogous to, e.g., football than to chess. It is not only the proper inferential transitions we make between sentences in a language but also the Sellarsian language-entry and language-exit transitions which are constitutive of meanings. As such the system of inferential rules is not blind to the causal realm.

Thus also the worry about spinning in the void is relinquished. The notion of the inferential rules connecting language with the world might not be for everyone's taste. Perhaps more discussion is needed regarding this specific point. Yet Peregrin shows that the notion of such rules is not particularly mysterious. Not much more than football, that is.

3. Wittgensteinian motifs

Saying that meaning is normative is not incompatible with our specifically human freedom. Rules of language are rather restrictive, not descriptive, they are something we can "bounce of" to do something we could not do without them, i.e. perform various speech acts. Furthermore, saying that such and such an expression has such and such a meaning is a specific speech act. Not only is it reporting some fact about the linguistic habits of the given community but it is also endorsing them. We can contribute to making the claims about meaning true by the very saying them.

Peregrin develops the Wittgensteinian turn to pragmatics in the philosophy of language. On the example of Lorenzen's game-theoretic accounts of semantics of logical constants (cf. Lorenzen – Lorenz 1978) he exhibits the general idea of meanings being instituted by sets of rules regulating a dialog. The rules regulating the use of non-logical vocabulary in our language games are overall more complex, though.

Unlike those of chess the rules of language cannot all be formulated fully explicitly, since this would already have to presuppose language. Before being able to say that thus and so is a correct or incorrect usage of given expression, we have to be able to treat some usages as correct or incorrect implicitly. Here I think we get more nuanced view on the distinction between the normative and descriptive inferentialism Peregrin describes at the outset of his book. Saying that it is the

rules of inferring rather than our actual acts of inferring which constitute meaning might sound as making the actual acts irrelevant. But they cannot be, as they contribute to the institution of the rules.

4. Relationship with natural sciences and evolution theory

Peregrin wants to make inferentialism plausible not only by confronting it with specifically philosophical discourse (and later with logic) but also with natural sciences and showing its relationship with them. It is true that we can describe a given community and its behaviour including normative stances towards inferences. Yet we also have to regard ourselves some inferences as valid or otherwise. Thus meaning cannot be captured exhaustively by natural sciences (even despite their possible great future achievements). Yet they are not banned from contributing to the study of it, either.

Peregrin discusses the possible ways language and rule-governed behaviour could have emerged in the course of evolution. He relates this to the question about the origins of cooperative behaviour (and altruism) in general. His solution is basically to turn to a more holistic perspective. Following, e.g., moral rules might seem a bad survival strategy from a perspective of individual situation, yet from a perspective of series of similar situations the picture changes because being a member of the community of rule-followers opens new ways of coping with hardships of life. The same holds in particular for the rules of language. And even if they might seem arbitrary seen individually, taken as a whole system they open new dimensions to us which it is already not arbitrary to enter. Peregrin is thus not a naturalist though he shows that inferentialism does not postulate anything supernatural.

5. Inferentialism and logic

Logic is for inferentialists most naturally presented in the proof-theoretical framework of the calculi of natural deduction or the sequent calculi. Though inferentialism seems to be a foe of model-theory, Peregrin does not shun set-theoretical constructions as explications of the inferential rules. The question for him is not whether model-theory or set-theory should be used in formal explication of our language, but rather how to interpret them. His interpretation is clearly not the one which would offer itself prima-facie. As an inferentialist he does not want to accept the relation of reference as explanatorily primary to the relation of inference. The notion of truth is not primary for him, as well. He understands it as something constituted by the notion of correct inference. The model-theoretical

constructions are thus ways, though rather indirect ones, of examining the inferential roles of expressions.

This is rather unproblematic when we have completeness theorems showing that the proof-theory and model-theory are basically doing the same thing.

Yet there are obvious problems with this. Take Gödel's incompleteness theorem which appears to show that we necessarily need model-theoretical devices which transcend the possibilities of the proof-theoretical ones. It would be too long a detour to go into details here (yet it has to be done elsewhere!) but Peregrin offers a very inspiring and illuminating account of logic, according to which the intuitionistic propositional logic can be seen as the very core logic because it most naturally arises from the inferentialist demands on logic to express the material inferential relations. The stronger logics, ultimately even the second-order logic or modal logics, are shown as a ways of relaxing demands put on the notion of inferential rules. The idea of different logics being truly logical in degrees with some logics as the core ones is in general a very appealing one. It depends how seriously one takes the attempt to demarcate what still is and what is not a logic anymore. But at least a very nice and neat overview and even systematization of some of the most important achievements of logic is reached.

Peregrin thus does a lot of great work in reconciling inferentialism with model theory and set theory, yet his account encourages further development.

And not just an inferentialist philosophy of logic but also one of mathematics is needed. Furthermore, some logics are left untouched regarding their status for inferentialism, e.g. the paraconsistent logics. But is this really an objection?

Obviously the book would have to be blown up considerably to deal with all these issues. Peregrin shows that inferentialism is suitable philosophical background for logic and for more concrete issues of philosophy of logic.

6. What is logic good for?

In the last two chapters Peregrin addresses two related questions. How can the logical rules be justified? And what is the overall purpose of logic?

Peregrin frames the topic of justifying logical rules into the mould of the dilemma of triviality and contingency. He chooses the first horn, saying that the logical rules are in fact trivial. At least, in the sense that they are not to be really discovered. Modus ponens, for example, simply belongs to the very concept of the conditional. Thus it actually does not make sense to doubt whether it is valid. This seems to be a rather obvious consequence of inferentialism about meaning. Without rules of the kind of modus ponens, we can hardly even identify an ex-

pression as a conditional. The issue of substantiating the rules thus cannot arise in the first place.

This is basically a Quinean move (cf. Quine 1986) and surely has got a lot to recommend it. Yet Quine uses a much more controversial example, which, though he himself does not really acknowledge it, also shows that this solution might not be fully satisfactory. Alternative logics give alternative answers to the questions regarding the validity of certain logical laws. The Quinean answer would be that they really speak of different concepts, when, e.g., the classical logician holds tertium non datur for valid, while the intuitionistic one does not. One speaks of the classical, the other of intuitionistic disjunction and negation. Yet this attempt at complete overthrowing the debates about logical laws is hardly acceptable. That the dispute between rivalling logics can be real is something which is difficult to doubt. Yet Peregrin probably does not want to tackle the specific issue of rivalry of logics in this chapter.

In the last chapter the normativity of logic is clarified. If logic is to be normative – and this is what Peregrin repeatedly made clear to be a part and parcel of his view – then it seems obvious that it has to guide our reasoning or argumentation practices in one way or another. MP thus seems to tell us what to do with the conditionals. But after reading the previous chapter it should be clear that this is not really Peregrin's view. At most it could be seen as a very misleading statement of his account.

Logical laws such as MP do not basically tell us what to do, e.g. what to say with the conditional, they rather tell us what the conditional is. In this sense they are rather like the constitutive rules of chess and unlike the tactical rules of the same game, which indeed hint us at doing smart steps during the game. They do not tell us how to maximize the number of true beliefs, they rather enable something like the concept of truth to emerge at all.

I have already mentioned that truth emerges out of the rules of inference which we create. This, at first, sounds like dangerous idealism. Yet we have to remind ourselves that not all the inferential rules can be explicit. Perhaps each one of them can be made explicit but we can never get rid of the rules which are only implicit to our practices. And such rules are followed, as Wittgenstein put it, blindly. Therefore they obviously are not consciously stipulated by us. All in all, it cannot be said that we simply decide what is and what is not true, as the immediate worry might be.

Someone might perhaps still protest that in order to understand the fact that thus and such an inferential move is correct, we already have to understand the notion of the truth, that is the truth exactly of the statement that the statement about the correctness of the move is true. Peregrin does not discuss this possible

objection (which can be linked to Frege's remarks about the irreducibility of the notion of truth in Frege 1918) but I think it does not have to be fatal for him. The notion of truth has to be understood, but it is enough that it be understood only implicitly. And the implicit understanding does not have to involve the treatment of the sentence and regarding it as true or otherwise.

The deceiving necessity of understanding logic as giving us directives about how to reason stems mostly from the need to somehow clarify the purpose of logic, i.e. why should it be better to have logic at all. In Peregrin's understanding, largely influenced by Brandom, logic is in a way both prior and subsequent to the rest of the meaningful discourse. That is, it is prior implicitly, as the sounds we emit cannot be regarded as truly meaningful without their standing in logical relationships, such as entailment or incompatibility. Yet logic as a tool of making these relationships explicit can come to the fore only after they exist (which of course does not mean that acquiring logic cannot lead us to changing the relations afterwards, exactly because we are then capable of judging them).

Asking what is the purpose of logic it thus very close to asking what is it good for to live as meaning-mongering creatures at all.

Pavel Arazim
pavel.arazim@centrum.cz

References

FREGE, G. (1918): Der Gedanke. Eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus* 2, 58-77.
LORENZEN, P. – LORENZ, K. (1978): *Dialogische logik*. Darmstadt: Wissenschaftliche Buchgesselschaft.
MCDOWELL, J. (1994): *Mind and World*. Harvard University Press.
QUINE, W. V. O. (1986): *Philosophy of Logic*. Cambridge: Harvard University Press.
ROSS, A. (1957): Tû-tû. *Harvard Law Review* 70, 812-825.

Theodore Sider: *Writing the Book of the World*
Oxford University Press 2011, 318 pages

This book brings arguments for metaphysical realism, and its aspiration appears to be to make them knock-down. Sider tries to convince the reader that we can identify the structure of the world; he believes that we can reach the access to

its fundamental level if we adopt a modified version of David Lewis's theory of naturalness. Such a high aspiration may easily lead to a deep disappointment. The present book, fortunately, is not deeply disappointing, though, unfortunately, I would not say it is so persuasive as to live up to its aspirations.

The book consists of thirteen chapters, but only in four "core" chapters (1, 2, 6, and 7) Sider explicates the idea of structure and fundamentality. In the rest of the chapters he focuses on possible applications of his ideas and the comparison of his theory with those of his rivals. Although someone interested in the particular topics Sider discusses (quantifier realism, philosophy of time, laws, intrinsic properties, modality…) may find them interesting, they do not really reinforce the main ideas introduced in core chapters. This imbalance between the space for presenting the idea of structure/fundamentality and its applications leads me to my first critical point. Only arguments presented in the core chapters back up Sider's views about fundamentality in a "positive" way – by means of explicit arguments. Arguments in the rest of the chapters are mostly meant to support his views by showing what we can potentially gain if we adopt his realism about structure. This may not be a bad idea by itself. But showing potential gains is not enough to convince the reader who does not believe that the kind of realism is well founded.

Sider calls himself a knee-jerk realist. He believes that we must conform to the world and the best way how to do it is by means of physics; especially by those parts of physics which try to explain our world on its most fundamental level of quarks and properties like mass and charge. At this point, Sider follows David Lewis and his theory of naturalness. According to Lewis, we can use the criterion of naturalness to decide which of the possible meaning candidates has relatively highest degree of naturalness and should be understood as the meaning of a predicate. The degree of relative naturalness of a meaning candidate (i.e. property) is a matter of the syntactic complexity of its definition in fundamental terms (mass and charge) – the lower the syntactic complexity, the higher the degree of naturalness. Let's say we want to find the most natural candidate for the predicate 'being lithium'. To achieve this we have to track the chain between what we call lithium to its micro-physical properties – mass and charge. The point is that if we have a "gruesome" candidate for this role (lithium until the next year and then iron), the definition will be longer than the definition of the "intended" meaning (because we must state all fundamental properties of lithium *and* of iron) and therefore it should not be understood as the meaning of the predicate 'being lithium'. But this is also the last point in which Sider follows Lewis.

According to Sider, predicates which are closest to fundamental properties carve the world at its joints. In other words, the distinctions suggested by the fun-

damental terms are in accordance with the structure of the world. The existence of the structure is supported by the "Quinean argument". According to this argument, we should accept that the ontology of our best theory (the theory of physics in our case) is in accordance with the structure of the world. But unlike Quine, Sider believes that the best theory vindicates also other notions included in our ideology (in particular: mathematical and logical notions) and therefore we should accept them as a part of the structure of the world too. Terms, in which the ideology is stated, are vindicated because they take their part in the success of the theory. This also means that – in contrast to Lewis – we should broaden the scope of fundamentality beyond predicates to logical and mathematical terms, because they are indispensable for the success of our best theory.

The acceptance of the whole project of Sider's realism relies on the willingness of the reader to accept this argument. Therefore it should be subjected to a high level of scrutiny. But as far as I can see Sider takes it for granted – his only attempt to back it up is by showing potential deficiencies of explanations of related topics (e.g. substantivity or modality) if we try to explain them without reference to the notion of structure.

But even if we are willing to accept Sider's argument and we accept that there *is* the structure of the world and we *can* identify its fundamental parts – proposed by physics, logic and mathematics – there are other important issues which emerge and which are not addressed in the book at all. If we accept mathematical and logical entities as fundamental, how does this influence the way in which we define the degree of naturalness of properties? As I said earlier, the degree of naturalness is stated by the syntactic complexity of definitions in *fundamental terms*. Since mathematical and logical terms are also fundamental, they don't need a definition. They are defined, as it were, "by themselves". But does this mean that numbers can be included in definitions of some ordinary language predicates like 'being human'? What is their role in those definitions in comparison to mass and charge? We can state the "quantity" of mass in some physical units, but what is "the unit for mathematical entities"? Moreover, logical terms have already been included in definitions – conjunction and disjunction were used to relate masses and charges of different physical particles within one definition. Did the position of logical terms in definitions change after we accepted them as fundamental? Can they be included in definitions not only as connectives, but also as fundamental terms?

I understand that if Sider wants to claim that logical terms are fundamental, he has to answer challenges stated by his opponents who undermine the fundamentality of some logical notions. But this is still only a strategy of showing gains. The final upshot would look far more convincing if it were based on a sat-

isfactory and more detailed explanation of "internal relations" between fundamental parts *within* the structure.

Sider could respond that his main aim is not a metasemantic theory for ordinary language; so he doesn't need to pay much attention to possible problems with fundamental definitions of ordinary terms. His aim is to build the Ontologese – a language of metaphysics which is primarily stated in fundamental terms. This is certainly true, as well as that the main feature of the Ontologese is an independence of its (meta)semantics from the (meta)semantics of ordinary language and vice versa. But despite their independence, Sider believes that there is no theoretical obstacle which could stop us in providing definitions of our ordinary terms in terms of the fundamental ones. Moreover, even the Ontologese will probably involve less-than-fundamental terms which need to be defined. And this seems problematic, until we answer basic questions about relations of fundamental terms.

This said, I must also stress that the book has several positive aspects. With regard to the length of the review, I am not going to focus on the discussion of possible applications of Sider's notion of a joint-carving/structure. Some of them could surely present a valuable contribution to particular topics – e.g. his application of joint-carving within Bayesianism to setting the prior probability distribution or his discussion of deflationism in terms of quantifier variance.

However, the most positive aspect of Sider's realism lies in its most general motivations. And as far as I can tell, those motivations are the right ones. First of all, it's the idea that being a realist doesn't mean to have a ready-made answer for every metaphysical question. It requires patient reconsidering of particular questions case by case. Secondly, it is the idea that it would be naïve to suppose that we can answer any metaphysical question without looking at the world – by making stipulations regarding how the world must and could be or by conceptual analysis of the ordinary language. If we want to say something about the world we must build on what we have already found out and look for answers we still don't have. Thirdly, it is the defeasible character of his views. Sider doesn't try to pretend that he uncovered the eternal and unchangeable truth about the world and its structure. His main assumption relies on the notion of the best theory of the world. And what Sider proposes is based on what our best theory tells us about the world. If we are forced to change the best theory, he is willing to accept consequences and possibly modify or abandon his views without any hesitation. This may sound like a weak point of the book, but it is actually its highest virtue. It is a step in a good direction for realism in metaphysics.

Furthermore, the book wants to offer something more than to be a one of many metaphysical theories. Throughout the book Sider draws a line – the line

delineating boundaries of metaphysical investigation. We can look at the book as on an attempt to create a safe ground for metaphysics. A ground deprived of any debts to ordinary language, modality or any other philosophical investigation. It is an attempt to redefine the scope and – most importantly – the methodology of metaphysics. Building "toy" metaphysical languages (meant as some earmarked parts of the Ontologese) could be a way how to approach various questions in metaphysics. After all, the book offers several illustrious examples. As Sider says – you are invited to enter the metaphysics room. The Ontologese and the methodology of the toy languages definitely need some enhancements to play the role of a common ground for metaphysics. But it is up to the reader if she wants to step in and try to improve it from within.

The book offers an interesting metaphysical theory advocating realism about structure and it is definitely obligatory reading for anyone interested in current metaphysics. It is a book with a potential to change a course of metaphysics in nearby future. But there are still open issues which need to be elaborated in Sider's idea of structure – mainly the role and relations of mathematical and logical fundamental terms. For now, I prefer sticking with a more sceptical view about the possibility of writing the book of the world.

Matej Drobňák
matej.drobnak@gmail.com

Organon F 23 (1) 2016: 139-140

Modal Metaphysics: Issues on the (Im)Possible III
(Institute of Philosophy, Slovak Academy of Sciences, Bratislava, September 16-17, 2015)

It is hard to believe that Bratislava hosted yet another instalment of a conference whose topic included, but did not exhaust, contemporary developments in the metaphysics of modality. *Modal Metaphysics: Issues on the (Im)Possible III* was organized by the Institute of Philosophy of Slovak Academy of Sciences, Slovak Metaphysical Society and Slovak Philosophical Association. This year was especially "impossible" due to the more or less recent work of our plenary speakers: Daniel Nolan (Australian National University, Australia) and Mark Jago (University of Nottingham, UK).

As usual, the list of contributors and commentators (henceforth in brackets) indicated that the focus of the conference was on metaphysical, logical, epistemological and methodological problems of modality. The first day started with Andy Yu's (Oxford University, UK) "*The Indefinite Extensibility of Proposition*" and Andriy Vasylchenko's (National Academy of Sciences of Ukraine, Ukraine) "*Identity and Existence in Intentionally Possible Worlds*" and continued with Sam Cowling's (Denison University, USA) "*Conceivability Arguments for Haecceitism*" (Jonathan Livingstone-Banks), Peter Marton's (Clark University, USA) "*Knowing Possibilities and the Possibility of Knowing (A Further Challenge for the Anti-Realist)*" (Igor Sedlár), keynote given by Daniel Nolan entitled "*Conditionals and Curry"*, Karen Green's (University of Melbourne, Australia) "*Natural Language and Ontological Illusions*" (Darragh Byrne/Naomi Thompson), Igor Sedlár's (Comenius University in Bratislava, Slovakia) "*Impossible Worlds in Epistemic Logic*" (Peter Marton), Louis deRosset's (University of Vermont, USA) "*Modal Logic for Contingentist Metaphysics*", Darragh Byrne's (University of Birmingham, UK) and Naomi Thompson's (University of Hamburg, Germany) "*Is the World Really Hyperintensional?*", Michael De's (University of Konstanz, Germany) "*Five-dimensionalism*" (Theodore Locke) and Brian Ball's (Oxford University, UK) "*Modality and Metaontology*" (Amy Karofsky).

The program of the second day commenced with a couple of talks, namely Nikk Effingham's (University of Birmingham, UK) "*Heterodox Ludovicianism*" (Louis deRosset) and Theodore D. Locke's (University of Miami, USA*)* "*Grounding and Impossible Worlds*" (Brian Ball). It then continued with Amy

Karofsky's (Hofstra University, USA) "*The Impossibility of Otherwisedness*" (Sam Cowling), Zsófia Zvolenszky's (Eötvös University, Hungary) "*Inadvertently Created Fictional Characters Are Innocuous*" (Jonathan Nassim), a keynote address by Mark Jago entitled "*Three Roads to the Impossible*", Jonathan Livingstone-Banks's "*Essence and Possibility*", Johannes Bulhof's (McNeese State University, USA) "*The 'Problem' of Alien Properties*" and finished with Cristina Nencha's (Northwest Philosophy Consortium, Italy) "*Essentialism and David Lewis*" (Andriy Vasylchenko).

We've been always thinking about the "Issues" as a conference that fills a certain gap in (at least) the region of central Europe. The experience shows however that it transcends Europe and attracts philosophers from all around the world. That gives us reasons to continue in organizing it and, of course, we will. In fact, we have already started and the reader can check out how it goes at our sites www.metaphysics.sk.

Martin Vacek
martinvacekphilosophy@gmail.com

Errata

The following list involves errors which inadvertently appeared in Duško Prelević's article "Modal Empiricism and Knowledge of De Re Possibilities: A Critique of Roca-Roye's Account" (*Organon F* 22, No. 4, 488-498):

Location	*Error*	*Correction*
p. 489, lines 7-8 from the bottom	"for explaining successfully *a posteriori* cases"	"for explaining successfully *de dicto* cases"
p. 489, lines 4-5 from the bottom	"*a posteriori* and *de re* reading respectively"	"*de dicto* and *de re* reading respectively"
p. 489, lines 3-4 from the bottom	"explaining *a posteriori* reading of necessary *a posteriori* statements"	"explaining *de dicto* reading of necessary *a posteriori* statements"
p. 490, line 2 from the top	"(*a posteriori* reading")	"(*de dicto* reading)"
p. 495, line 22 from the top	"the possibility *a posteriori*"	"the possibility *de dicto*"
p. 495, line 23-24 from the top	"knowledge of modality *a posteriori*"	"knowledge of modality *de dicto*"

We would like to apologize to the author as well as to the readers for the above mistakes.

Editors